The genetic background of bovine αs1- and αs2-casein phosphorylation

Zih-Hua Fang

▶ **To cite this version:**

Zih-Hua Fang. The genetic background of bovine αs1- and αs2-casein phosphorylation. Animal genetics. Institut agronomique, vétérinaire et forestier de France; Wageningen Universiteit en Researchcentrum (Wageningen, Nederland), 2017. English. NNT : 2017IAVF0020 . tel-02879335

HAL Id: tel-02879335

https://pastel.hal.science/tel-02879335

Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2017IAVF0020

THESE DE DOCTORAT

préparée à l'Institut des sciences et industries du vivant et de l'environnement (AgroParisTech)

pour obtenir le grade de

Docteur de l'Institut agronomique, vétérinaire et forestier de France

Spécialité : Génétique Animale

École doctorale n°581
Agriculture, alimentation, biologie, environnement et santé (ABIES)

par

Zih-Hua FANG

The genetic background of bovine α_{s1}- and α_{s2}-casein phosphorylation

Directeur de thèse : **Patrice MARTIN**
Co-encadrement de la thèse : **Marleen VISKER**

Thèse présentée et soutenue à AgroParisTech 16 Rue Claude Bernard, 75005 Paris, le 15 Decembre 2017:

Composition du jury :

M. Etienne VERRIER, Professor, AgroParisTech - UMR GABI	Président
Mme Lotte BACH LARSEN, Professor, Aarhus University	Rapporteur
M. Tiny Van BOEKEL, Professor, Wageningen University & Research	Rapporteur
M. Fabio PILLA, Professor, Università degli Studi del Molise	Rapporteur/ Examinateur
M. Laurent SCHIBLER, Directeur de Recherche, Allice	Examinateur
M. Henk BOVENHUIS, Professor, Wageningen University & Research	Co-Directeur de thèse
Mme Marleen VISKER, Researcher, Wageningen University & Research	Co-Directrice de thèse

INRA-UMR 1313 GABI
Domaine de Vilvert, F-78352 Jouy en Josas, France

Animal Breeding and Genomics
Droevendaalsesteeg 1,6708 PB, Wageningen, NL

Contents

Chapter 1 General Introduction **5**

Chapter 2 The relationships among bovine α_s-casein phosphorylation isoforms suggest different phosphorylation pathways **19**

Chapter 3 Genetic and nongenetic factors contributing to differences in α_s-casein phosphorylation isoforms and other major milk proteins **37**

Chapter 4 Genetic parameters for α_{s1}- and α_{s2}-casein phosphorylation isoforms in Dutch Holstein Friesian **63**

Chapter 5 Genome-wide association study for α_{s1}- and α_{s2}-casein phosphorylation in Dutch Holstein Friesian **83**

Chapter 6 General Discussion **103**

Summary **119**

Acknowledgments/Remerciements **123**

Curriculum vitae **126**

Abstract **134**

Résumé **135**

Chapter 1

General Introduction

Protein phosphorylation regulates nearly every aspect of cell life, including disease states, by altering the structural confirmation of proteins to activate, deactivate or modify their function. Caseins from cows' milk are the well-studied group of phosphoproteins and play an important role in human nutrition as well as manufacturing of dairy products, such as yogurt and cheese. Phosphorylation of caseins is a crucial post-translational modification that allows caseins to form colloid particles called casein micelles, which enables transport of large amounts of calcium and phosphate above what is normally soluble. As a result, these minerals can be delivered efficiently to the neonate without damaging the mammary gland of the mother by evoking either pathological calcification or amyloidosis (Holt et al., 2013). Using caseins as model phosphoproteins can increase our knowledge of protein phosphorylation. Besides the high biological importance of phosphorylation of caseins, the formation and stability of casein micelles are highly relevant to manufacturing of dairy products. Therefore, it is of great interest to explore the variation in phosphorylation of caseins, in particular α_{s1}- and α_{s2}-caseins that are heavily phosphorylated, and study to what extent genetic and other factors contribute to this variation. This thesis aimed to investigate the genetic background of bovine milk protein composition with a focus on phosphorylation of α_{s1}- and α_{s2}-caseins.

1.1 Bovine milk protein composition

Bovine milk contains 3 to 4 % protein, of which 90% comprises the six major milk proteins, i.e. the two whey proteins [α-lactalbumin (α-LA) and β-lactoglobulin (β-LG)] and the four caseins (α_{s1}-, α_{s2}-, β-, and κ-CN). The caseins account for around 80% of total protein in cows' milk and are the phosphoproteins that precipitate from raw skim milk by acidification to pH=4.6 at 20°C (Jenness et al., 1956). The relative amounts of α_{s1}-, α_{s2}-, β-, and κ-CN occur in a ratio of approximately 4.0:1.0:3.5:1.5 in cow's milk. Most of caseins in milk, together with calcium, phosphate and other minerals, are assembled in colloidal structures known as casein micelles. The casein micelles act as transporters of calcium and phosphorous. Moreover, the casein micelle properties, such as their size and structure, have a major influence on the technological properties of milk (Glantz et al., 2010; Gustavsson et al., 2014; Logan et al., 2015). For example, the destabilization of casein micelles by rennet or acid to form gel networks is highly relevant for manufacturing cheese and yogurt. As for the whey proteins, α-LA and β-LG are soluble at pH=4.6. The β-LG is the most abundant whey protein in cows' milk, accounting for around 10% of total protein or 50% of whey proteins. The α-LA has an important biological function as a regulatory protein in the biosynthesis of lactose (Brodbeck and Ebner, 1966; Brodbeck et al., 1967).

Different methods have been developed to analyze milk protein composition: gel electrophoresis (Ng-Kwai-Hang and Kroeker, 1984; Seibert et al., 1985), high-performance liquid chromatography (HPLC) by ion-exchange (Hollar et al., 1991), hydrophobic interactions (Bramanti et al., 2002) and reversed-phase methods (Visser et al., 1991; Bobe et al., 1998; Bordin et al., 2001; Bonfatti et al., 2008), capillary zone electrophoresis (de Jong et al., 1993; Recio et al., 1997; Heck et al., 2008), and liquid chromatography coupled with mass spectrometry (LC-MS; Léonil et al., 1995). In general, gel-based methods are less suitable for quantifying milk

protein composition on a large scale due to the difficulty in automation, poor dynamic range and limited sensitivity (O'Donnell et al., 2004). Particular advantageous method is LC-MS, as it allows simultaneous qualitative and quantitative analysis of the six major milk proteins including their isoforms and genetic variants. The LC method allows rapid and automated analysis with high resolutions, accuracy and reproducibility, and combined with MS allows identification of proteins based on their molecular masses even when proteins co-elute in LC.

1.2 Variation in milk protein composition

Detailed milk protein composition exhibits high heterogeneity because of quantitative variation in the content of different milk proteins, numerous genetic variants, and isoforms with different degrees of post-translational modifications such as glycosylation of κ-CN and phosphorylation of all caseins (Caroli et al., 2004; Holland and Boland, 2014). The next sections will provide a detailed overview regarding natural variation in milk protein composition.

Genetic variants. Milk proteins are products of milk protein genes. The genes encoding caseins are clustered within a 250 kb region on chromosome 6 with the following order: α_{s1}-, β-, α_{s2}-, and κ-CN (Threadgill and Womack, 1990). The gene encoding α-LA is located on chromosome 5 (Hayes et al., 1993), and the one encoding β-LG is located on chromosome 11 (Hayes and Petit, 1993). The DNA sequence determines the amino acid sequence of the protein and influences the amount of protein that is produced. Genetic differences can be due to point mutations such as single nucleotide polymorphisms (SNP) or due to DNA rearrangements such as structure variations, insertions or deletions. When changes occur in regulatory regions (e.g. enhancers and promoters), alteration may occur at the transcription level, and may result in different levels of expression and different amounts of proteins. When non-synonymous changes occur in coding regions, this will result in amino acid substitutions. One or more differences in amino acid sequence e.g. due to mutations, deletions or insertions in the DNA sequence will result in different genetic (protein) variants, which may possess different physical or chemical properties. Genetic variants of the six major milk proteins detected so far in bovine are shown in Table 1.1. In most western dairy breeds, α_{s2}-CN is monomorphic for the A variant, and α-LA is monomorphic for the B variant. The important functional role of α-LA in lactose biosynthesis suggests that less genetic variation of this protein is allowed. Furthermore, α_{s1}-CN B, β-CN A1 and A2, κ-CN A and B, and β-LG A and B are the most common genetic variants.

Table 1.1. Detected genetic variants for the six major bovine milk proteins

Protein	Genetic variants
α_{s1}-CN	A, B, C, D, E, F, G, H, I
α_{s2}-CN	A, B, C, D
β-CN	A1, A2, A3, B, C, D, E, F, G, H, I, J
κ-CN	A, B, B2, C, D, E, F1, F2, G1, G2, H, I, J, K, L
α-LA	A, B, C, D
β-LG	A, B, B*, C, D, Dr, E, F, G, H, I, J, W

(Visker et al., 2012; Martin et al., 2013)

Genetic variants may affect cheese-making properties of milk (Bittante et al., 2012), either through changed protein structure (e.g. α_{s1}-CN A) or through changed casein content (e.g. β-LG B) or composition (e.g. κ-CN B). The α_{s1}-CN A variant is a variant with a deletion of 13 amino acid residues compared to the α_{s1}-CN B variant. This deletion reduces the hydrophobicity of this protein and is responsible for the loss of the primary cleavage site (F^{37}-F^{38}, see Figure1.2) by chymosin. Milk containing α_{s1}-CN A variant exhibits longer clotting time than milk containing α_{s1}-CN B variant (Coker et al., 1999). The β-LG B variant is associated with higher casein content (Bobe et al., 1999; Heck et al., 2009). Milk containing β-LG B variant results in higher cheese yield (Van Den Berg et al., 1992; Wedholm et al., 2006; Hallén et al., 2008). The κ-CN B variant is associated with a greater proportion of κ-CN including glycosylated κ-CN and smaller casein micelle size (Bijl et al., 2014a; Bonfatti et al., 2014). Milk containing the κ-CN B variant exhibits shorter renneting time compared with milk containing the κ-CN A variant (Ng-Kwai-Hang, 2006; Jensen et al., 2015).

Post-translational modifications. Unlike two main whey proteins, caseins are present in diverse isoforms because of different degrees of post-translational modifications such as glycosylation of κ-CN, phosphorylation of all caseins, and disulphide bridging of α_{s2}-CN and κ-CN. These post-translational modifications are crucial for micelle formation and stability as shown in the schematic representation of a casein micelle in Figure 1.1A. Although casein micelle structure is not fully elucidated, it is now well established that most of κ-CN is present on the surfaces of the micelles (Dalgleish et al., 1989). The highly hydrophilic part of κ-CN (some are glycosylated), known as the hairy layer, causes electrostatic and steric repulsion between micelles to avoid aggregation of caseins (De Kruif and Zhulina, 1996). The micellar interior consists of a matrix of caseins and calcium phosphate nanoclusters (De Kruif and Holt, 2003). Besides glycosylation and phosphorylation, another post-translational modification is the disulphide bond formed between two cysteine residues in α_{s2}-CN or κ-CN. As a result, α_{s2}-CN is present as monomers or dimers, and κ-CN is present as monomers or oligomers (Rasmussen et al., 1994).

Phosphorylation. Phosphorylation of caseins occurs after the synthesis of polypeptide chains in the Golgi apparatus of the mammary epithelial cell and is catalyzed by protein kinases (Bingham et al., 1972). These kinases phosphorylate serine (Ser/S) or threonine (Thr/T) by recognizing the tripeptide sequence as Ser or Thr-X-glutamic acid (Glu/E) or phosphorylated Ser (SerP) or Aspartic acid (Asp/D), where X represents any amino acid residue (Mercier, 1981). This post-translational modification allows caseins to interact with calcium phosphate in the assembly of casein micelles. Phosphoserine clusters of α_{s1}-, α_{s2}- and β-CN are involved in stabilizing calcium phosphate nanoclusters and enable micellar growth by crosslinking between phosphorylated residues of caseins and calcium phosphate as shown in Figure 1.1B (De Kruif and Holt, 2003; Dalgleish and Corredig, 2012; De Kruif et al., 2012). So far, the molecular mechanism and factors controlling casein phosphorylation are still largely unknown.

Although caseins are all phosphoproteins, they differ greatly in their degree of phosphorylation. The known phosphorylation sites of the different caseins are shown in Figure 1.2. The α_{s1}-CN possesses 2 phosphoserine clusters and has 2 common phosphorylation isoforms: α_{s1}-CN-8P as the predominant form and α_{s1}-CN-9P as the minor form (Farrell, Jr. et al., 2004). The extra phosphorylated serine of α_{s1}-CN-9P is present on Ser^{56} in the Ser-X-Asp motif (Manson

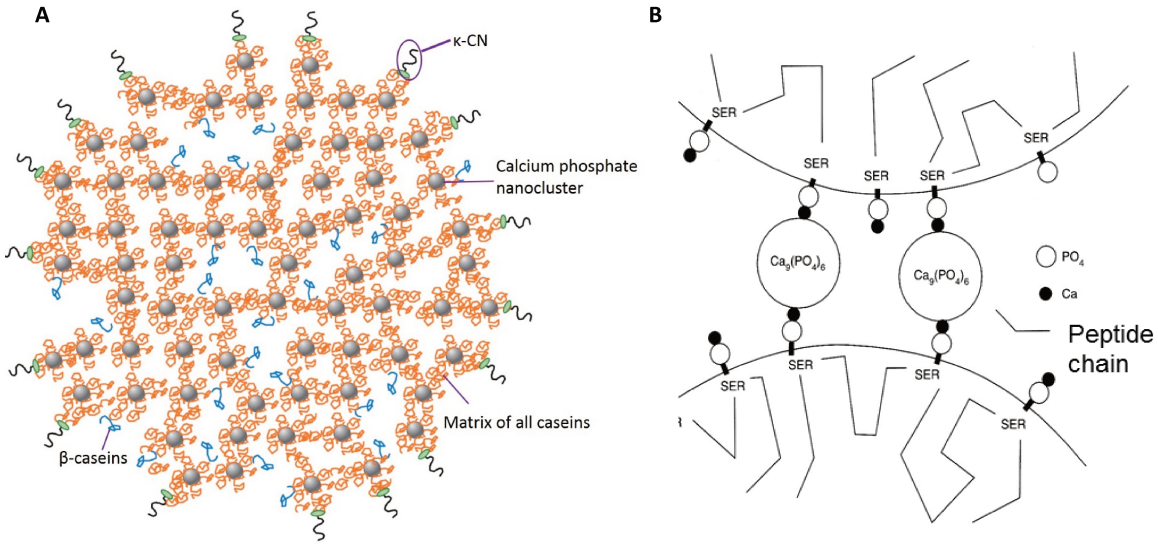

Figure 1.1. (A) Schematic representation of the casein micelle structure adopted from Dalgleish and Corredig (2012). The internal structure is the matrix of α_s- and β-caseins linked to the calcium phosphate nanoclusters. The κ-CN (the hairy layer) is on the outermost parts of the surface. (B) Interaction between calcium phosphate nanoclusters and phosphopeptides of caseins (Schmidt, 1982).

et al., 1977). The α_{s2}-CN possesses 3 phosphoserine clusters and is present with isoforms from 10P to 13P (Farrell, Jr. et al., 2004), with α_{s2}-CN-11P as the major form. Based on the amino acid sequence of α_{s2}-CN, 12 Ser and 6 Thr residues can be phosphorylated in theory, but not all the phosphorylation sites have been confirmed by experiments. Up to now, only 11 Ser have been confirmed as phosphorylation sites of α_{s2}-CN (Imanishi et al., 2007; Baum et al., 2013). The β-CN possesses only 1 phosphoserine cluster and is usually present with 5P and occasionally with 4P. The κ-CN does not possess any phosphoserine cluster and is present with 1P to 2P and occasionally with 3P (Farrell, Jr. et al., 2004; Holland and Boland, 2014). The 2 Ser residues in κ-CN are phosphorylated first in Ser^{170} followed by Ser^{148} (Holland and Boland, 2014). In κ-CN-B-3P, the additional phosphorylated residue is on Thr^{166} (Holland et al., 2006).

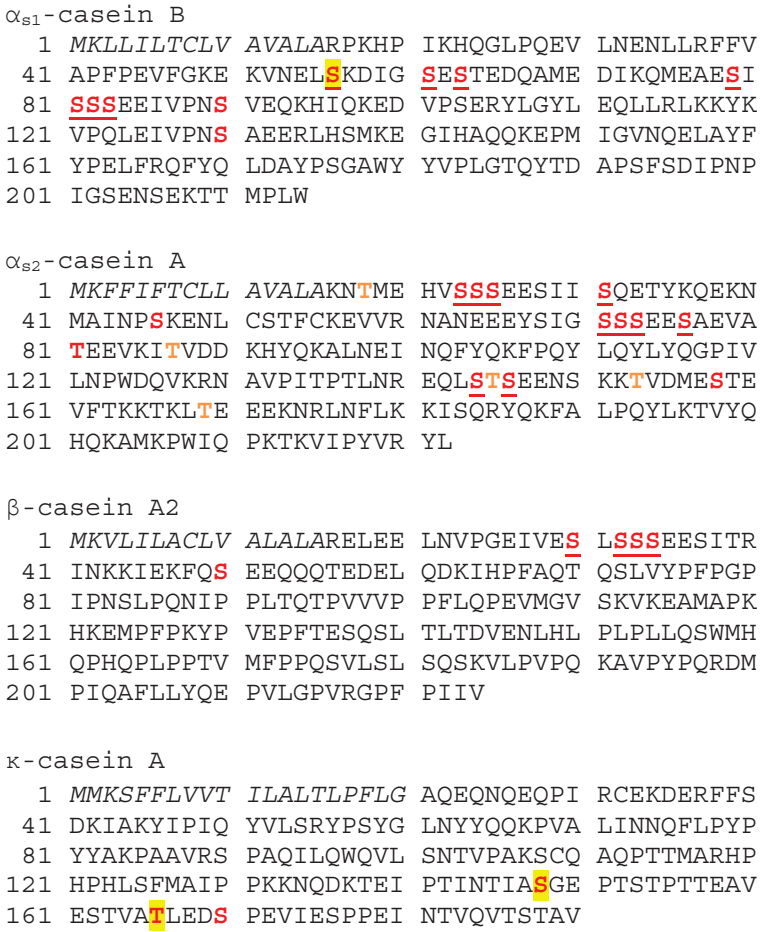

Figure 1.2. Amino acid sequences of α_{s1}-CN B, α_{s2}-CN A, β-CN A2 and κ-CN A variants. The signal peptides are in italics. The known phosphorylated residues are marked bold and red (Holland and Boland, 2014). Phosphoserine clusters are marked underlined (De Kruif and Holt, 2003). The extra phosphorylated residue in α_{s1}-CN-9P, κ-CN-2P and κ-CN-3P are marked bold and highlighted in yellow (Manson et al., 1977; Holland et al., 2006). For α_{s2}-CN, not all the phosphorylation sites are confirmed by experiments. The potential (theoretical) phosphorylation sites are marked in orange (Mercier, 1981).

1.3 Genomic regions involved in milk protein composition

Detailed milk protein composition contributes largely to the nutritional value and technological properties of milk. Therefore, detailed milk protein composition tailored according to desired products is of great interest to the dairy industry, especially for cheese production (Wedholm

et al., 2006; Hallén et al., 2008; Caroli et al., 2009). Relative concentrations of individual milk proteins and their isoforms vary considerably among milk of individual cows (Heck et al., 2008; Bonfatti et al., 2011; Bonfatti et al., 2014; Bijl et al., 2014b; Boichard et al., 2014; Gebreyesus et al., 2016). Factors contributing to the variation in milk protein composition include breed, feed, seasonal changes, stage of lactation and health status of individual cows (Ng-Kwai-Hang et al., 1987; Groen et al., 1994; Coulon et al., 1998). Additionally, several studies have shown that genetic factors contribute a major part of the variation in milk protein composition; heritability estimates, as the proportion of phenotypic variation explained by genetic variation, for the six major milk proteins range from 0.25 to 0.79 across different breeds (Schopen et al., 2009; Bonfatti et al., 2011; Boichard et al., 2014). This implies that selecting cows that produce milk with the desired composition is one possible way to modify detailed milk protein composition. While genetic variation for the six major milk proteins has been reported, information about genetic variation in phosphorylation of caseins is scarce. To date, only two studies have reported heritability estimates for relative concentrations of α_{s1}-CN-8P and α_{s1}-CN-9P in Dutch Holstein Friesian (Bijl et al., 2014b), Danish Holstein and Danish Jersey (Buitenhuis et al., 2016). Heritability estimates for α_{s2}-CN phosphorylation isoforms have not been reported. The limited number of studies reflects the technical difficulties for their quantification. Given the substantial genetic variation in bovine milk protein composition, identifying genomic regions responsible for this variation will increase our understanding of the biological mechanisms involved in milk protein synthesis. To identify such regions, known as quantitative trait loci (QTL), the whole genome is screened using SNP with known location. Up to now, there are only three genome-wide association studies (GWAS) for milk protein composition measured with different analytical methods (Schopen et al., 2011; Buitenhuis et al., 2016; Sanchez et al., 2016). The main identified regions for milk protein composition are on chromosomes 6 where casein gene cluster is located, 11 where the gene encoding β-LG is located, 14 where *DGAT1* gene is located.

The formation and stability of casein micelles are highly relevant to manufacturing of dairy products. Therefore, exploring the variation in phosphorylation degree of α_s-CN is of great interest because caseins may play different roles in stabilizing the internal micellar structure (Hoagland et al., 2001; De Kruif and Holt, 2003), and affect cheese-making properties of milk (Frederiksen et al., 2011; Jensen et al., 2012). Furthermore, finding genomic regions associated with different phosphorylation isoforms of caseins, in particular α_{s1}- and α_{s2}-CN, might help to understand the mechanism of phosphorylation of caseins. Up to now, only a single study has reported QTL associated with α_{s1}-CN-8P and α_{s1}-CN-9P (Bijl et al., 2014b), and there is no information regarding QTL associated with individual α_{s2}-CN phosphorylation isoforms. Interestingly, Bijl et al. (2014b) showed that one QTL on BTA 11 is associated with α_{s1}-CN-8P concentration, and one QTL on chromosome 14 is associated with α_{s1}-CN-9P concentration. Consequently, they hypothesize that different sets of genes regulate the phosphorylation of α_{s1}-CN-8P and α_{s1}-CN-9P. Furthermore, Heck et al. (2008) reported strong phenotypic correlations of α_{s1}-CN-9P with α_{s2}-CN-10P and -11P, suggesting that phosphorylation of α_{s1}- and α_{s2}-CN might be related. However, the relations between α_{s2}-CN phosphorylation isoforms as well as the relations between α_{s1}- and α_{s2}-CN isoforms have not been reported.

1.4 Aim and outline of this thesis

This thesis aimed to unravel the genetic background of bovine milk protein composition with special emphasis on the phosphorylation of caseins. In **Chapter 2**, we quantified the variation in α_{s1}- and α_{s2}-CN phosphorylation isoforms among milk of French Montbéliarde cows and quantified the relations among these isoforms. These results were further explored in chapter 3. In **Chapter 3**, we investigated to what extent genetic and other factors contribute to the variation in relative concentrations of α_{s1}- and α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN defined as the proportion of isoforms with higher degrees of phosphorylation in total α_{s1}- and total α_{s2}-CN, respectively. We also investigated associations of genetic variants of milk proteins and casein haplotypes with relative concentrations of α_{s1}- and α_{s2}-CN phosphorylation isoforms and with the phosphorylation degrees of α_{s1}- and α_{s2}-CN. In **Chapter 4**, we estimated genetic parameters for individual α_{s1}- and α_{s2}-CN phosphorylation isoforms in Dutch Holstein Friesian and the genetic correlations among these isoforms. In **Chapter 5**, we performed a genome-wide association study in Dutch Holstein Friesian for α_{s1}- and α_{s2}-CN phosphorylation isoforms to identify genomic regions involved in phosphorylation of caseins. In **Chapter 6**, the general discussion, the possible reasons for discrepancies across studies regarding milk protein composition were discussed, and the results obtained from this thesis were used to speculate on the phosphorylation mechanism of the caseins.

References

Baum, F., J. Ebner, and M. Pischetsrieder (2013). Identification of multiphosphorylated peptides in milk. *J. Agric. Food Chem.* 61.38, pp. 9110–9117.

Bijl, E., R. de Vries, H. van Valenberg, T. Huppertz, and T. van Hooijdonk (2014a). Factors influencing casein micelle size in milk of individual cows: Genetic variants and glycosylation of κ-casein. *Int. Dairy J.* 34.1, pp. 135–141.

Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk, and H. Bovenhuis (2014b). Phosphorylation of αS1-casein is regulated by different genes. *J. Dairy Sci.* 97.11, pp. 7240–7246.

Bingham, E. W., H. M. Farrell, Jr., and J. J. Basch (1972). Phosphorylation of Casein: ROLE OF THE GOLGI APPARATUS. *J. Biol. Chem.* 247, pp. 8193–8194.

Bittante, G., M. Penasa, and A. Cecchinato (2012). Invited review: Genetics and modeling of milk coagulation properties. *J. Dairy Sci.* 95.12, pp. 6843–70.

Bobe, G., D. Beitz, A. Freeman, and G. Lindberg (1998). Separation and quantification of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J. Agric. Food Chem.* 46.2, pp. 458–463.

Bobe, G., D. Beitz, A. Freeman, and G. Lindberg (1999). Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. *J. Dairy Sci.* 82.12, pp. 2797–2804.

Boichard, D., A. Govignon-Gion, H. Larroque, C. Maroteau, I. Palhière, G. Tosser-Klopp, R. Rupp, M.-P. Sanchez, and M. Brochard (2014). Déterminisme génétique de la composition en acides gras et protéines du lait des ruminants, et potentialités de sélection. *INRA Prod. Anim.* 27.4, pp. 283–298.

Bonfatti, V., A. Cecchinato, L. Gallo, A. Blasco, and P. Carnier (2011). Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. *J. Dairy Sci.* 94.10, pp. 5183–5193.

Bonfatti, V., G. Chiarot, and P. Carnier (2014). Glycosylation of κ-casein: Genetic and nongenetic variation and effects on rennet coagulation properties of milk. *J. Dairy Sci.* Pp. 1–9.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier (2008). Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. *J. Chromatogr. A* 1195.1-2, pp. 101–106.

Bordin, G., F. Cordeiro Raposo, B. de La Calle, and A. R. Rodriguez (2001). Identification and quantification of major bovine milk proteins by liquid chromatography. *J. Chromatogr. A* 928.1, pp. 63–76.

Bramanti, E., C. Sortino, and G. Raspi (2002). New chromatographic method for separation and determination of denatured αs1-, αs2-, β-and κ-caseins by hydrophobic interaction chromatography. *J. Chromatogr. A* 958.1, pp. 157–166.

Brodbeck, U., W. L. Denton, N. Tanahashi, and K. E. Ebner (1967). The isolation and identification of the B protein of lactose synthetase as α-lactalbumin. *J. Biol. Chem.* 242.7, pp. 1391–1397.

Brodbeck, U. and K. E. Ebner (1966). Resolution of a soluble lactose synthetase into two protein components and solubilization of microsomal lactose synthetase. *J. Biol. Chem.* 241.3, pp. 762–764.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen (2016). Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genet.* 17.

Caroli, A. M., S. Chessa, and G. J. Erhardt (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92.11, pp. 5335–5352.

Caroli, A., S. Chessa, P. Bolla, E. Budelli, and G. C. Gandini (2004). Genetic structure of milk protein polymorphisms and effects on milk production traits in a local dairy cattle. *J. Anim. Breed. Genet.* 121.2, pp. 119–127.

Coker, C., L. Creamer, R. Burr, and J. Hill (1999). The hydrolysis of the α s1-casein A, B and C variants by plasmin and chymosin. *Int. Dairy J.* 9, pp. 371–372.

Coulon, J. B., C. Hurtaud, B. Remond, and R. Verite (1998). Factors contributing to variation in the proportion of casein in cows' milk true protein: a review of recent INRA experiments. *J. Dairy Res.* 65.3, pp. 375–387.

Dalgleish, D., D. Horne, and A. Law (1989). Size-related differences in bovine casein micelles. *Biochim. Biophys. Acta - Gen. Subj.* 991.3, pp. 383–387.

Dalgleish, D. G. and M. Corredig (2012). The structure of the casein micelle of milk and its changes during processing. *Annu. Rev. Food Sci. Technol.* 3.1, pp. 449–467.

De Kruif, C. G. and E. B. Zhulina (1996). κ-casein as a polyelectrolyte brush on the surface of casein micelles. *Colloids Surfaces A Physicochem. Eng. Asp.* 117.1, pp. 151–159.

De Kruif, C. and C. Holt (2003). "Casein micelle structure, functions and interactions". In: *Adv. Dairy Chem. Proteins*. Ed. by P. F. Fox and P. L. H. McSweeney. 3rd. Vol. 1. JANUARY 2003. Springer US. Chap. 5, pp. 233–276.

De Kruif, C. G., T. Huppertz, V. S. Urban, and A. V. Petukhov (2012). Casein micelles and their internal structure. *Adv. Colloid Interface Sci.* 171-172, pp. 36–52.

De Jong, N., S. Visser, and C. Olieman (1993). Determination of milk proteins by capillary electrophoresis. *J. Chromatogr. A* 652.1, pp. 207–13.

Farrell, Jr., H. M., G. T. Bleck, E. M. Brown, J. E. Butler, L. K. Creamer, C. L. Hicks, C. M. Hollar, K. F. Ng-Kwai-Hang, and H. E. Swaisgood (2004). Nomenclature of the proteins of cows' milk–sixth revision. *J. Dairy Sci.* 87.6, pp. 1641–74.

Frederiksen, P. D., K. K. Andersen, M. Hammershøj, H. D. Poulsen, J. Sørensen, M. Bakman, K. B. Qvist, and L. B. Larsen (2011). Composition and effect of blending of noncoagulating, poorly coagulating, and well-coagulating bovine milk from individual Danish Holstein cows. *J. Dairy Sci.* 94.10, pp. 4787–4799.

Gebreyesus, G., M. S. Lund, L. Janss, N. A. Poulsen, L. B. Larsen, H. Bovenhuis, and A. J. Buitenhuis (2016). Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the Danish Holstein. *J. Dairy Sci.* 99.4, pp. 2863–2866.

Glantz, M., T. Devold, G. Vegarud, H. Lindmark Månsson, H. Stålhammar, and M. Paulsson (2010). Importance of casein micelle size and milk composition for milk gelation. *J. Dairy Sci.* 93.4, pp. 1444–1451.

Groen, A. F., R. Van der Vegt, M. Van Boekel, O. De Rouw, and H. Vos (1994). Case study on individual animal variation in milk protein composition as estimated by high-pressure liquid chromatography. *Ned. melk en Zuiveltijdschr.* 48.4, pp. 201–212.

Gustavsson, F., M. Glantz, A. J. Buitenhuis, H. Lindmark-Månsson, H. Stålhammar, A. Andrén, and M. Paulsson (2014). Factors influencing chymosin-induced gelation of milk from individual dairy cows: Major effects of casein micelle size and calcium. *Int. Dairy J.* 39.1, pp. 201–208.

Hallén, E., A Wedholm, A Andrén, A Lundén, A. Wedholm, A. Andrén, and A. Lundén (2008). Effect of beta-casein, kappa-casein and beta-lactoglobulin genotypes on concentration of milk protein variants. *J. Anim. Breed. Genet.* 125.2, pp. 119–129.

Hayes, H. C. and E. J. Petit (1993). Mapping of the beta-lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 4.4, pp. 207–210.

Hayes, H. C., P. Popescu, and B. Dutrillaux (1993). Comparative gene mapping of lactoperoxidase, retinoblastoma, and alpha-lactalbumin genes in cattle, sheep, and goats. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 4.10, pp. 593–597.

Heck, J. M. L., A. Schennink, H. J. F. Van Valenberg, H. Bovenhuis, M. H. P. W. Visker, J. A. M. Van Arendonk, and A. C. M. van Hooijdonk (2009). Effects of milk protein variants on the protein composition of bovine milk. *J. Dairy Sci.* 92.3, pp. 1192–1202.

Heck, J., C. Olieman, A. Schennink, H. van Valenberg, M. Visker, R. Meuldijk, and A. van Hooijdonk (2008). Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *Int. Dairy J.* 18.5, pp. 548–555.

Hoagland, P. D., J. J. Unruh, E. D. Wickham, and H. M. Farrell (2001). Secondary structure of bovine αS2-casein: theoretical and experimental approaches. *J. Dairy Sci.* 84.9, pp. 1944–1949.

Holland, J. W. and M. J. Boland (2014). "Post-translational modifications of caseins". In: *Milk proteins From Expr. to food*. Ed. by H. Singh, A. Thompson, and M. Boland. Second Edi. Elsevier Inc., pp. 141–168.

Holland, J. W., H. C. Deeth, and P. F. Alewood (2006). Resolution and characterisation of multiple isoforms of bovine κ-casein by 2-DE following a reversible cysteine-tagging enrichment strategy. *Proteomics* 6.10, pp. 3087–3095.

Hollar, C. M., A. J. R. Law, D. G. Dalgleish, and R. J. Brown (1991). Separation of Major Casein Fractions Using Cation-Exchange Fast Protein Liquid Chromatography1. *J. Dairy Sci.* 74.8, pp. 2403–2409.

Holt, C., J. A. Carver, H. Ecroyd, and D. C. Thorn (2013). Invited review: Caseins and the casein micelle: Their biological functions, structures, and behavior in foods. *Joural Dairy Sci.* 96.10, pp. 6127–6146.

Imanishi, S. Y., V. Kochin, S. E. Ferraris, A. de Thonel, H.-M. Pallari, G. L. Corthals, and J. E. Eriksson (2007). Reference-facilitated phosphoproteomics: fast and reliable phosphopeptide validation by microLC-ESI-Q-TOF MS/MS. *Mol. Cell. Proteomics* 6, pp. 1380–1391.

Jenness, R., B. Larson, T. McMeekin, A. Swanson, C. Whitnah, and R. Whitney (1956). Nomenclature of the Proteins of Bovine Milk. *J. Dairy Sci.* 39.5, pp. 536–541.

Jensen, H. B., N. A. Poulsen, K. K. Andersen, M. Hammershøj, H. D. Poulsen, and L. B. Larsen (2012). Distinct composition of bovine milk from Jersey and Holstein-Friesian cows with good, poor, or noncoagulation properties as reflected in protein genetic variants and isoforms. *J. Dairy Sci.* 95.12, pp. 6905–6917.

Jensen, H. B., K. S. Pedersen, L. B. Johansen, N. A. Poulsen, M. Bakman, D. E. Chatterton, and L. B. Larsen (2015). Genetic variation and posttranslational modification of bovine κ-casein: Effects on caseino-macropeptide release during renneting. *J. Dairy Sci.* 98.2, pp. 747–758.

Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes (1987). Variation in milk protein concentrations associated with genetic polymorphism and environmental factors. *J. Dairy Sci.* 70.3, pp. 563–570.

Ng-Kwai-Hang, K. (2006). Genetic variants of milk proteins and their effects on the yield and quality of cheese. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 1.056, pp. 1–11.

Ng-Kwai-Hang, K. and E. Kroeker (1984). Rapid Separation and Quantification of Major Caseins and Whey Proteins of Bovine Milk by Polyacrylamide Gel Electrophoresis. *J. Dairy Sci.* 67.12, pp. 3052–3056.

Léonil, J., D. Mollé, F. Gaucheron, P. Arpino, P. Guénot, and J. L. Maubois (1995). Analysis of major bovine milk proteins by on-line high-performance liquid chromatography and electrospray ionization-mass spectrometry. *Lait* 75.3, pp. 193–210.

Logan, A., A. Leis, L. Day, S. K. Øiseth, A. Puvanenthiran, and M. A. Augustin (2015). Rennet gelation properties of milk: Influence of natural variation in milk fat globule size and casein micelle size. *Int. Dairy J.* 46, pp. 71–77.

Manson, W., T. Carolan, and W. D. Annan (1977). Bovine αS0 casein; a phosphorylated homologue of αS1 casein. *Eur. J. Biochem.* 417, pp. 411–417.

Martin, P., L. Bianchi, C. Cebo, and G. Miranda (2013). "Genetic polymorphism of milk protein". In: *Adv. Dairy Chem. Proteins Basic Asp.* Ed. by P. Fox and P. McSweeney. 4th ed. New York: Springer. Chap. 15, pp. 463–514.

Mercier, J. C. (1981). Phosphorylation of caseins, present evidence for an amino acid triplet code posttranslationally recognized by specific kinases. *Biochimie* 63.1, pp. 1–17.

O'Donnell, R., J. Holland, H. Deeth, and P. Alewood (2004). Milk proteomics. *Int. Dairy J.* 14.12, pp. 1013–1023.

Rasmussen, L. K., P. Højrup, and T. E. Petersen (1994). Disulphide arrangement in bovine caseins: localization of intrachain disulphide bridges in monomers of κ-and αs2-casein from bovine milk. *J. Dairy Res.* 61.4, pp. 485–493.

Recio, I., M. L. Pérez-Rodríguez, M. Ramos, and L. Amigo (1997). Capillary electrophoretic analysis of genetic variants of milk proteins from different species. *J. Chromatogr. A* 768.1, pp. 47–56.

Sanchez, M., A. Govignon-Gion, M. Ferrand, M. Gelé, D. Pourchet, Y. Amigues, S. Fritz, M. Boussaha, A. Capitan, D. Rocha, G. Miranda, P. Martin, M. Brochard, and D. Boichard (2016). Whole-genome scan to detect quantitative trait loci associated with milk protein composition in 3 French dairy cattle breeds. *J. Dairy Sci.* 99.10, pp. 8203–8215.

Schmidt, D. G. (1982). Association of caseins and casein micelle structure. *Dev. dairy Chem.* 1, pp. 61–86.

Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. van Valenberg, and J. A. M. van Arendonk (2009). Genetic parameters for major milk proteins in Dutch Holstein-Friesians. *J. Dairy Sci.* 92.3, pp. 1182–1191.

Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and H. Bovenhuis (2011). Whole-genome association study for milk protein composition in dairy cattle. *J. Dairy Sci.* 94.6, pp. 3148–3158.

Seibert, B., G. Erhardt, and B. Senft (1985). Procedure for simultaneous phenotyping of genetic variants in cow's milk by isoelectric focusing. *Anim. Blood Groups Biochem. Genet.* 16.3, pp. 183–191.

Threadgill, D. W. and J. E. Womack (1990). Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Res.* 18.23, pp. 6935–6942.

Van Den Berg, G., J. T. M. Escher, P. J. De Koning, and H. Bovenhuis (1992). Genetic polymorphism of κ-casein and β-lactoglobulin in relation to milk composition and processing properties. eng. *Ned. melk en Zuiveltijdschr.* 46.3-4, pp. 145–168.

Visker, M. H. P. W., J. M. L. Heck, H. J. F. van Valenberg, J. A. M. van Arendonk, and H. Bovenhuis (2012). Short communication: A new bovine milk-protein variant: α-lactalbumin variant D. *J. Dairy Sci.* 95.4, pp. 2165–9.

Visser, S., C. J. Slangen, and H. S. Rollema (1991). Phenotyping of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J. Chromatogr. A* 548.

Wedholm, A., L. B. Larsen, H. Lindmark-Månsson, A. H. Karlsson, and A. Andrén (2006). Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *J. Dairy Sci.* 89.9, pp. 3296–305.

Chapter 2

The relationships among bovine α_s-casein phosphorylation isoforms suggest different phosphorylation pathways

Z.H. Fang[*†], M.H.P.W. Visker[†], G. Miranda[*], A. Delacroix-Buchet[*], H. Bovenhuis[†] and P. Martin[*]

[*]Génétique Animale et Biologie Intégrative, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[†]Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

Journal of Dairy Science (2016) 99:8168-8177

Abstract

Casein (CN) phosphorylation is an important post-translational modification, and it is one of the key factors responsible for constructing and stabilizing casein micelles. Variation in phosphorylation degree of α_s-CN is of great interest because it is suggested to affect milk technological properties. This study aimed to investigate the variation in phosphorylation degree of α_s-CN among milk of individual cows and to explore relationships among different phosphorylation isoforms of α_s-CN. For this purpose, we analyzed morning milk samples from 529 French Montbéliarde cows using Liquid Chromatography coupled with Electrospray Ionization Mass Spectrometry (LC/ESI-MS). We detected three new phosphorylation isoforms: α_{s2}-CN-9P, -14P, and -15P in bovine milk, in addition to the known isoforms α_{s1}-CN-8P and -9P, and α_{s2}-CN-10P, -11P, -12P, and -13P. The relative concentrations of each α_s-CN phosphorylation isoform varied considerably among individual cows. Furthermore, the phenotypic correlations and hierarchical clustering suggest at least two regulatory systems for phosphorylation of α_s-CN: one responsible for isoforms with lower levels of phosphorylation (α_{s1}-CN-8P, α_{s2}-CN-10P and -11P), and another responsible for isoforms with higher levels of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-12P, -13P and -14P). Identifying all phosphorylation sites of α_{s2}-CN and investigating the genetic background of different α_{s2}-CN phosphorylation isoforms may provide further insight into phosphorylation mechanism of caseins.

Key words: phosphorylation, casein, milk protein composition, LC/ESI-MS

2.1 Introduction

Detailed milk protein composition contributes largely to the nutritional value and technological properties of milk. Casein accounts for about 80% of total protein in cow milk, and its content and composition largely influence cheese manufacturing properties of milk (Hallén et al., 2008; Wedholm, 2008; Caroli et al., 2009). Casein arises from the expression of 4 genes encoding 4 distinct polypeptide chains: α_{s1}-CN, α_{s2}-CN, β-CN, and κ-CN, of which the relative amounts in milk are approximately at the ratio 4:1:4:1. Phosphorylation of caseins is an important post-translational modification occurring after the synthesis of the polypeptide chains in the Golgi apparatus of the mammary epithelial cell under the action of protein kinases (Bingham et al., 1972). These kinases phosphorylate Ser or Thr by recognizing the tripeptide sequence Ser/Thr-X-Glu/SerP/Asp, where X represents any AA residue and P indicates phosphorylation (Mercier, 1981). This posttranslational modification allows caseins to interact with calcium phosphate to form large colloidal structures called casein micelles. Phosphoserine clusters of α_{s1}-CN, α_{s2}-CN, and β-CN are involved in stabilizing calcium phosphate nanoclusters and enable micellar growth by crosslinking between phosphorylated residues of caseins and calcium phosphate (De Kruif and Holt, 2003; De Kruif et al., 2012; Dalgleish and Corredig, 2012).

Although α_{s1}-CN, α_{s2}-CN, β-CN, and κ-CN are all phosphoproteins, they vary strikingly in their degree of phosphorylation. In bovine milk, α_{s1}-CN has 2 common phosphorylation isoforms: α_{s1}-CN-8P and α_{s1}-CN-9P; α_{s2}-CN is present with isoforms from 10P to 13P, β-CN is usually present with 5P and occasionally with 4P, and κ-CN is present with 1P to 2P and occasionally with 3P (Farrell, Jr. et al., 2004; Holland and Boland, 2014). Exploring the variation in phosphorylation degree of α_s-CN is of great interest because caseins may play different roles in stabilizing the internal micellar structure (Hoagland et al., 2001; De Kruif and Holt, 2003), and affect cheese-making properties of milk (Frederiksen et al., 2011). Phosphorylation degree of α_{s2}-CN was found negatively correlated with phosphorylation degree of α_{s1}-CN (Heck et al., 2008). Additionally, α_{s1}-CN-8P and α_{s1}-CN-9P are suggested to be regulated by different sets of genes (Bijl et al., 2014).

This study aimed to investigate the variation in phosphorylation degree of α_s-CN among milk of individual cows and to explore the relationships among different phosphorylation isoforms of α_s-CN in Montbéliarde cows in the Franche-Comté region in France.

2.2 Materials and Methods

2.2.1 Milk samples

We sampled 576 Montbéliarde cows located on 430 herds across 3 departments (Doubs, Jura, and Haute-Saône) to obtain a good representation of the variation in milk protein composition from the current Montbéliarde cattle population in the Franche-Comté region in the east of France. We sampled cows across different parities (1-5) and lactation stages (7-652 d) based on paternal pedigree (to maximize genetic diversity) and on protein and calcium content in milk (to increase milk content diversity). During autumn/winter (October-December, 2014)

and spring/summer (April-July, 2015), fresh morning milk samples (25 mL) were preserved with Bronopol after collection, transported on ice to the laboratory, and then frozen at -20°C until analysis by liquid chromatography/electrospray ionization-mass spectrometry **(LC/ESI-MS)**.

2.2.2 Milk protein profiling

All chemicals used in the LC/ESI-MS analysis were of the highest purity commercially available and were used without further purification. Trifluoroacetic acid **(TFA)**, urea, Bis-Tris, dithiothreitol, and acetonitrile were purchased from Sigma-Aldrich (St. Louis, MO). Ultrapure water (Milli-Q Plus System, >18.3 MΩcm) was produced in the laboratory.

Prior to LC/ESI-MS analysis, milk samples were skimmed by centrifugation at 2,500 × g for 20 min at 4°C. Skim milk samples (20 μL) were then clarified by adding 180 μL of 0.1 M Bis-Tris buffer pH 8.0, containing 8 M urea, 1.3% trisodium citrate, and 0.3% dithiothreitol (Visser et al., 1991). Next, 15 μL of clarified milk samples were injected into a Discovery BIOWide Pore (Supelco) C5 column (150 × 2.10 mm, 300Å). Reversed-phase HPLC was carried out with an Ultimate LC 3000 system (Thermo Fisher Scientific, Waltham, MA). During the analysis, the autosampler was kept at 10°C, and the column was maintained at 42°C. The column's mobile phase consisted of a gradient mixture of solvent A (0.025% TFA in ultrapure water, vol/vol) and solvent B (0.02% TFA in acetonitrile, vol/vol) at a flow rate of 0.2 mL/min. The elution condition was a linear gradient from 29.5 to 34% B in 16 min, from 34 to 35.5% B in 0.1 min, from 35.5 to 37.5% B in 14.9 min, from 37.5 to 42% B in 14 min, from 42 to 95% B in 0.1 min, followed by an isocratic elution at 95% B for 5 min, and a linear return to 29.5% B in 0.1 min. The column was then re-equilibrated at 29.5% B as the starting condition for 10 min.

Protein elutes were detected by UV absorbance at 214 nm. The column was directly interfaced with an ESI-TOF mass spectrometer micrOTOF II focus (Bruker Daltonics, Wissembourg, France). The positive ion mode was used and mass scans were acquired over a range of 50 to 3,000 m/z. End plate offset voltage was set at -500 V and capillary voltage to 4,500 V. Nebulizer gas (N_2) pressure was maintained at 250 kPa and drying gas (N_2) flow was set at 4.0 L/min at 200°C. The LC/ESI-MS system was controlled by Hystar software v.2.3 (Bruker Daltonics). The charge number of multicharged ions, the deconvoluted mass spectra, and the determination of average molecular mass ($\boldsymbol{M}_r$) were obtained from Data Analysis v.3.4 software (Bruker Daltonics). A blank sample was injected after every milk sample to avoid carryover effects. A reference milk sample was analyzed after every 10 milk samples to determine reproducibility. In total, 58 aliquots of this reference sample were analyzed during the whole series of analyses.

2.2.3 Identification and quantification of milk protein fractions

We implemented the LC/ESI-MS method developed at INRA to simultaneously measure the relative concentrations of the major milk proteins and their isoforms, notably their phosphorylation isoforms (Miranda et al., 2013). Peak profiles from UV 214 nm were analyzed by Chromeleon software (Chromeleon 7.0.0, Dionex, Thermo Fisher Scientific). Protein variants

and isoforms of the 6 major milk proteins (α_{s1}-CN, α_{s2}-CN, β-CN, κ-CN, α-LA, and β-LG) were identified by matching measured molecular masses with an in-house calculated mass database on bovine milk proteins (Miranda et al., 2011).

A milk sample was considered degraded if peptides arising from the proteolysis of milk proteins, such as γ-casein and proteose-peptone component 5, accounted for more than 10% of total protein in milk. Degraded milk samples were discarded because degradation products co-eluted with some major protein fractions and interfered with the accuracy of quantification. As a result, 47 milk samples were discarded and 529 samples qualified for downstream quantification of the relative concentration of each protein fraction. Fractions of individual proteins and of α_{s1}-CN phosphorylation isoforms were estimated based on dividing the peak area of an individual protein by the total integrated peak area in the chromatogram of an individual milk sample. The LC could not completely separate the α_{s2}-CN phosphorylation isoforms in our study, so quantification based on peak area in the chromatogram would not be possible for individual α_{s2}-CN phosphorylation isoforms. Coupling mass spectrometry allows identification of a specific isoform in a mixture based on its molecular mass and further quantification via its signal intensity (Tuli and Ressom, 2009). Therefore, we used mass signal intensity to estimate the proportion of each α_{s2}-CN phosphorylation isoform as a fraction of total α_{s2}-CN, and the following equation was used to estimate the relative concentration of each α_{s2}-CN phosphorylation isoform in total milk protein:

$$\frac{\text{isoform mass signal intensity}}{\sum \text{isoform mass signal intensity}} \times \text{total } \alpha_{s2} - \text{CN (\%wt/wt)}.$$

The reproducibility of the method was assessed by calculating the coefficient of variation of the relative protein concentration for each protein and isoform fraction from the 58 aliquots of the reference milk sample.

2.2.4 Divisive hierarchical clustering analysis

A dendrogram was constructed using divisive hierarchical clustering analysis to graphically visualize phenotypic correlations among α_s-CN phosphorylation isoforms. The dissimilarity between the isoforms was calculated as $\sqrt{1-r}$, where r is the correlation coefficient between isoforms, to account for the direction and intensity of the relationships. The distance matrix was calculated using Euclidean distance. The analysis was performed using the R package Cluster (Maechler et al., 2015; R Core Team, 2015).

2.3 Results and Discussions

2.3.1 Milk protein profiling by LC/ESI-MS

Milk protein composition of 529 individual cows was determined by the LC/ESI-MS method. In this method, the 6 major milk proteins were separated by RP-HPLC, and their protein variants and posttranslational isoforms were identified by measured molecular masses (M_r) using ESI-

MS. The LC/ESI-MS reproducibility was based on the coefficient of variation of 58 aliquots of a reference milk sample analyzed during the whole series of sample analyses, which were below 5% for the 6 major milk proteins and below 8% for the α_{s1}-CN and α_{s2}-CN phosphorylation isoforms (Table 2.1). The LC/ESI-MS reproducibility for the 6 major milk proteins obtained in this study is within the range of reproducibilities reported for previous studies (Bobe et al., 1998; Bordin et al., 2001; Bonfatti et al., 2008; Heck et al., 2008).

Table 2.1. Mean, SD, CV, liquid chromatography coupled with electrospray ionization mass spectrometry (LC/ESI-MS) reproducibility, and frequencies of protein variants for the 6 major milk proteins and the different phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 529 Montbéliarde cows

Trait	Mean	SD	CV	LC-ESI/MS reproducibility[1]	Frequency of protein variant[2]
Milk protein composition (%wt/wt)					
α_{s1}-CN					
total	32.92	2.22	7	1	B(0.935); C(0.065)
8P	25.27	2.18	8	1	
9P	7.65	0.96	13	5	
α_{s2}-CN					
total	8.41	0.72	11	4	A(0.999); D(0.001)
9P[3]	0.24	0.05	21	NA	
10P	0.72	0.3	42	6	
11P	3.03	0.55	18	8	
12P	2.68	0.34	13	7	
13P	1.56	0.31	20	8	
14P	0.4	0.14	36	8	
15P[4]	0.2	0.1	49	NA	
β-CN	28.14	2.69	10	2	B(0.355);A1(0.087); A2(0.405); I(0.155)
κ-CN	9.03	0.86	9	2	A(0.537); B(0.459); E(0.004)
α-LA	3.54	0.7	20	5	B (1)
β- LG	12.16	1.87	15	4	A(0.444); B(0.554); D(0.002)
∑casein[5]	78.5	2.6	3		
∑whey[6]	15.71	1.83	12		
Milk production					
Milk yield (kg/day)	23.9	7.3	31		
Protein (%)	3.32	0.37	11		
Fat (%)	3.84	0.59	16		
Lactose (%)	4.93	0.31	6		

[1] Calculated based on the CV (%) of 58 repeated analyses of a reference milk sample.
[2] Frequencies of protein variants were calculated from 576 Montbéliarde cows.
[3] α_{s2}-CN-9P is 0.0 in 508 out of 529 samples; mean, SD, and CV are calculated based on 9 samples.
[4] α_{s2}-CN-15P is 0.0 in 525 out of 529 samples; mean, SD, and CV are calculated based on 4 samples.
[5] ∑casein = α_{s1}-CN + α_{s2}-CN + β-CN + κ-CN.
[6] ∑whey = α-LA + β-LG.

The 6 major milk proteins eluted in the following order: κ-CN, α_{s2}-CN, α_{s1}-CN, β-CN, α-LA, and β-LG (Figure 2.1A). κ-Casein eluted as 3 separate peaks: the first one contained glycosylated κ-CN from different protein variants, whereas the second one with a minor shoulder contained nonglycosylated κ-CN A-1P+2P (M_r = 19,037 Da for 1P; M_r = 19,117 Da for 2P) or κ-CN E-1P+2P (M_r = 19,007 Da for 1P; M_r = 19,087 Da for 2P), and the last peak with a minor shoulder contained nonglycosylated κ-CN B-1P+2P (M_r = 19,005 Da for 1P; M_r = 19,085 Da for

2P). Although κ-CN A-1P+2P and E-1P+2P had similar retention times, they could be discriminated based on their molecular masses. The α_{s2}-CN eluted into 2 major peaks as the retention time increased with its phosphorylation degree: one with α_{s2}-CN-11P (M_r = 25,228 Da) as the major component, together with α_{s2}-CN-10P (M_r = 25,148 Da) and occasionally with α_{s2}-CN-9P (M_r = 25,069 Da; Figure 2.1.B1), and the other with α_{s2}-CN-12P (M_r = 25,308 Da) as the major component and α_{s2}-CN-13P (M_r = 25,388 Da), α_{s2}-CN-14P (M_r = 25,468 Da), and sometimes α_{s2}-CN-15P (M_r = 25,548 Da; Figure 2.1.C2). To our knowledge, this is the first study to show the presence of α_{s2}-CN-9P, α_{s2}-CN-14P, and α_{s2}-CN-15P (Figure 2.1.B1 and 2.1.C2). The α_{s1}-CN separated into α_{s1}-CN-8P (M_r = 23,615 Da for the B variant; M_r = 23,543 Da for the C variant) and α_{s1}-CN-9P (M_r = 23,695 Da for the B variant; M_r = 23,623 Da for the C variant). The B and C variants had the same retention times, but they could be differentiated by their molecular masses. Four protein variants of β-CN eluted in the following order: β-CN B-5P (M_r = 24,092 Da), β-CN A1–5P (M_r = 24,023 Da), β-CN A2–5P (M_r = 23,983 Da), and β-CN I-5P (M_r = 23,965 Da). Two whey proteins without posttranslational modifications eluted after the caseins in the following order: α-LA (M_r = 14,186 Da), β-LG D (M_r = 18,280 Da), β-LG B (M_r = 18,281 Da), and β-LG A (M_r = 18,367 Da).

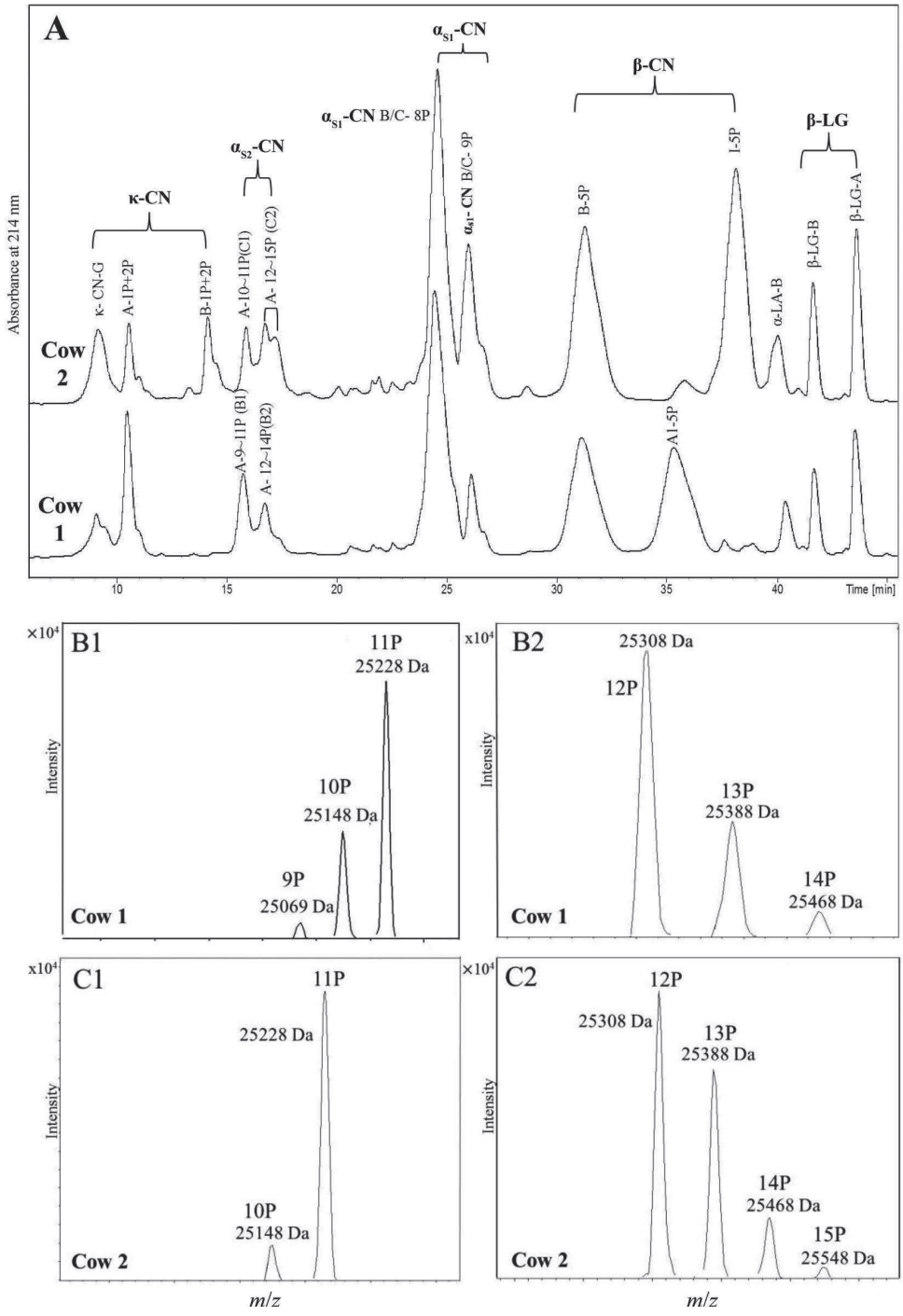

Figure 2.1. Milk protein profiling by liquid chromatography coupled with electrospray ionization mass spectrometry (LC/ESI-MS) of 2 individual bovine milk samples with distinct α_{s2}-casein profiles. (A) Peaks of protein variants and isoforms (G = glycan, P = phosphate group) of major milk proteins are identified by LC/ESI-MS. (B1 and B2) α_{s2}-Casein profile of cow 1 from deconvolution of multicharged-ion spectra corresponding to chromatographic peak B1 and B2. The B1 spectra show the presence of α_{s2}-CN-9P. (C1 and C2) α_{s2}-Casein profile of cow 2 from deconvolution of multicharged-ion spectra corresponding to chromatographic peak C1 and C2. The C2 spectra show the presence of α_{s2}-CN-15P.

2.3.2 Variation in major milk proteins

Detailed milk protein composition of Montbéliarde cows is of great interest because cheese-making properties are considered important in this breed. In Franche-Comté, milk from the Montbéliarde breed is used to produce regional cheeses with Protected Designation of Origin certification. However, not much information is available about the Montbéliarde breed. Table 2.1 summarizes the descriptive statistics for milk production traits and milk protein composition, and frequencies of milk protein variants. α-Lactalbumin was the least abundant protein, and α_{s1}-CN the most abundant protein among milk samples from 529 Montbéliarde cows. The coefficient of variation values ranged between 7 and 20%, implying substantial variation in milk protein composition among milk of individual cows. The average relative protein concentrations for the 6 major milk proteins are in the range of results from previous studies in other breeds (Bobe et al., 1998; Heck et al., 2008).

Two protein variants were found for α_{s1}-CN (B and C) and α_{s2}-CN (A and D), 3 for κ-CN (A, B, and E) and β-LG (A, B, and D), and 4 for β-CN (A1, A2, B, and I). Protein variants κ-CN E and β-CN I have not been reported in the French Montbéliarde population before. These variants may have been missed in earlier studies (Grosclaude, 1988) because the analytical technique used (gel electrophoresis) could not detect these 2 protein variants (Miranda et al., 1993; Jann et al., 2002). The frequency of κ-CN E is lower, whereas the frequency of κ-CN B is higher in the studied Montbéliarde population compared with other cattle breeds, such as Dutch Holstein Friesian, Danish Holstein, Italian Holstein Friesian, and Swedish Red (Heck, 2009; Gustavsson et al., 2014; Chessa et al., 2014). κ-Casein B has been associated with better coagulation properties of milk, whereas for κ-CN E the opposite has been shown (Hallén et al., 2008). Milk from the Montbéliarde breed is considered to possess better cheese-making properties than milk from Holstein and Nordic cattle breeds because of this characteristic (Montbéliarde, 2014). Regarding β-CN, the frequency of the B variant was moderate (0.355) and has increased at the expense of the A1 variant since 1976 (Grosclaude, 1988). The β-CN B variant has been positively associated with rennet coagulation and cheese-making properties (Caroli et al., 2009), whereas the β-CN A1 variant has been negatively associated with milk yield and protein yield (Ikonen et al., 1999; Visker et al., 2011). Because selection based on specific milk protein variants has not been implemented in Montbéliarde breed, the increased frequency of the β-CN B variant could be either due to genetic drift or selection on milk production traits.

2.3.3 Phenotypic correlations among six major milk proteins

Table 2.2 presents the phenotypic correlations among the 6 major milk proteins. Strong negative correlation was found between β-LG and $\sum$casein (-0.74). This correlation supports previous findings (Wedholm et al., 2006; Ng-Kwai-Hang, 2006; Schopen et al., 2009). As for the relationships among the 4 caseins, the phenotypic correlations were relatively low and ranged from -0.33 to +0.11. Generally, these low correlations among the caseins are in line with those reported by (Gebreyesus et al., 2016; Schopen et al., 2009) with 2 exceptions. Schopen et al. (2009) reported a moderate and negative phenotypic correlation between α_{s1}-CN and α_{s2}-CN (-0.50 vs. 0.07 in this study), and Gebreyesus et al. (2016) reported a moderate and positive

phenotypic correlation between α_{s1}-CN and β-CN (0.48 vs. -0.33 in this study). The discrepancies between these studies might be explained by genetic differences between the 3 breeds (Montbéliarde, Holstein Friesian, and Danish Holstein) and use of different analytical methods [capillary zone electrophoresis (CZE) and LC/ESI-MS)]. As some protein variants have been shown to be associated with the relative concentrations of certain protein fractions, differences in frequencies of protein variants between breeds could affect the concentrations and, subsequently, the phenotypic correlations between protein fractions (Heck et al., 2009; Bonfatti, 2010). In terms of analytical methods, CZE may have provided ambiguous estimations of α_{s2}-CN, β-CN, and κ-CN because α_{s2}-CN-13P was not detected, measurement of β-CN was partially disturbed by compounds co-migrating with β-CN such as glycosylated and multi-phosphorylated κ-CN, and measurement of κ-CN involved only the mono-phosphorylated isoform (Heck et al., 2008). Furthermore, measurements of protein fractions using the same analytical method, such as LC, could differ because of different separation conditions.

Table 2.2. Phenotypic correlations[1] among the 6 major milk proteins (% wt/wt of total protein in milk) measured on test-day morning milk samples from 529 Montbéliarde cows

Trait	α_{s1}-CN	α_{s2}-CN	β-CN	κ-CN	∑casein	α-LA	β-LG
α_{s2}-CN	0.07						
β-CN	−0.33	−0.32					
κ-CN	−0.14	0.11	−0.17				
∑casein[2]	0.41	0.00	0.63	0.07			
α-LA	0.15	0.14	−0.16	−0.10	−0.03		
β-LG	−0.37	0.01	−0.35	−0.18	−0.74	−0.24	
∑whey[3]	−0.32	0.06	−0.42	−0.22	−0.77	0.14	0.93

[1] SE (0.004–0.04).
[2] ∑casein = α_{s1}-CN + α_{s2}-CN + β-CN + κ-CN.
[3] ∑whey = α-LA + β-LG.

2.3.4 Variation in α_{s1}-CN and α_{s2}-CN phosphorylation isoforms

Among milk samples from 529 Montbéliarde cows, we found 77% of α_{s1}-CN carrying 8P and 23% carrying 9P, on average. Thus, α_{s1}-CN-8P was the abundant isoform whose mean concentration accounted for 25.3% of total protein in milk (Table 2.1). For α_{s2}-CN phosphorylation isoforms, we estimated only the relative concentrations of α_{s2}-CN-10P to α_{s2}-CN-14P from the A variant because only one milk sample contained the D variant (which is present with isoforms from 7P to 11P). The α_{s2}-CN-9P and α_{s2}-CN-15P occurred in trace amounts, and they were not observed in all milk samples. Out of 529 milk samples, α_{s2}-CN-9P was detected in 21 samples, and α_{s2}-CN-15P was detected in 4 different samples. For the remaining samples, the concentrations of these 2 isoforms may have been too low to be detected by MS. Among milk samples from 529 Montbéliarde cows, we found 8% of α_{s2}-CN having 10P, 36.5% having 11P, 32% hav-

ing 12P, 19% having 13P, and 4.5% having 14P, on average. Thus, α_{s2}-CN-11P and α_{s2}-CN-12P were the abundant isoforms, whose mean concentrations accounted for 3.1 and 2.7% of the total protein in milk, whereas α_{s2}-CN-13P had a mean concentration of 1.6%. The coefficient of variation values of α_s-CN phosphorylation isoforms ranged between 8 and 42%, suggesting large variation in relative concentrations of different α_s-CN phosphorylation isoforms. The estimated mean concentrations of α_{s1}-CN-8P and α_{s1}-CN-9P and of α_{s2}-CN-10P to α_{s2}-CN-12P, as well as their coefficient of variation values, were comparable with the results reported by Heck et al. (2008). However, Heck et al. (2008) did not detect α_{s2}-CN-9P, α_{s2}-CN-13P, α_{s2}-CN-14P, and α_{s2}-CN-15P because of the technique used (CZE) in that study.

2.3.5 Phenotypic correlations among α_s-CN phosphorylation isoforms and the six major milk proteins

As 3 new α_{s2}-CN phosphorylation isoforms were detected, we further explored the relationships among different α_s-CN phosphorylation isoforms and with other major milk proteins (Table 2.3). Correlations between the relative concentrations of different α_s-CN phosphorylation isoforms and those of the other 4 major milk proteins were relatively low (from -0.34 to 0.25). Correlations related to α_{s1}-CN-8P concentration were in line with those reported by Gebreyesus et al. (2016) except for the correlation between α_{s1}-CN-8P and β-CN (0.42 vs. -0.28 in this study). The discrepancy between the 2 studies could be explained by differences between studied breeds and analytic methods used as mentioned above. A strong positive correlation was found between α_{s2}-CN-13P and α_{s2}-CN-14P (0.92). The α_{s2}-CN-10P correlated positively with α_{s2}-CN-11P (0.62), but correlated negatively with α_{s2}-CN-13P and α_{s2}-CN-14P (-0.58 and -0.55, respectively). The α_{s2}-CN-12P correlated weakly with the rest of α_{s2}-CN phosphorylation isoforms (-0.28 to 0.19). Correlation between α_{s1}-CN-8P and α_{s1}-CN-9P was weak (-0.12), which agrees with previous findings reported by Bijl et al. (2014).

Table 2.3. Phenotypic correlations[1] among α_s-CN phosphorylation isoforms and the 6 major milk proteins (% wt/wt of total protein in milk) measured on test-day morning milk samples from 529 Montbéliarde cows

Trait	α_{s1}-CN-8P	α_{s1}-CN-9P	α_{s2}-CN-10P	α_{s2}-CN-11P	α_{s2}-CN-12P	α_{s2}-CN-13P	α_{s2}-CN-14P
Major milk protein							
α_{s1}-CN	0.90	0.33	0.11	−0.08	−0.01	−0.19	−0.14
α_{s2}-CN	0.22	−0.32	0.49	0.66	0.17	0.14	0.08
β-CN	−0.28	−0.16	−0.09	−0.08	−0.24	−0.10	−0.04
κ-CN	−0.17	0.04	−0.17	0.10	0.01	0.23	0.17
α-LA	0.22	−0.12	0.25	0.19	0.18	−0.30	−0.32
β-LG	−0.34	−0.11	−0.09	−0.09	0.06	0.01	−0.04
Phosphorylation isoform							
α_{s1}-CN-9P	−0.12						
α_{s2}-CN-10P	0.43	−0.68					
α_{s2}-CN-11P	0.22	−0.67	0.62				
α_{s2}-CN-12P	−0.10	0.19	−0.07	−0.28			
α_{s2}-CN-13P	−0.43	0.50	−0.58	−0.28	0.11		
α_{s2}-CN-14P	−0.38	0.49	−0.55	−0.27	−0.11	0.92	

[1] SE (0.01–0.04).

The relationships between phosphorylation of α_{s1}-CN and phosphorylation of α_{s2}-CN have rarely been investigated before due to the limitations of analytic methods for quantifying α_{s2}-CN phosphorylation isoforms. The dendrogram constructed from hierarchical analysis shows 2 main groups: one group of α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P, and another group of α_{s1}-CN-9P, α_{s2}-CN-12P, α_{s2}-CN-13P, and α_{s2}-CN-14P (Figure 2.2). Both phenotypic correlations and hierarchical clustering suggest that α_s-CN phosphorylation might involve at least 2 different regulatory systems. One system might be involved in forming isoforms with a lower degree of phosphorylation (α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P), whereas another system favors a higher degree of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-12P, α_{s2}-CN-13P, and α_{s2}-CN-14P). This is in line with Bijl et al. (2014), who showed that 2 different chromosomal regions are associated with α_{s1}-CN-8P and α_{s1}-CN-9P concentration, respectively. Consequently, they hypothesize that a different set of genes regulates the phosphorylation of α_{s1}-CN-8P and α_{s1}-CN-9P. Our results support this hypothesis and further suggest that one regulatory system is more effective than the other depending on the individual cow. As illustrated by the relationships between α_{s1}-CN-9P and the other α_s-CN isoforms, the proportion of α_s-CN isoforms with a lower degree of phosphorylation correlated negatively with the proportion of α_s-CN isoforms with a higher degree of phosphorylation. In other words, in the milk of an individual cow, either α_s-CN isoforms with a lower degree of phosphorylation would be more abundant than those with a higher degree of phosphorylation or vice versa.

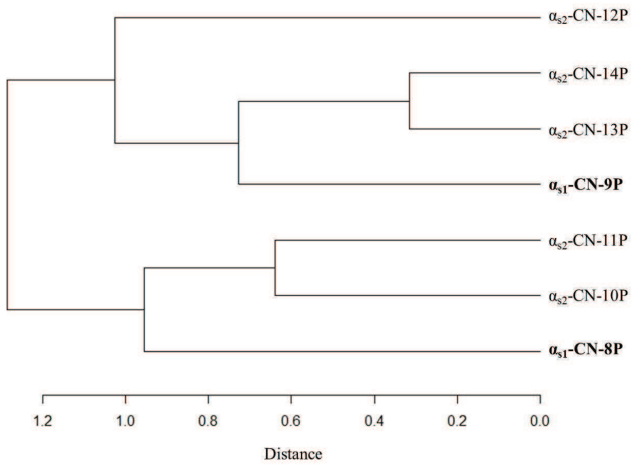

Figure 2.2. Dendrogram based on divisive hierarchical clustering analysis of phenotypic correlations among α_{s2}-CN phosphorylation isoforms (% wt/wt of total protein in milk) measured on test-day morning milk samples from 529 Montbéliarde cows.

Phosphorylation could be influenced by many factors, such as genetic expression and efficiency of protein kinases, substrate availability, and accessibility of phosphorylation sites as a result of the conformational structure of proteins (Holland and Boland, 2014). It is still unclear what factors regulate the differences in α_s-CN phosphorylation among milk of individual cows. Two casein kinases with different substrate specificity could explain the hypothesis of one regulatory system responsible for isoforms with a lower degree of phosphorylation and

another responsible for isoforms with a higher degree of phosphorylation. The casein kinases from the Golgi-enriched fraction of the lactating mammary gland (termed G-CK) are thought to phosphorylate caseins in vivo (Bingham and Farrell, Jr., 1974; Moore et al., 1985). The G-CK has been shown to phosphorylate consensus phosphorylation sites (Ser/Thr-X-Glu/SerP/Asp) with different efficiency. Ser-X-Glu is the most effective motif, Thr-X-Glu and Ser-X-Asp are less effective to be phosphorylated, and Thr-X-Asp has never been shown to be phosphorylated (Mercier, 1981; Meggio et al., 1988; Lasa-Benito et al., 1996). Bovine α_{s1}-CN has 10 potential phosphorylation sites, but only α_{s1}-CN-8P and α_{s1}-CN-9P are the common isoforms. The α_{s1}-CN-9P carries an extra phosphorylated serine at Ser41 in the Ser-X-Asp motif of the mature peptide chain (Manson et al., 1977). Furthermore, bovine α_{s2}-CN has 18 potential phosphorylation sites, but only 12 serine residues are in the consensus motifs. Until now, only 10 phosphorylation sites of α_{s2}-CN are identified, and they are in Ser-X-Glu/SerP motifs (Imanishi et al., 2007; Baum et al., 2013). Consequently, some threonine residues among the 2 Thr-X-Asp and 4 Thr-X-Glu motifs in α_{s2}-CN have to be phosphorylated for α_{s2}-CN carrying more than 12P. Additionally, threonine in Thr-X-Glu motifs would be more likely to be phosphorylated because phosphorylated threonine in Thr-X-Glu motifs has been identified in caprine, human, and equine β-CN when they are phosphorylated at the highest level (Greenberg et al., 1984; Neveu et al., 2002; Matéos et al., 2010). Recently, Fam20C has been identified as a bona fide G-CK, and FAM20C is also highly expressed in human breast (Lizio et al., 2015; The GTEx Consortium, 2015). However, Fam20C phosphorylates only Ser-X-Glu/SerP motifs (Ishikawa et al., 2012; Tagliabracci et al., 2012; Tibaldi et al., 2015; Tagliabracci et al., 2015), so it could be responsible only for forming α_s-CN isoforms with a lower degree of phosphorylation. To form α_s-CN isoforms with a higher degree of phosphorylation, another mammary G-CK should be responsible for phosphorylating serine or threonine in less favorable motifs, such as those in Ser/Thr-X-Asp or Thr-X-Glu motifs. This hypothesis is further supported by the phosphorylated serine in the Ser-X-Asp motif of α_{s1}-CN-9P. Identifying all phosphorylation sites of α_{s2}-CN and investigating the genetic background of different α_{s2}-CN phosphorylation isoforms may provide further insight into the phosphorylation mechanism of caseins.

2.4 Conclusions

We characterized the bovine α_{s1}-CN and α_{s2}-CN phosphorylation profile using the LC/ESI-MS method. This is the first study to show the presence of α_{s2}-CN-9P, α_{s2}-CN-14P, and α_{s2}-CN-15P, in addition to α_{s1}-CN-8P and α_{s1}-CN-9P and α_{s2}-CN-10P, α_{s2}-CN-11P, α_{s2}-CN-12P, and α_{s2}-CN-13P. Relative concentrations of α_{s1}-CN and α_{s2}-CN phosphorylation isoforms varied considerably among milk of individual cows. Interestingly, our results suggest 2 regulatory systems for the phosphorylation of α_s-CN: one responsible for isoforms with a lower degree of phosphorylation and the other responsible for isoforms with a higher degree of phosphorylation. Investigating the genetic background of α_{s2}-CN phosphorylation would help to elucidate the phosphorylation mechanism of caseins.

Acknowledgments

The first author benefited from an Erasmus-Mundus fellowship and a grant from APIS-GENE (Paris, France), within the framework of the European Graduate School in Animal Breeding and Genetics (EGS- ABG, Paris, France). This work was part of the French "FROM'MIR" project, funded by the National Agricultural and Rural Development Program (CASDAR), Regional Union of Protected Designation cheeses of Franche-Comté (URFAC, Poligny, France), CEL25-90 (Conseil en Elevage, Roulans, France), Centre National Interprofessionnel de l'Economie Laitière (CNIEL, Paris, France), and Conseil Régional de Franche Comté (Besançon, France).

References

Baum, F., J. Ebner, and M. Pischetsrieder (2013). Identification of multiphosphorylated peptides in milk. *J. Agric. Food Chem.* 61.38, pp. 9110–9117.

Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk, and H. Bovenhuis (2014). Phosphorylation of αS1-casein is regulated by different genes. *J. Dairy Sci.* 97.11, pp. 7240–7246.

Bingham, E. W. and H. M. Farrell, Jr. (1974). Casein kinase from the Golgi apparatus of lactating mammary gland. *J. Biol. Chem.* 249, pp. 3647–3651.

Bingham, E. W., H. M. Farrell, Jr., and J. J. Basch (1972). Phosphorylation of Casein: ROLE OF THE GOLGI APPARATUS. *J. Biol. Chem.* 247, pp. 8193–8194.

Bobe, G., D. Beitz, A. Freeman, and G. Lindberg (1998). Separation and quantification of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J. Agric. Food Chem.* 46.2, pp. 458–463.

Bonfatti, V. (2010). "Genetic analysis of milk protein composition and of its relationship with renneting properties of individual cow milk". PhD Thesis. Università degli Studi di Padova, Italy, p. 125.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier (2008). Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. *J. Chromatogr. A* 1195.1-2, pp. 101–106.

Bordin, G., F. Cordeiro Raposo, B. de La Calle, and A. R. Rodriguez (2001). Identification and quantification of major bovine milk proteins by liquid chromatography. *J. Chromatogr. A* 928.1, pp. 63–76.

Caroli, A. M., S. Chessa, and G. J. Erhardt (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92.11, pp. 5335–5352.

Chessa, S., O. Bulgari, R. Rizzi, L. Calamari, P. Bani, S. Biffani, and A. M. Caroli (2014). Selection for milk coagulation properties predicted by Fourier transform infrared spectroscopy in the Italian Holstein-Friesian breed. *J. Dairy Sci.* 97.7, pp. 4512–21.

Dalgleish, D. G. and M. Corredig (2012). The structure of the casein micelle of milk and its changes during processing. *Annu. Rev. Food Sci. Technol.* 3.1, pp. 449–467.

De Kruif, C. and C. Holt (2003). "Casein micelle structure, functions and interactions". In: *Adv. Dairy Chem. Proteins*. Ed. by P. F. Fox and P. L. H. McSweeney. 3rd. Vol. 1. JANUARY 2003. Springer US. Chap. 5, pp. 233–276.

De Kruif, C. G., T. Huppertz, V. S. Urban, and A. V. Petukhov (2012). Casein micelles and their internal structure. *Adv. Colloid Interface Sci.* 171-172, pp. 36–52.

Farrell, Jr., H. M., G. T. Bleck, E. M. Brown, J. E. Butler, L. K. Creamer, C. L. Hicks, C. M. Hollar, K. F. Ng-Kwai-Hang, and H. E. Swaisgood (2004). Nomenclature of the proteins of cows' milk–sixth revision. *J. Dairy Sci.* 87.6, pp. 1641–74.

Frederiksen, P. D., K. K. Andersen, M. Hammershøj, H. D. Poulsen, J. Sørensen, M. Bakman, K. B. Qvist, and L. B. Larsen (2011). Composition and effect of blending of noncoagulating, poorly coagulating, and well-coagulating bovine milk from individual Danish Holstein cows. *J. Dairy Sci.* 94.10, pp. 4787–4799.

Gebreyesus, G., M. S. Lund, L. Janss, N. A. Poulsen, L. B. Larsen, H. Bovenhuis, and A. J. Buitenhuis (2016). Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the Danish Holstein. *J. Dairy Sci.* 99.4, pp. 2863–2866.

Greenberg, R., M. L. Groves, and H. J. Dower (1984). Human beta-casein. Amino acid sequence and identification of phosphorylation sites. *J. Biol. Chem.* 259.8, pp. 5132–5138.

Grosclaude, F. (1988). Le polymorphisme génétique des principales lactoprotéines bovines. Relations avec la quantité, la composition et les aptitudes fromagères du lait. *INRA Prod. Anim* 1.tableau 1, pp. 5–17.

Gustavsson, F., A. J. Buitenhuis, M. Johansson, H. P. Bertelsen, M. Glantz, N. A. Poulsen, H. L. Mansson, H. Stalhammar, L. B. Larsen, C. Bendixen, M. Paulsson, and A. Andren (2014). Effects of breed and casein genetic variants on protein profile in milk from Swedish Red, Danish Holstein, and Danish Jersey cows. *J. Dairy Sci.* 97.6, pp. 3866–3877.

Hallén, E., A Wedholm, A Andrén, A Lundén, A. Wedholm, A. Andrén, and A. Lundén (2008). Effect of beta-casein, kappa-casein and beta-lactoglobulin genotypes on concentration of milk protein variants. *J. Anim. Breed. Genet.* 125.2, pp. 119–129.

Heck, J. M. L., A. Schennink, H. J. F. Van Valenberg, H. Bovenhuis, M. H. P. W. Visker, J. A. M. Van Arendonk, and A. C. M. van Hooijdonk (2009). Effects of milk protein variants on the protein composition of bovine milk. *J. Dairy Sci.* 92.3, pp. 1192–1202.

Heck, J. M. L. (2009). *Phd: Milk Genomics fatty acid composition in raw milk*, p. 144.

Heck, J., C. Olieman, A. Schennink, H. van Valenberg, M. Visker, R. Meuldijk, and A. van Hooijdonk (2008). Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *Int. Dairy J.* 18.5, pp. 548–555.

Hoagland, P. D., J. J. Unruh, E. D. Wickham, and H. M. Farrell (2001). Secondary structure of bovine αS2-casein: theoretical and experimental approaches. *J. Dairy Sci.* 84.9, pp. 1944–1949.

Holland, J. W. and M. J. Boland (2014). "Post-translational modifications of caseins". In: *Milk proteins From Expr. to food*. Ed. by H. Singh, A. Thompson, and M. Boland. Second Edi. Elsevier Inc., pp. 141–168.

Ikonen, T., M. Ojala, and O. Ruottinen (1999). Associations Between Milk Protein Polymorphism and First Lactation Milk Production Traits in Finnish Ayrshire Cows. *J. Dairy Sci.* 82.5, pp. 1026–1033.

Imanishi, S. Y., V. Kochin, S. E. Ferraris, A. de Thonel, H.-M. Pallari, G. L. Corthals, and J. E. Eriksson (2007). Reference-facilitated phosphoproteomics: fast and reliable phosphopeptide validation by microLC-ESI-Q-TOF MS/MS. *Mol. Cell. Proteomics* 6, pp. 1380–1391.

Ishikawa, H. O., A. Xu, E. Ogura, G. Manning, and K. D. Irvine (2012). The raine syndrome protein FAM20C is a golgi kinase that phosphorylates bio-mineralization proteins. *PLoS One* 7.8.

Jann, O., G. Ceriotti, A. Caroli, and G. Erhardt (2002). A new variant in exon VII of bovine β-casein gene (CSN2) and its distribution among European cattle breeds. *J. Anim. Breed. Genet.* 119, pp. 65–68.

Ng-Kwai-Hang, K. (2006). Genetic variants of milk proteins and their effects on the yield and quality of cheese. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 1.056, pp. 1–11.

Lasa-Benito, M., O. Marin, F. Meggio, and L. A. Pinna (1996). Golgi apparatus mammary gland casein kinase: monitoring by a specific peptide substrate and definition of specificity determinants. *FEBS Lett.* 382.1-2, pp. 149–52.

Lizio, M., J. Harshbarger, H. Shimoji, J. Severin, T. Kasukawa, S. Sahin, I. Abugessaisa, S. Fukuda, F. Hori, S. Ishikawa-Kato, C. J. Mungall, E. Arner, J. Baillie, N. Bertin, H. Bono, M. de Hoon, A. D. Diehl, E. Dimont, T. C. Freeman, K. Fujieda, W. Hide, R. Kaliyaperumal, T. Katayama, T. Lassmann, T. F. Meehan, K. Nishikata, H. Ono, M. Rehli, A. Sandelin, E. A. Schultes, P. 't Hoen, Z. Tatum, M. Thompson, T. Toyoda, D. W. Wright, C. O. Daub, M. Itoh, P. Carninci, Y. Hayashizaki, A. Forrest, and H. Kawaji (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 16.1, p. 22.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik (2015). *cluster: cluster analysis basics and extensions. R Package version 2.0. 1*.

Manson, W., T. Carolan, and W. D. Annan (1977). Bovine αS0 casein; a phosphorylated homologue of αS1 casein. *Eur. J. Biochem.* 417, pp. 411–417.

Matéos, A., J. Girardet, D. Mollé, C. Corbier, J. Gaillard, and L. Miclo (2010). Identification of phosphorylation sites of equine β-casein isoforms. *Rapid Commun. Mass Spectrom.* 24.11, pp. 1533–1542.

Meggio, F., A. P. Boulton, F. Marchiori, G. Borin, D. P. Lennon, A. Calderan, and L. A. Pinna (1988). Substrate-specificity determinants for a membrane-bound casein kinase of lactating mammary gland. A study with synthetic peptides. *Eur. J. Biochem.* 177.2, pp. 281–4.

Mercier, J. C. (1981). Phosphorylation of caseins, present evidence for an amino acid triplet code posttranslationally recognized by specific kinases. *Biochimie* 63.1, pp. 1–17.

Miranda, G., P. Anglade, M. F. Mahé, and G. Erhardt (1993). Biochemical characterization of the bovine genetic K-casein C and E variants. *Anim. Genet.* 24, pp. 27–31.

Miranda, G., L. Bianchi, and P. Martin (2011). *Bovine Milk Protein Masses Database (Genetic and Splicing variants, PTM isoforms and Plasmin peptides). INRA-France. APP interdeposit certification : IDDN.FR.001.460019.000.R.C.2011.000.10300, November 2011*.

Miranda, G., Z. Krupova, L. Bianchi, and P. Martin (2013). "A novel LC-MS protein profiling method to characterize and quantify individual milk proteins and multiple isoforms". In: *10th Annu. Int. Milk Genomics Consort. Symp.* October 1-3, 2013 U.C. Davis Conference Center, Davis, California USA.

Montbéliarde, O. S. (2014). *Etude réalisée sur 651,573 femelles à partir des typages de leurs 488 pères et grands-pères*.

Moore, A., A. P. Boulton, H. W. Heid, E. D. Jarasch, and R. K. Craig (1985). Purification and tissue-specific expression of casein kinase from the lactating guinea-pig mammary gland. *Eur. J. Biochem.* 152.3, pp. 729–37.

Neveu, C., D. Mollé, J. Moreno, P. Martin, and J. Léonil (2002). Heterogeneity of caprine beta-casein elucidated by RP-HPLC/MS: genetic variants and phosphorylations. *J. Protein Chem.* 21.8, pp. 557–567.

R Core Team (2015). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*.

Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. van Valenberg, and J. A. M. van Arendonk (2009). Genetic parameters for major milk proteins in Dutch Holstein-Friesians. *J. Dairy Sci.* 92.3, pp. 1182–1191.

Tagliabracci, V. S., J. L. Engel, J. Wen, S. E. Wiley, C. A. Worby, L. N. Kinch, J. Xiao, N. V. Grishin, and J. E. Dixon (2012). Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* 336.6085, pp. 1150–3.

Tagliabracci, V. S., S. E. Wiley, X. Guo, L. N. Kinch, E. Durrant, J. Wen, J. Xiao, J. Cui, K. B. Nguyen, J. L. Engel, J. J. Coon, N. Grishin, L. A. Pinna, D. J. Pagliarini, and J. E. Dixon (2015). A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* 161.7, pp. 1619–1632.

The GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.).* 348.6235, pp. 648–660.

Tibaldi, E., G. Arrigoni, G. Cozza, L. Cesaro, L. A. Pinna, E. Tibaldi, and G. Arrigoni (2015). “” Genuine ” Casein Kinase: The False Sister of CK2 That Phosphorylates Secreted Proteins at S-x-E/pS Motifs”. In: *Protein Kinase CK2 Cell. Funct. Norm. Dis. States*. Springer International Publishing, pp. 227–237.

Tuli, L. and H. W. Ressom (2009). LC–MS based detection of differential protein expression. *J. Proteomics Bioinform.* 02.202, pp. 416–438.

Visker, M. H. P. W., B. W. Dibbits, S. M. Kinders, H. J. F. van Valenberg, J. A. M. van Arendonk, and H. Bovenhuis (2011). Association of bovine β-casein protein variant I with milk production and milk protein composition. *Anim. Genet.* 42.2, pp. 212–218.

Visser, S., C. J. Slangen, and H. S. Rollema (1991). Phenotyping of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J. Chromatogr. A* 548.

Wedholm, A., L. B. Larsen, H. Lindmark-Månsson, A. H. Karlsson, and A. Andrén (2006). Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *J. Dairy Sci.* 89.9, pp. 3296–305.

Wedholm, A. (2008). “Variation in milk protein composition and its importance for the quality of cheese milk”. PhD Thesis. Swedish University of Agricultural Sciences, p. 53.

Chapter 3

Genetic and nongenetic factors contributing to differences in α_S-casein phosphorylation isoforms and other major milk proteins

Z.H. Fang[*†], H. Bovenhuis[†], A. Delacroix-Buchet[*], G. Miranda[*], D. Boichard[*], M.H.P.W. Visker[†] and P. Martin[*]

[*]Génétique Animale et Biologie Intégrative, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[†]Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

Journal of Dairy Science (2017) 100:5564-5577

Abstract

Relative concentrations of α_s-casein (α_s-CN) phosphorylation isoforms vary considerably among milk of individual cows. We aimed to explore to what extent genetic and other factors contribute to the variation in relative concentrations of α_s-CN phosphorylation isoforms and the phosphorylation degree of α_s-CN defined as the proportion of isoforms with higher degrees of phosphorylation. We also investigated the associations of genetic variants of milk proteins and casein haplotypes with relative concentrations of α_s-CN phosphorylation isoforms and with the phosphorylation degree of α_s-CN in French Montbéliarde cattle from the cheese production area of Franche-Comté. Detailed milk protein composition was determined by liquid chromatography coupled with electrospray ionization mass spectrometry from 531 test-day morning milk samples. Parity, lactation stage, and genetic variation of cows contributed to the phenotypic variation in relative concentrations of individual α_s-CN phosphorylation isoforms and in the phosphorylation degree of α_s-CN. As lactation progressed, we observed a significant increase for relative concentrations of α_s-CN isoforms with higher degrees of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-13P, and α_{s2}-CN-14P) as well as for the phosphorylation degree of both α_{s1}-CN and α_{s2}-CN. Furthermore, the β-CN I variant was associated with a greater proportion of isoforms with lower degrees of phosphorylation (α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P); the β-CN B variant was associated with a greater proportion of isoforms with higher degrees of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-12P to α_{s2}-CN-14P). The heritability estimates were low to moderate for relative concentrations of α_{s2}-CN phosphorylation isoforms (0.07 to 0.32), high for relative concentrations of α_{s1}-CN-8P (0.84) and α_{s1}-CN-9P (0.56), and moderate for phosphorylation degrees of α_{s1}-CN (0.37) and α_{s2}-CN (0.23). Future studies investigating relations between the phosphorylation degree of α_s-CN and technological properties of milk will be beneficial for the dairy industry.

Key words: casein haplotype; genetic variant; lactation stage; French Montbéliarde

3.1 Introduction

Detailed milk protein composition exhibits high heterogeneity because of quantitative variation in the content of the different milk proteins, numerous genetic variants, and isoforms with different degrees of posttranslational modifications such as glycosylation in κ-CN and phosphorylation in all caseins (Caroli et al., 2009; Holland and Boland, 2014). Phosphorylation of caseins is one of the key factors responsible for constructing and stabilizing casein micelles (De Kruif and Holt, 2003). Although all 4 caseins (α_{s1}-CN, α_{s2}-CN, β-CN, and κ-CN) are phosphoproteins, α_{s1}-CN and α_{s2}-CN are more phosphorylated, and their phosphorylation profiles are more heterogeneous than those of β-CN and κ-CN. In bovine milk, α_{s1}-CN accounts for about 35% of the total casein and has 2 common phosphorylation isoforms: α_{s1}-CN-8P and α_{s1}-CN-9P (where *P* indicates phosphate group attached); α_{s2}-CN accounts for about 10% of the total casein and is present with isoforms from 10 to 14P and occasionally with 9P or 15P (Holland and Boland, 2014; Fang et al., 2016). Relative concentrations of α_s-CN phosphorylation isoforms vary considerably among milk of individual cows (Bijl et al., 2014a; Fang et al., 2016).

The phosphorylation degree (**PD**) of α_s-CN is one of the factors affecting technological properties of milk. Bijl et al. (2014b) demonstrated that high α_{s1}-CN-8P concentration in bovine milk is a great benefit for the production of uncooked curd cheese because α_{s1}-CN-8P is hydrolyzed more efficiently by chymosin during ripening. Additionally, 2 studies have shown that poorly or noncoagulating milk was associated with a greater proportion of α_s-CN isoforms with higher degrees of phosphorylation (e.g., α_{s1}-CN-9P, α_{s2}-CN-12P, and α_{s2}-CN-13P) when compared with well-coagulating milk, although the numbers of observations were limited (Frederiksen et al., 2011; Jensen et al., 2012). Recently, Fang et al. (2016) suggested that α_s-CN isoforms with lower degrees of phosphorylation (e.g., α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P) might be regulated differently compared with α_s-CN isoforms with higher degrees of phosphorylation (e.g., α_{s1}-CN-9P, α_{s2}-CN-12P to α_{s2}-CN-14P). Therefore, it is of great interest to explore to what extent genetic and other factors contribute to the variation in α_s-CN phosphorylation profile. Bijl et al. (2014a) and Buitenhuis et al. (2016) reported the genetic parameters for relative concentrations of α_{s1}-CN-8P and α_{s1}-CN-9P in Dutch Holstein Friesians and in Danish Holstein and Jersey, respectively. However, no information is available regarding α_{s2}-CN phosphorylation profile as well as detailed milk protein composition in the Montbéliarde breed.

The objective of this study was to investigate the genetic and nongenetic sources of variation in the phosphorylation degree of α_s-CN, and in relative concentrations of α_s-CN phosphorylation isoforms and other major milk proteins in French Montbéliarde cattle from Franche-Comté cheese production area. We also investigated the associations of genetic variants of milk proteins and casein haplotypes with detailed milk protein composition.

3.2 Material and Methods

3.2.1 Milk samples

Test-day morning milk samples from 531 Montbéliarde cows were collected from 430 commercial herds across 3 French departments (239 herds in Doubs, 160 herds in Jura, and 31 herds in Haute-Saône) located in the production area of protected designation of origin (PDO) cheeses: Comté, Morbier, Mont d'Or, and Bleu de Gex, and of protected geographical indication (PGI) French Gruyère cheese. The sampling periods were during October-December 2014 and April-July 2015. The objective of the sampling was to maximize genetic diversity and milk content diversity to obtain optimal representation of the variation in milk protein composition from the current French Montbéliarde cattle population in the Franche-Comté region. For this purpose, we sampled cows across different parities (1-5) and lactation stages (7-652 d), and based on paternal pedigree and on protein and calcium content in milk from previous lactation records. Cows descended from 191 sires and 68 paternal grandsires, 52 of which were also maternal grandsires. Milk (25 mL) was preserved with Bronopol after collection, transported on ice to the laboratory, and then frozen at -20°C until analyzed by liquid chromatography **(LC)**/electrospray ionization **(ESI)**-MS.

3.2.2 Milk protein profiling

Milk protein composition was determined by LC/ESI-MS method developed at INRA as described in detail by Fang et al. (2016). Briefly, milk proteins were separated by reversed-phase HPLC using an Ultimate LC 3000 system (Thermo Fisher Scientific, Waltham, MA) with a Discovery BIOWide Pore (Supelco, Bellefonte, PA) C5 column (150 × 2.10 mm, 300 Å). Genetic variants and isoforms of the 6 major milk proteins (α_{s1}-CN, α_{s2}-CN, κ-CN, β-CN, α-LA, and β-LG) were identified using an ESI-TOF mass spectrometer micrOTOF II focus (Bruker Daltonics, Wissembourg, France). Relative concentrations of individual milk proteins and of α_{s1}-CN phosphorylation isoforms were estimated based on dividing the peak area of an individual protein by the total integrated peak area in the chromatogram (absorbance at 214 nm) of an individual milk sample. Mass signal intensity obtained from mass spectrometry was used to estimate the proportion of each α_{s2}-CN phosphorylation isoform as a fraction of total α_{s2}-CN. The relative concentration of each α_{s2}-CN phosphorylation isoform was estimated using the following equation:

$$\frac{\text{isoform mass signal intensity}}{\sum \text{isoform mass signal intensity}} \times \text{total } \alpha_{s2}\text{—CN (\%wt/wt).} \tag{3.1}$$

The phosphorylation degrees of α_{s1}-CN and α_{s2}-CN were defined as the proportion of isoforms with higher degrees of phosphorylation, which were calculated as α_{s1}-CN PD = (α_{s1}-CN-9P/total α_{s1}-CN) × 100 and α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

3.2.3 Statistical analyses

To estimate variance components and genetic parameters, the following animal model was used:

$$y_{ijklm} = \mu + \text{region}_i + \text{parity}_j + \text{lstage}_k + \text{season}_l + \text{animal}_m + \text{e}_{ijklm}, \quad (3.2)$$

where y_{ijklm} is the observation of the trait of interest; μ is the overall mean; region_i is the fixed effect of the *i*th region class (3 classes for the 3 French departments: Doubs, Jura, and Haute-Saône); parity_j is the fixed effect of the *j*th parity class (3 classes: 1st, 2nd, ≥3rd parity); lstage_k is the fixed effect of the *k*th lactation stage class [0-50 DIM (n = 95), 51-100 DIM (n = 73), 101-150 DIM (n = 69), 151-200 DIM (n = 76), 201-250 DIM (n = 82), 251-300 DIM (n = 66), >300 DIM (n = 68)]; season_l is the fixed effect of the *l*th season class (2 classes for 2 sampling seasons: October-December 2014 and April-July 2015); animal_m is the random additive genetic effect of animal *m* and is assumed to be distributed as $N(\mathbf{0},\mathbf{A}\sigma_a^2)$, where **A** is the additive genetic relationships matrix consisting of 5,546 animals with pedigree traced back for 5 generations, and σ_a^2 is the additive genetic variance; e_{ijklm} is the random residual effect and is assumed to be distributed as $N(\mathbf{0},\mathbf{I}\sigma_e^2)$, where **I** is the identity matrix, and σ_e^2 is the residual variance. The heritability was defined as

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2},$$

where the phenotypic variance $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$.

To estimate the effects of genetic variants of milk proteins on detailed milk protein composition, model 3.2 was extended with a genotype effect where the genotypes of the milk proteins were inferred from the genetic variants obtained from LC/ESI-MS analysis. The genotype classes containing less than 5 observations were excluded from the analysis. The proportion of additive genetic variance explained by the genetic variants was calculated as

$$\frac{\sigma_{genotype}^2}{\sigma_a^2} \times 100,$$

where $\sigma_{genotype}^2$ is calculated based on the estimated genotype effects obtained from the extended mode 3.2 and the observed frequencies of genetic variants, and σ_a^2 is the additive genetic variance estimated from model 3.2.

The α_{s1}-β-κ-CN haplotypes were inferred with Beagle 4.1 (Browning and Browning, 2007) adapted to multiallelic variants from α_{s1}-CN, β-CN, and κ-CN genetic variants. A multiallelic locus with n alleles (n > 2) was replaced by (n - 1) virtual fully linked biallelic loci. The allele coded as 1 from the *i*th biallelic virtual locus (1 ≤*i* <n) corresponded the i*th* allele of the original locus and the allele coded as 2 corresponded to any other allele. To estimate the effects of α_{s1}-β-κ-CN haplotypes on detailed milk protein composition, the following model was used:

$$y_{ijklmno} = \mu + \text{region}_i + \text{parity}_j + \text{lstage}_k + \text{season}_l + \text{animal}_m + \text{haplo1}_n + \text{haplo2}_o + \text{e}_{ijklmno}; \quad (3.3)$$

model 3.3 was model 3.2 extended with haplo1$_n$ and haplo2$_o$, where haplo1$_n$ is the effect of the first α_{s1}-β-κ-CN haplotype, and haplo2$_o$ is the effect of the second α_{s1}-β-κ-CN haplotype. The 2 haplotypes per animal were randomly assigned to haplo1$_n$ or haplo2$_o$, and the design matrix for haplo1$_n$ was added to the design matrix for haplo2$_o$. The haplo1$_n$ or haplo2$_o$ were modeled as fixed effects to estimate the effect of having 1 copy of each haplotype. The haplotype classes containing less than 5 observations were excluded from the analysis. The proportion of additive genetic variance explained by haplotypes was calculated as

$$\frac{\sigma^2_{haplotype}}{\sigma^2_a},$$

where haplotypes were modeled as random effects assumed to be distributed as $N(\mathbf{0}, \mathbf{I}\sigma^2_{haplo})$ All statistical analyses were performed using ASReml (Gilmour et al., 2009).

3.3 Results

3.3.1 Genetic parameters

Table 3.1 summarizes the means, standard deviations, and genetic parameters for relative concentrations of the 6 major milk proteins and α_s-CN phosphorylation isoforms, and for the phosphorylation degree of α_s-CN (α_s-CN PD). Heritability estimates for relative concentrations of the 6 major milk proteins were moderate to high and ranged from 0.22 (α-LA) to 1.00 (α_{s1}-CN). The standard error of the heritability estimate for α_{s1}-CN concentration could not be approximated accurately as the estimate was at the boundary of the parameter space. Likelihood ratio test suggested that the 95% confidence interval of the heritability estimate for α_{s1}-CN concentration ranged from 0.75 to 1.00. For α_s-CN phosphorylation isoforms, heritability estimates for relative concentrations of α_{s1}-CN-8P (0.84) and α_{s1}-CN-9P (0.56) were high, and for α_{s1}-CN PD (0.37) was moderate. The heritability estimates for relative concentrations of α_{s2}-CN phosphorylation isoforms and for α_{s2}-CN PD were low to moderate (0.07 to 0.32).

3.3.2 Effects of parity and lactation stage on detailed milk protein composition

Parity significantly affected relative concentrations of β-CN and α-LA (all P <0.001; Table 3.2). For α_s-CN phosphorylation isoforms, parity significantly affected α_{s1}-CN PD, α_{s2}-CN PD, and relative concentrations of all α_s-CN phosphorylation isoforms except α_{s2}-CN-11P and α_{s2}-CN-12P. Lactation stage significantly affected α_{s1}-CN PD, α_{s2}-CN PD, and relative concentrations of all α_s-CN phosphorylation isoforms except α_{s2}-CN-12P (all P <0.001, Figure 3.1). The changes

Table 3.1. Mean, SD, phenotypic variance[1] (σ_p^2) and h^2 for the 6 major milk proteins, the individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and the phosphorylation degrees (PD)[2] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows (SE in parentheses)

Trait	Mean	SD	σ_p^2	h^2
Major milk protein (% wt/wt)				
α_{s1}-CN	32.92	2.18	4.51	1.00 (-)
α_{s2}-CN	8.22	0.88	0.74	0.29 (0.14)
κ-CN	9.03	0.86	0.73	0.62 (0.20)
β-CN	28.14	2.69	7.48	0.75 (0.19)
α-LA	3.54	0.70	0.32	0.22 (0.12)
β-LG	12.16	1.87	3.41	0.73 (0.19)
Phosphorylation isoform (% wt/wt)				
α_{s1}-CN-8P	25.27	2.08	3.60	0.84 (0.18)
α_{s1}-CN-9P	7.65	0.96	0.84	0.56 (0.18)
α_{s1}-CN PD	23.27	2.80	6.19	0.37 (0.15)
α_{s2}-CN-10P	0.72	0.30	0.08	0.11 (0.09)
α_{s2}-CN-11P	3.04	0.55	0.28	0.32 (0.14)
α_{s2}-CN-12P	2.68	0.34	0.11	0.09 (0.12)
α_{s2}-CN-13P	1.57	0.31	0.07	0.07 (0.11)
α_{s2}-CN-14P	0.40	0.14	0.02	0.14 (0.13)
α_{s2}-CN PD	57.05	8.36	55.94	0.23 (0.12)

[1] Phenotypic variance after adjusting for systematic effects: sampling region and season, parity, and lactation stage.
[2] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

in relative concentrations of individual α_s-CN phosphorylation isoforms during lactation are shown in Figure 3.1. The magnitude of lactation stage effects is expressed as a fold change in respective phenotypic standard deviations of relative concentrations of α_s-CN isoforms (as given in Table 3.1) to facilitate comparison across isoforms. The magnitude of the effects varied from 0.01 to 1.5 phenotypic standard deviation. As lactation progressed, we observed a significant decrease in relative concentrations for the group of isoforms with lower degrees of phosphorylation (α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P) and a significant increase in relative concentrations for the group of isoforms with higher degrees of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-13P, and α_{s2}-CN-14P) as well as a significant increase in both α_{s1}-CN PD and α_{s2}-CN PD.

Table 3.2. Effects of parity on the 6 major milk proteins, individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and phosphorylation degrees (PD)[1] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows (SE in parentheses)

Trait	Parity			
	1 (n = 64)	2 (n = 197)	≥3(n = 269)	-Log$_{10}$(P)
Major milk protein (% wt/wt)				
α_{s1}-CN	0	-0.19 (0.15)	-0.23 (0.15)	0.5^{NS}
α_{s2}-CN	0	-0.09 (0.14)	-0.11 (0.14)	0.1^{NS}
κ-CN	0	0.14 (0.13)	0.14 (0.14)	0.2^{NS}
β-CN	0	-0.54 (0.42)	1.45 (0.44)	3.7^{***}
α-LA	0	-0.16 (0.09)	-0.13 (0.09)	31.9^{***}
β-LG	0	-0.14 (0.29)	0.16 (0.30)	0.6^{NS}
Phosphorylation isoform (% wt/wt)				
α_{s1}-CN-8P	0	-0.47 (0.29)	-0.72 (0.30)	1.3^{*}
α_{s1}-CN-9P	0	0.04 (0.15)	0.26 (0.15)	1.5^{*}
α_{s1}-CN PD	0	0.45 (0.40)	1.20 (0.41)	3.0^{***}
α_{s2}-CN-10P	0	-0.03 (0.05)	-0.11 (0.05)	2.4^{**}
α_{s2}-CN-11P	0	-0.04 (0.08)	-0.09 (0.08)	0.4^{NS}
α_{s2}-CN-12P	0	0.08 (0.05)	0.09 (0.05)	0.7^{NS}
α_{s2}-CN-13P	0	0.02 (0.04)	0.14 (0.04)	5.3^{***}
α_{s2}-CN-14P	0	0.00 (0.02)	0.05 (0.02)	3.6^{***}
α_{s2}-CN PD	0	1.87 (1.22)	4.30 (1.22)	3.8^{***}

[1] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.
* P <0.05 ** P <0.01 *** P <0.001

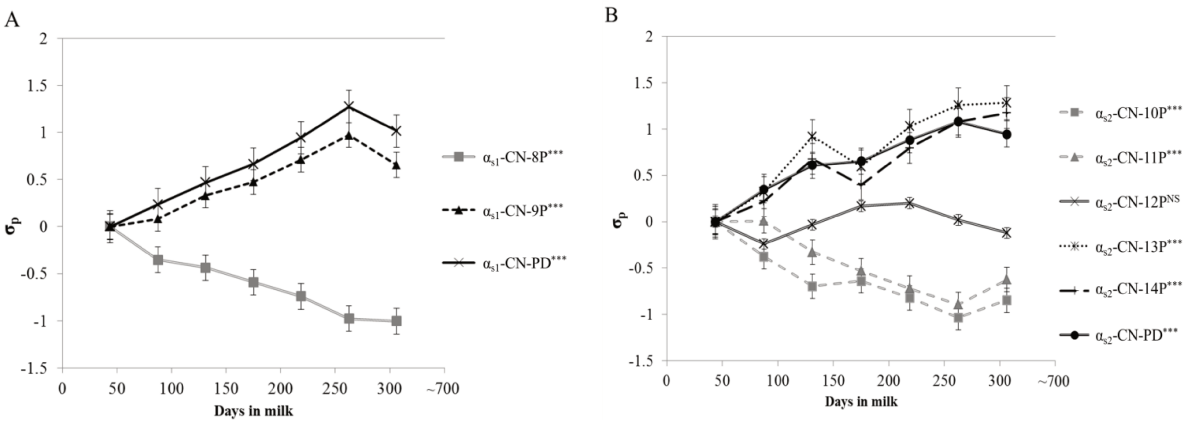

Figure 3.1. Effects of lactation stage on α_{s1}-CN (A) and α_{s2}-CN (B) phosphorylation profile and their phosphorylation degree (PD) throughout lactation. The x-axis shows days in milk. The y-axis shows the effect of lactation stage expressed as a fold change in phenotypic standard deviation (σp) of the relative concentration of each phosphorylation isoform. α_{s1}-CN PD = (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100. ***P <0.001.

3.3.3 Effects of genetic variants and casein haplotypes on detailed milk protein composition

We investigated the effects of genetic variants of milk proteins and α_{s1}-β-κ-CN haplotypes on relative concentrations of the 6 major milk proteins and α_s-CN phosphorylation isoforms, and α_s-CN PD. Two genetic variants were detected for α_{s1}-CN (B and C) and α_{s2}-CN (A and D), 3 for κ-CN (A, B, and E) and β-LG (A, B, and D), and 4 for β-CN (A1, A2, B, and I) in the sampled population (Fang et al., 2016).

Effects of α_{s1}-CN variants. The α_{s1}-CN genotypes affected relative concentrations of α_{s1}-CN, κ-CN, β-CN, and α-LA (all P <0.001; Table 3.3); the α_{s1}-CN genotype explained 22% of additive genetic variance of α_{s1}-CN concentration. The C variant was positively associated with α_{s1}-CN and α-LA concentrations and negatively associated with κ-CN and β-CN concentrations. For α_s-CN phosphorylation profile, α_{s1}-CN genotypes affected relative concentrations of all α_s-CN phosphorylation isoforms except α_{s2}-CN-12P and α_{s2}-CN-14P (P <0.001 to P = 0.046; Table 3.3). The C variant was positively associated with α_{s1}-CN-8P, α_{s1}-CN-9P, and α_{s2}-CN-10P concentrations and negatively associated with α_{s2}-CN-11P and α_{s2}-CN-13P concentrations. The α_{s1}-CN genotypes did not significantly affect α_s-CN PD.

Table 3.3. Effects of α_{s1}-CN genotypes on the 6 major milk proteins, individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and phosphorylation degrees (PD)[1] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows (SE in parentheses)

Trait	α_{s1}-CN genotype			
	BB (n =469)	BC (n = 57)	CC (n = 5)	-Log$_{10}$(P)
Major milk protein (% wt/wt)				
α_{s1}-CN	0	2.75 (0.25)	5.80 (0.85)	29.0***
α_{s2}-CN	0	-0.17 (0.12)	0.39 (0.40)	0.6^{NS}
κ-CN	0	-0.87 (0.11)	-1.35 (0.37)	14.0***
β-CN	0	-2.26 (0.36)	-4.01 (1.21)	9.7***
α-LA	0	0.23 (0.08)	0.73 (0.26)	3.1***
β-LG	0	-0.12 (0.25)	-1.71 (0.86)	0.9^{NS}
Phosphorylation isoform (% wt/wt)				
α_{s1}-CN-8P	0	2.21 (0.23)	5.17 (0.79)	23.6***
α_{s1}-CN-9P	0	0.55 (0.13)	0.71 (0.42)	4.4***
α_{s1}-CN PD	0	-0.22 (0.36)	-1.61 (1.17)	0.5^{NS}
α_{s2}-CN-10P	0	0.05 (0.04)	0.29 (0.13)	1.3*
α_{s2}-CN-11P	0	-0.26 (0.07)	-0.05 (0.23)	2.8**
α_{s2}-CN-12P	0	0.00 (0.04)	0.04 (0.14)	0.0^{NS}
α_{s2}-CN-13P	0	-0.10 (0.04)	-0.12 (0.13)	1.6*
α_{s2}-CN-14P	0	-0.03 (0.02)	-0.03 (0.06)	0.7^{NS}
α_{s2}-CN PD	0	-0.89 (1.07)	-4.61 (3.50)	0.5^{NS}

[1] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

* P <0.05 ** P <0.01 *** P <0.001

Effects of β-CN variants. The β-CN genotypes affected relative concentrations of all 6 major milk proteins (P <0.001 to P = 0.003, Supplemental Table S3.1); the β-CN genotypes explained 81% of additive genetic variance of β-CN concentration. The associations of individual genetic variants of β-CN with relative concentrations of the 6 major milk proteins are present in Table 3.4 where the number of alleles was modeled as a class variable. The β-CN B variant showed positive association with β-CN concentration and negative association with α_{s1}-CN, α_{s2}-CN, κ-CN, α-LA, and β-LG concentrations, whereas the A2 variant showed opposite associations. The A1 variant was positively associated with β-CN concentration and negatively associated with α_{s2}-CN, α-LA, and β-LG concentrations. The I variant was positively associated with α_{s2}-CN, κ-CN, and β-LG concentrations and negatively associated with β-CN concentration.

Table 3.4. Associations of β-CN genotypes with the 6 major milk proteins measured on test-day morning milk samples from 531 Montbéliarde cows[1] (SE in parentheses)

β-CN genotype[2]	α_{s1}-CN	α_{s2}-CN	κ-CN	β-CN	α-LA	β-LG
B						
0 (n = 210)	0	0	0	0	0	0
1 (n = 271)	-1.75 (0.15)	-0.31 (0.08)	-0.22 (0.07)	2.95 (0.19)	-0.18 (0.05)	-0.40 (0.16)
2 (n = 50)	-3.46 (0.26)	-0.95 (0.13)	-0.94 (0.13)	5.29 (0.32)	-0.42 (0.09)	-0.02 (0.28)
-Log$_{10}$(P)	39.2***	11.1***	11.3***	61.9***	5.4***	1.5*
A2						
0 (n = 180)	0	0	0	0	0	0
1 (n = 279)	1.63 (0.16)	0.34 (0.08)	0.16 (0.08)	-2.49 (0.20)	0.13 (0.05)	0.27 (0.17)
2 (n = 72)	3.89 (0.24)	0.25 (0.13)	0.47 (0.12)	-5.33 (0.31)	0.41 (0.09)	0.20 (0.27)
-Log$_{10}$(P)	46.8***	3.5***	3.1***	53.8***	5.3***	0.5^{NS}
A1						
0 (n = 437)	0	0	0	0	0	0
1 (n = 93)	-0.18 (0.23)	-0.51 (0.10)	-0.10 (0.10)	2.15 (0.30)	-0.14 (0.07)	-0.57 (0.21)
-Log$_{10}$(P)	1.3^{NS}	6.4***	0.5^{NS}	11.8***	1.5*	2.2**
I						
0 (n = 368)	0	0	0	0	0	0
1 (n = 153)	0.02 (0.20)	0.67 (0.08)	0.35 (0.08)	-1.50 (0.26)	0.10 (0.05)	0.35 (0.18)
2 (n = 10)	-0.36 (0.61)	1.22 (0.26)	0.44 (0.26)	-2.74 (0.79)	0.05 (0.18)	1.03 (0.55)
-Log$_{10}$(P)	0.1^{NS}	16.9***	4.1***	8.8***	0.7^{NS}	1.4*

[1] One cow with A1A1 genotype was excluded from the analysis.
[2] The number below the β-CN genotype indicates the number of alleles (0, 1, or 2) that a cow carried.
* P <0.05 ** P <0.01 *** P <0.001

For α_s-CN phosphorylation profile, β-CN genotypes affected relative concentrations of all α_s-CN phosphorylation isoforms (P <0.001 to P = 0.037; Supplemental Table S3.2) and α_{s2}-CN PD (P <0.001; Supplemental Table S3.2). The associations of individual genetic variants of β-CN with α_s-CN phosphorylation profile are present in Table 3.5. The B variant showed positive associations with both α_{s1}-CN PD and α_{s2}-CN PD, whereas the I variant showed opposite associations.

Table 3.5. Associations of β-CN genotypes with individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and the phosphorylation degree (PD)[1] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows[2] (SE in parentheses)

β-CN genotype[3]	α_{s1}-CN			α_{s2}-CN					
	8P	9P	PD	10P	11P	12P	13P	14P	PD
B									
0 (n = 210)	0	0	0	0	0	0	0	0	0
1 (n = 271)	-1.55 (0.14)	-0.23 (0.08)	0.53 (0.23)	-0.04 (0.03)	-0.08 (0.05)	-0.13 (0.03)	0.05 (0.03)	0.03 (0.01)	1.76 (0.67)
2 (n = 50)	-3.05 (0.25)	-0.45 (0.14)	1.12 (0.40)	-0.14 (0.05)	-0.28 (0.08)	-0.17 (0.05)	0.06 (0.04)	0.06 (0.02)	6.52 (1.18)
-$\text{Log}_{10}(P)$	35.9^{***}	2.9^{**}	2.1^{**}	2.1^{**}	2.7^{**}	5.7^{***}	1.1^{NS}	2.7^{**}	6.5^{***}
A2									
0 (n = 180)	0	0	0	0	0	0	0	0	0
1 (n = 279)	1.22 (0.15)	0.41 (0.08)	0.10 (0.24)	-0.03 (0.03)	-0.05 (0.05)	0.11 (0.03)	0.02 (0.03)	0.00 (0.01)	-1.00 (0.73)
2 (n = 72)	3.11 (0.23)	0.81 (0.13)	-0.26 (0.37)	-0.02 (0.04)	-0.19 (0.07)	0.22 (0.04)	-0.04 (0.04)	-0.03 (0.02)	-0.29 (1.10)
-$\text{Log}_{10}(P)$	34.2^{***}	9.4^{***}	0.3^{NS}	0.2^{NS}	1.4^{*}	6.2^{***}	0.5^{NS}	0.9^{NS}	0.4^{NS}
A1									
0 (n = 437)	0	0	0	0	0	0	0	0	0
1 (n = 93)	-0.04 (0.21)	-0.13 (0.11)	-0.25 (0.29)	0.01 (0.03)	-0.04 (0.06)	-0.11 (0.04)	-0.08 (0.03)	-0.03 (0.01)	1.09 (0.88)
-$\text{Log}_{10}(P)$	0.6^{NS}	0.8^{NS}	0.4^{NS}	0.2^{NS}	0.3^{NS}	2.5^{**}	1.9^{*}	0.9^{NS}	0.7^{NS}
I									
0 (n = 368)	0	0	0	0	0	0	0	0	0
1 (n = 153)	0.21 (0.18)	-0.23 (0.09)	-0.62 (0.25)	0.11 (0.03)	0.35 (0.05)	0.06 (0.03)	0.00 (0.00)	-0.01 (0.01)	-4.85 (0.72)
2 (n = 10)	-0.04 (0.56)	-0.47 (0.28)	-1.12 (0.78)	0.19 (0.09)	0.64 (0.15)	0.23 (0.10)	0.04 (0.09)	-0.10 (0.04)	-5.07 (2.29)
-$\text{Log}_{10}(P)$	0.3^{NS}	1.7^{*}	1.6^{*}	4.0^{***}	13.8^{***}	1.2^{NS}	0.1^{NS}	0.3^{NS}	9.8^{***}

[1] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

[2] One cow with A1A1 genotype was excluded from the analysis.

[3] The number below the β-CN genotype indicates the number of alleles (0, 1, or 2) that a cow carried.

* P <0.05 ** P <0.01 *** P <0.001

Effects of κ-CN variants. The κ-CN genotypes showed highly significant effects on milk protein composition, especially on κ-CN concentration (Table 3.6); the κ-CN genotype explained 65% of additive genetic variance of κ-CN concentration. The B variant was positively associated with relative concentrations of all major milk proteins except β-CN concentration for which the association was negative. For α_s-CN phosphorylation profile, κ-CN genotypes affected α_{s1}-CN-8P, α_{s2}-CN-11P and α_{s2}-CN-12P concentrations (P <0.001; Table 3.6), and the B variant was positively associated with relative concentrations of aforementioned isoforms. The κ-CN genotypes significantly affected α_{s2}-CN PD, and the A variant was positively associated with α_{s2}-CN PD.

Table 3.6. Effects of κ-CN genotypes on the 6 major milk proteins, individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and phosphorylation degrees (PD)[1] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows[2] (SE in parentheses)

Trait	κ-CN genotype			
	AA (n = 144)	AB (n = 268)	BB (n = 116)	-Log$_{10}$(P)
Major milk protein (% wt/wt)				
α_{s1}-CN	-0.15 (0.21)	0	0.82 (0.21)	3.9***
α_{s2}-CN	-0.40 (0.09)	0	0.44 (0.09)	12.7***
κ-CN	-1.12 (0.07)	0	0.25 (0.07)	53.0***
β-CN	1.87 (0.24)	0	-2.10 (0.25)	33.3***
α-LA	-0.12 (0.06)	0	0.14 (0.06)	2.6**
β-LG	-0.08 (0.19)	0	0.49 (0.19)	1.6*
Phosphorylation isoform (% wt/wt)				
α_{s1}-CN-8P	-0.18 (0.19)	0	0.78 (0.19)	4.4***
α_{s1}-CN-9P	0.06 (0.10)	0	0.05 (0.10)	0.1^{NS}
α_{s1}-CN PD	0.25 (0.26)	0	-0.42 (0.28)	0.9^{NS}
α_{s2}-CN-10P	0.01 (0.03)	0	0.01 (0.03)	0.0^{NS}
α_{s2}-CN-11P	-0.22 (0.05)	0	0.17 (0.05)	8.1***
α_{s2}-CN-12P	-0.01 (0.03)	0	0.17 (0.03)	5.9***
α_{s2}-CN-13P	0.02 (0.03)	0	0.02 (0.03)	0.4^{NS}
α_{s2}-CN-14P	0.00 (0.01)	0	-0.02 (0.01)	0.3^{NS}
α_{s2}-CN PD	2.49 (0.79)	0	-1.01 (0.83)	3.2***

[1] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

[2] One cow with AE genotype and 2 cows with BE genotype were excluded from the analysis.

* P <0.05 ** P <0.01 *** P <0.001

Effects of β-LG variants. The β-LG genotypes significantly affected relative concentrations of all major milk proteins except α_{s2}-CN concentration (P <0.001 to P = 0.002; Table 3.7); the β-LG genotypes explained 95% of additive genetic variance of β-LG concentration. The difference in β-LG concentration between AA and BB genotypes was about 4%. The B variant was positively associated with α_{s1}-CN, κ-CN, β-CN, and α-LA concentrations and negatively associated with β-LG concentration. For α_s-CN phosphorylation profile, β-LG genotypes affected

α_{s1}-CN-8P and α_{s1}-CN-9P concentrations (*P* <0.001; Table 3.7), and the B variant was positively associated with both isoforms' concentrations. The β-LG genotypes did not significantly affect α_s-CN PD.

Table 3.7. Effects of β-LG genotypes on the 6 major milk proteins, individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and phosphorylation degrees (PD)[1] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows[2] (SE in parentheses)

Trait	β-LG genotype			
	AA (n = 103)	AB (n = 259)	BB (n = 167)	-Log$_{10}$(*P*)
Major milk protein (% wt/wt)				
α_{s1}-CN	-0.84 (0.21)	0	1.32 (0.18)	3.6^{***}
α_{s2}-CN	0.01 (0.10)	0	0.17 (0.09)	0.9^{NS}
κ-CN	-0.27 (0.09)	0	0.33 (0.08)	7.2^{***}
β-CN	-0.84 (0.30)	0	0.41 (0.26)	3.0^{**}
α-LA	-0.09 (0.07)	0	0.21 (0.06)	4.3^{***}
β-LG	1.88 (0.11)	0	-2.41 (0.10)	135.9^{***}
Phosphorylation isoform (% wt/wt)				
α_{s1}-CN-8P	-0.61 (0.20)	0	0.99 (0.17)	11.4^{***}
α_{s1}-CN-9P	-0.29 (0.10)	0	0.32 (0.09)	6.1^{***}
α_{s1}-CN PD	-0.21 (0.29)	0	0.05 (0.25)	0.2^{NS}
α_{s2}-CN-10P	0.01 (0.03)	0	0.03 (0.03)	0.3^{NS}
α_{s2}-CN-11P	0.02 (0.06)	0	0.08 (0.05)	0.5^{NS}
α_{s2}-CN-12P	0.03 (0.04)	0	0.01 (0.03)	0.3^{NS}
α_{s2}-CN-13P	-0.01 (0.03)	0	0.03 (0.03)	0.4^{NS}
α_{s2}-CN-14P	0.01 (0.01)	0	0.01 (0.01)	0.3^{NS}
α_{s2}-CN PD	0.14 (0.87)	0	-0.61 (0.76)	0.2^{NS}

[1] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

[2] Two cows with BD genotype were excluded from the analysis.

* *P* <0.05 ** *P* <0.01 *** *P* <0.001

Effects of casein haplotypes. Inferring α_{s1}-β-κ-CN casein haplotypes resulted in 10 haplotypes: CA2A, BA2A, BA1A, BBA, BA1B, BA2B, BIB, BBB, BIA, and BA1E (frequencies provided in Supplemental Table S3.3). The α_{s1}-CN C variant occurred only with the β-CN A2 and κ-CN A variants. The κ-CN E variant occurred only with the β-CN A1 variant. The β-CN I variant occurred predominately with the κ-CN B variant. The α_{s1}-β-κ-CN haplotypes significantly affected relative concentrations of the 6 major milk proteins (*P* <0.001 to *P* = 0.0025); the proportions of additive genetic variance explained by haplotypes were high (0.46-0.89) for the 4 caseins and low to moderate (0.04-0.22) for the 2 whey proteins (Table 3.8). The haplotype carrying the α_{s1}-CN C variant was positively associated with α_{s1}-CN concentration. The haplotypes carrying the β-CN A2 and I variants were negatively associated with β-CN concentration and positively associated with α_{s1}-CN and α_{s2}-CN concentrations. The haplotypes carrying the κ-CN B variant were positively associated with κ-CN concentration. For the α_s-CN phosphory-

lation profile, the α_{s1}-β-κ-CN haplotypes significantly affected both α_{s1}-CN PD and α_{s2}-CN PD, and relative concentrations of all α_s-CN phosphorylation isoforms (*P* <0.001 to *P* = 0.03); the proportion of additive genetic variance explained by haplotypes was as high as 0.72 for α_{s1}-CN-8P concentration, and ranged from 0.04 to 0.39 for relative concentrations of the other α_s-CN phosphorylation isoforms and α_s-CN PD (Table 3.9). The haplotype carrying the β-CN I variant showed negative associations with both α_{s1}-CN PD and α_{s2}-CN PD, whereas the haplotype carrying the β-CN B variant showed opposite associations.

Table 3.8. Effects of α_{s1}-β-κ-CN haplotypes proportions of additive genetic variance explained by haplotypes (h_{haplo}) for the 6 major milk proteins measured on test-day morning milk samples from 531 Montbéliarde cows[1] (SE in parentheses)

α_{s1}-β-κ-CN haplotype[2]	α_{s1}-CN	α_{s2}-CN	κ-CN	β-CN	α-LA	β-LG
CA2A (n = 31)	4.08 (0.18)	0.11 (0.10)	-0.59 (0.09)	-3.81 (0.22)	0.38 (0.07)	-0.26 (0.24)
BA2A (n = 34)	2.87 (0.20)	0.23 (0.11)	-0.17 (0.09)	-3.04 (0.23)	0.24 (0.08)	0.22 (0.26)
BA1A (n = 26)	1.98 (0.21)	-0.07 (0.12)	-0.14 (0.10)	-0.24 (0.25)	0.07 (0.09)	0.14 (0.28)
BBA (n = 178)	0	0	0	0	0	0
BA1B (n = 22)	0.73 (0.24)	-0.45 (0.13)	0.35 (0.12)	-0.08 (0.28)	0.17 (0.10)	-0.52 (0.31)
BA2B (n = 141)	1.91 (0.12)	0.42 (0.07)	0.68 (0.06)	-3.29 (0.14)	0.21 (0.05)	0.24 (0.15)
BIB (n = 94)	0.91 (0.13)	0.81 (0.08)	0.53 (0.06)	-2.94 (0.16)	0.19 (0.06)	0.49 (0.18)
-Log_{10}(*P*)	85.7***	27.0***	45.8***	106.7***	7.1***	2.6**
h_{haplo}[3]	0.69 (0.15)	0.46 (0.20)	0.89 (0.21)	0.76 (0.14)	0.22 (0.15)	0.04 (0.04)

[1] Cows that carried BIA, BBB, and BA1E haplotypes were excluded from the analysis due to low numbers of observations.

[2] The effects of haplotypes are estimated by modeling the haplotypes as fixed effects, and the estimate is the effect of having 1 copy of that haplotype.

[3] The proportion of additive genetic variance explained by haplotypes, where the haplotypes are modeled as random effects.

* *P* <0.05 ** *P* <0.01 *** *P* <0.001

Table 3.9. Effects of α_{s1}-β-κ-CN haplotypes and proportion of additive genetic variance explained by haplotypes (h_{haplo}) for individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and phosphorylation degrees (PD)[1] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows[2] (SE in parentheses)

α_{s1}-β-κ-CN haplotype[3]	α_{s1}-CN			α_{s2}-CN					
	8P	9P	PD	10P	11P	12P	13P	14P	PD
CA2A (n = 31)	3.39 (0.18)	0.69 (0.12)	-0.70 (0.34)	0.10 (0.04)	-0.15 (0.06)	0.07 (0.04)	-0.13 (0.04)	-0.05 (0.02)	-2.60 (0.96)
BA2A (n = 34)	2.30 (0.20)	0.58 (0.13)	-0.28 (0.36)	0.10 (0.04)	-0.13 (0.07)	0.20 (0.04)	-0.11 (0.04)	-0.06 (0.02)	-1.82 (1.03)
BA1A (n = 26)	1.78 (0.21)	0.22 (0.14)	-0.76 (0.39)	0.08 (0.04)	0.01 (0.07)	-0.05 (0.05)	-0.10 (0.04)	-0.05 (0.02)	-2.30 (1.12)
BBA (n = 178)	0	0	0	0	0	0	0	0	0
BA1B (n = 22)	0.72 (0.23)	0.00 (0.15)	-0.47 (0.44)	0.01 (0.05)	-0.01 (0.08)	0.02 (0.05)	-0.15 (0.05)	-0.06 (0.02)	0.91 (1.26)
BA2B (n = 141)	1.60 (0.11)	0.32 (0.07)	-0.37 (0.21)	-0.00 (0.02)	0.08 (0.04)	0.14 (0.03)	0.00 (0.02)	-0.02 (0.01)	-1.72 (0.60)
BIB (n = 94)	0.96 (0.13)	-0.06 (0.09)	-0.83 (0.25)	0.12 (0.03)	0.37 (0.05)	0.10 (0.03)	-0.02 (0.03)	-0.03 (0.01)	-5.28 (0.71)
$-\text{Log}_{10}(P)$	64.9^{***}	9.5^{***}	1.6^{*}	4.9^{***}	14.6^{***}	7.7^{***}	4.1^{***}	3.1^{*}	9.9^{***}
h_{haplo}[4]	0.72 (0.17)	0.20 (0.12)	0.04 (0.04)	0.39 (0.43)	0.36 (0.19)	0.38 (0.33)	0.21 (0.19)	0.12 (0.12)	0.32 (0.23)

[1] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

[2] One cow with A1A1 genotype was excluded from the analysis.

[3] The number below the β-CN genotype indicates the number of alleles (0, 1, or 2) that a cow carried.

* P <0.05 ** P <0.01 *** P <0.001

3.4 Discussion

We report heritability estimates for detailed milk protein composition and explored factors affecting detailed milk protein composition, especially for α_s-CN phosphorylation isoforms and the phosphorylation degree of α_s-CN (α_s-CN PD). Accurate quantification of relative concentrations of α_{s2}-CN phosphorylation isoforms has been a challenge due to limitations of the analytical techniques. As a result, limited information was available about factors contributing to the variation in relative concentrations of α_{s2}-CN phosphorylation isoforms and the phosphorylation degree of α_{s2}-CN.

3.4.1 Heritability estimates

To our knowledge, this is the first study to report heritability estimates for relative concentrations of individual α_{s2}-CN phosphorylation isoforms (0.07-0.32) and for α_{s2}-CN PD (0.23). Information about the 6 major milk proteins is scarce as well because their quantification is time consuming and costly. A few studies have reported heritability estimates for relative concentrations of the 6 major milk proteins (Ikonen et al., 1997; Schopen et al., 2009; Boichard et al., 2014; Buitenhuis et al., 2016), and only 2 studies have reported heritability estimates for α_{s1}-CN-8P and α_{s1}-CN-9P concentrations (Bijl et al., 2014a; Buitenhuis et al., 2016). In our study, we did not adjust for the herd effect because most of the cows were located in different herds. Consequently, we could not estimate the proportion of phenotypic variance due to differences between herds, so heritability estimates of this study were expected to be lower than the intra-herd heritability estimates reported in previous studies. However, differences between herds contribute only a relatively small part of phenotypic variances of the 6 major milk proteins, α_{s1}-CN-8P and α_{s1}-CN-9P concentrations according to previous studies (Schopen et al., 2009; Bonfatti et al., 2011; Bijl et al., 2014a).

The heritability estimates for relative concentrations of the 6 major milk proteins were similar to or higher than those reported by Boichard et al. (2014), who analyzed major protein fractions predicted by mid-infrared spectra in the Montbéliarde breed. In comparison with other breeds, the heritability estimates for α_{s1}-CN (1.00) and β-CN (0.75) in our study were higher than those reported by Schopen et al. (2009, 0.47 for α_{s1}-CN and 0.25 for β-CN in Dutch Holstein-Friesian) and by Buitenhuis et al. (2016, 0 for α_{s1}-CN and 0.05 for β-CN in Danish Holstein; 0.05 for α_{s1}-CN and 0.29 for β-CN in Danish Jersey). The heritability estimate for κ-CN (0.62) in our study is similar to those reported by Schopen et al. (2009, p. 0.64) and Buitenhuis et al. (2016, 0.77 for Danish Holstein) but higher than the one for Danish Jersey (0.29; Buitenhuis et al., 2016). The heritability estimate for β-LG (0.73) in our study is similar to the one reported by Schopen et al. (2009, p. 0.80) but higher than those reported by Buitenhuis et al. (2016, 0.58 for Danish Holstein and 0.16 for Danish Jersey). As shown in previous studies (Bobe et al., 1999; Heck et al., 2009; Schopen et al., 2009), the α_{s1}-CN, κ-CN, β-CN, and β-LG genotypes contributed a major part of the genetic variation in α_{s1}-CN, κ-CN, β-CN, and β-LG concentration, respectively.

For α_{s1}-CN phosphorylation profile, the heritability estimate for α_{s1}-CN-8P concentration

(0.84) in our study is higher than those reported by Bijl et al. (2014a, 0.48 in Dutch Holstein-Friesian) and Buitenhuis et al. (2016, 0.01 in Danish Holstein and 0.41 in Danish Jersey). The heritability estimate for α_{s1}-CN-9P concentration (0.56) in our study is within the range of the one reported by Bijl et al. (0.76 2014a) and is higher than those reported by Buitenhuis et al. (2016, 0.25 in Danish Holstein and 0.23 in Danish Jersey). Discrepancies between studies might be due to genetic differences between breeds, limited sample sizes in the study of Buitenhuis et al. (2016) , and in this study, and use of different analytical methods. In terms of analytical methods, Schopen et al. (2009) and Bijl et al. (2014a) measured protein fractions by capillary zone electrophoresis, so β-CN was measured together with glycosylated and multi-phosphorylated κ-CN, and κ-CN was only measured as the mono-phosphorylated isoform (Heck et al., 2008). Furthermore, protein fractions measured with the same analytical method, such as LC (as used by Buitenhuis et al. (2016)), could differ because of differences in separation conditions.

3.4.2 Effects of genetic variants and casein haplotypes on the 6 major milk proteins

We investigated the associations of genetic variants of milk proteins and casein haplotypes with detailed milk protein composition, and this is the first study of such association analysis done with the Montbéliarde breed. Our study confirms the effect of the κ-CN B variant on κ-CN concentration and the effect of the β-LG B variant on β-LG concentration, which seem to be the true effects as they are consistent across studies and breeds (Van Eenennaam and Medrano, 1991; Bobe et al., 1999; Graml and Pirchner, 2003; Hallén et al., 2008; Heck et al., 2009; Bonfatti et al., 2010). Discrepancies for other associations among studies could be due to reasons mentioned above (i.e., genetic differences between breeds, limited sample sizes in some studies, and use of different analytical methods). Regarding genetic differences between breeds, the casein loci are in close proximity, and linkage disequilibrium between genetic variants has been reported (Ikonen et al., 1999; Heck et al., 2009; Bonfatti et al., 2010). Differences in linkage disequilibrium between breeds will affect the associations of the genetic variants with detailed milk protein composition. Therefore, investigating the associations of casein haplotypes with detailed milk protein composition provides better insights into the associations of individual genetic variants.

α_{s1}-CN. Effects of α_{s1}-CN genotypes on milk protein composition are rarely reported due to the low frequency of the C variant in western cattle breeds such as Dutch Holstein-Frisian, Italian Simmental, Danish Holstein and Swedish Red (Heck et al., 2009; Bonfatti et al., 2010; Gustavsson et al., 2014). The C variant was associated with higher α_{s1}-CN concentration and lower β-CN concentration, which is in line with the results of Graml and Pirchner (2003). However, we did not detect the dominance effect reported by those authors. This could be due to the small number of homozygous CC cows in our population, genetic differences between breeds, or differences in the analytical methods.

β-CN. For the 6 major milk proteins, the associations of the B and A2 variants were in opposite directions, which agrees with associations reported by Visker et al. (2011). Positive as-

sociation of the B variant with β-CN concentration and negative association of the B variant with α_{s1}-CN and α_{s2}-CN concentrations are consistent with previous studies (Bonfatti et al., 2010; Visker et al., 2011). However, we found a negative association of the B variant with κ-CN concentration, whereas Visker et al. (2011) reported a positive association and Bonfatti et al. (2010) found no association. Positive association of the I variant with α_{s2}-CN concentration is in line with association reported by Visker et al. (2011). As for the effect of the I variant on α_{s1}-CN concentration, we found no association, whereas Bonfatti et al. (2010) and Visker et al. (2011) reported negative associations.

κ-CN. Besides a favorable effect of the κ-CN B variant on κ-CN concentration, we also report a favorable effect of the κ-CN B variant on α_{s2}-CN concentration, which is similar to what Heck et al. (2009) reported, whereas Bobe et al. (1999) found no effect of κ-CN genotypes on α_{s2}-CN concentration. Contrary to previous studies, we found a positive association of the B variant with α_{s1}-CN concentration and a negative association of the B variant with β-CN concentration, whereas other authors reported a negative association of the B variant with α_{s1}-CN concentration and no association with β-CN concentration (Bobe et al., 1999; Hallén et al., 2008; Heck et al., 2009; Bonfatti et al., 2010).

β-LG. We report that the β-LG genotypes accounted for 95% of genetic variation of β-LG concentration, which agrees with the results reported by Bobe et al. (1999) and Heck et al. (2009). These authors concluded that β-LG genotypes predominately regulate β-LG concentration in total milk protein. We show a favorable effect of the B variant on α_{s1}-CN, κ-CN, β-CN, and α-LA concentrations, and a negative effect on β-LG concentration. These findings agree with those from previous studies, as the B variant of β-LG decreases the proportion of β-LG which results in increased proportions of the other milk proteins (Bobe et al., 1999; Hallén et al., 2008; Heck et al., 2009; Bonfatti et al., 2010). Note that the effect of the β-LG BB genotype on β-LG concentration in this study was about 1.5 times larger than the one reported by Heck et al. (2009). One explanation for this difference might be that the observed genotype effect is the combination of the effect of β-LG genotypes and the effect of one or multiple linked causal mutations. Multiple genetic polymorphisms in the coding and promoter regions have been detected in the β-LG gene, and many of them are in linkage disequilibrium with the β-LG genotypes (Ganai et al., 2009). The effect of a linked causal mutation might vary across breeds and populations due to differences in linkage disequilibrium with the β-LG genotypes.

Casein haplotypes. The frequency and effects of the CA2A haplotype on detailed milk protein composition are reported for the first time in this study in French Montbéliarde cows due to the low frequency of the α_{s1}-CN C variant in other breeds. The β-CN B variant occurred predominantly with the κ-CN A variant in French Montbéliarde cows as reported for Simmental cows (Bonfatti et al., 2010), whereas the β-CN B variant occurred only with the κ-CN B variant in Dutch Holstein-Friesian cows (Visker et al., 2011).

Haplotype analysis confirmed positive association of the α_{s1}-CN C variant with α_{s1}-CN concentration, positive association of the κ-CN B variant with κ-CN concentration, and positive association of the β-CN I variant with α_{s2}-CN concentrations. These results are consistent across studies and breeds, which seem to be direct associations of the genetic variants or causal mutations that are closely linked to the genetic variants (Van Eenennaam and Medrano, 1991;

Bobe et al., 1999; Graml and Pirchner, 2003; Hallén et al., 2008; Heck et al., 2009; Bonfatti et al., 2010; Visker et al., 2011). We report a positive association of the β-CN I variant with α_{s1}-CN concentration, whereas Visker et al. (2011) reported a negative association. Due to its low frequency, the BIA haplotype was excluded from the association analysis. Hence, it is not straightforward to infer if associations of the BIB haplotype resulted from the β-CN I variant or from the κ-CN B variant. Positive associations of the BIB haplotype with κ-CN and α-LA concentrations most likely resulted from associations of the κ-CN B variant because associations of BA1B, BA2B, and BIB haplotypes with κ-CN and α-LA concentrations were in the same direction and of about similar magnitude.

3.4.3 Factors affecting phosphorylation of α_{s1}-CN and α_{s2}-CN

Effects of parity and lactation stage. We show changes of α_s-CN phosphorylation profile across parity and lactation. Cows in parity 3 or higher produced milk with higher degrees of phosphorylation of α_s-CN as relative concentrations of α_{s1}-CN-9P, α_{s2}-CN-13P, and α_{s2}-CN-14P increased. Furthermore, both α_{s1}-CN PD and α_{s2}-CN PD increased as lactation progressed, whereas the total α_{s1}-CN and α_{s2}-CN concentrations were not affected by lactation stage. Note that relative concentrations of the group of the isoforms with lower degrees of phosphorylation (α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P) changed in the opposite direction compared with those of the group of isoforms with higher degrees of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-13P, and α_{s2}-CN-14P) during lactation. This observation supports the hypothesis that different sets of genes regulate phosphorylation of α_s-CN (Bijl et al., 2014a; Fang et al., 2016). Such different sets of genes involved in phosphorylation of caseins could show different expression throughout lactation as the mammary transcriptome is known to vary at different lactation stages due to physiological changes such as pregnancy (Bionaz et al., 2012; Wickramasinghe et al., 2012). Moreover, Bijl et al. (2014a) detected association between the *DGAT1* gene and α_{s1}-CN-9P concentration, and Bovenhuis et al. (2015) showed that the magnitude of the *DGAT1* effect on milk production traits changes during lactation. The *DGAT1* A allele is almost fixed in the French Montbéliarde breed (0.96) (Gautier et al., 2007), so the K232A polymorphism in this gene is not responsible for the observed change during lactation in the French Montbéliarde. Nevertheless, the effects of other genes involved in the phosphorylation mechanism of caseins may show similar variation throughout lactation.

Effects of genetic variants and casein haplotypes. Phosphorylation of caseins occurs in the Golgi apparatus after the synthesis of polypeptide chains. If casein-phosphorylating enzymes are not saturated, the associations of genetic variants with individual phosphorylation isoforms can be ascribed to the associations of genetic variants with total α_s-CN concentrations. This is supported by the observation that the effects of genetic variants on α_{s1}-CN-8P, α_{s1}-CN-9P, and total α_{s1}-CN concentration were in the same direction and of similar magnitude. The same explanation applies to the associations of genetic variants with individual α_{s2}-CN phosphorylation isoforms. This is confirmed by estimating the effects of genetic variants on the phosphorylation degree of α_{s1}-CN and α_{s2}-CN. Only the negative effect of the β-CN I variant and the positive effects of the κ-CN A variant and the β-CN B variant seem to be direct effects

on α_{s1}-CN PD and α_{s2}-CN PD. These effects of the β-CN I and B variant are also confirmed by haplotype analysis as the BIB haplotype showed positive associations with α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P concentrations and a negative association with α_s-CN PD, and the BBA haplotype showed opposite associations. In contrast, Bijl et al. (2014a) reported negative associations of β-κ-CN haplotype IB with both α_{s1}-CN-8P and α_{s1}-CN-9P concentrations in Dutch Holstein-Friesian cows. Combining the negative association of β-κ-CN haplotype IB with α_{s1}-CN concentration reported by Visker et al. (2011), the associations of the IB haplotype with α_{s1}-CN-8P and α_{s1}-CN-9P concentrations could be ascribed to its association with total α_s-CN concentration as discussed above. Moreover, Bijl et al. (2014a) showed that β-LG genotypes affect only α_{s1}-CN-8P concentration, whereas we showed that β-LG genotypes affected both α_{s1}-CN-8P and α_{s1}-CN-9P concentrations. Inconsistencies between the 2 studies might be due to the reasons mentioned above (i.e., genetic differences between the 2 breeds and differences in linkage disequilibrium between the β-LG genotypes and other causal variants across breeds).

3.5 Conclusions

We report the difference in relative concentrations of α_s-CN phosphorylation isoforms and in the phosphorylation degree of α_s-CN between cows due to systematic environmental effects (parity and lactation stage) and genetic variation. We show that α_s-CN phosphorylation profile changed across parity and lactation, and exploitable genetic variation for the phosphorylation degrees of α_{s1}-CN and α_{s2}-CN exists in the French Montbéliarde cattle. Furthermore, the β-CN I variant is associated with a greater proportion of isoforms with lower degrees of phosphorylation (α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P); the β-CN B variant is associated with a greater proportion of isoforms with higher degrees of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-12P to α_{s2}-CN-14P). Currently, knowledge regarding the effects of the phosphorylation degree of α_s-CN on technological properties of milk is limited and requires further investigation.

Acknowledgments

The first author benefited from an Erasmus-Mundus fellowship and a grant by APIS-GENE (Paris, France), within the framework of the European Graduate School in Animal Breeding and Genetics (EGS-ABG, Paris, France). This study was funded by the French Ministry of Agriculture, Agro-food and Forest, the French Dairy Interbranch Organization (CNIEL, Paris, France), the Regional Union of Protected Designation cheeses of Franche-Comté (URFAC, Poligny, France), and the Regional Council of Bourgogne Franche-Comté (Besançon, France), under the project FROM'MIR.

References

Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk, and H. Bovenhuis (2014a). Phosphorylation of αS1-casein is regulated by different genes. *J. Dairy Sci.* 97.11, pp. 7240–7246.

Bijl, E., H. van Valenberg, S. Sikkes, S. Jumelet, G. Sala, K. Olieman, T. van Hooijdonk, and T. Huppertz (2014b). Chymosin-induced hydrolysis of caseins: Influence of degree of phosphorylation of alpha-s1-casein and genetic variants of beta-casein. *Int. Dairy J.* 39.2, pp. 215–221.

Bionaz, M., K. Periasamy, S. L. Rodriguez-Zas, R. E. Everts, H. A. Lewin, W. L. Hurley, and J. J. Loor (2012). Old and new stories: Revelations from functional analysis of the bovine mammary transcriptome during the lactation cycle. *PLoS One* 7.3, e33268.

Bobe, G., D. Beitz, A. Freeman, and G. Lindberg (1999). Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. *J. Dairy Sci.* 82.12, pp. 2797–2804.

Boichard, D., A. Govignon-Gion, H. Larroque, C. Maroteau, I. Palhière, G. Tosser-Klopp, R. Rupp, M.-P. Sanchez, and M. Brochard (2014). Déterminisme génétique de la composition en acides gras et protéines du lait des ruminants, et potentialités de sélection. *INRA Prod. Anim.* 27.4, pp. 283–298.

Bonfatti, V., A. Cecchinato, L. Gallo, A. Blasco, and P. Carnier (2011). Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. *J. Dairy Sci.* 94.10, pp. 5183–5193.

Bonfatti, V., G. Di Martino, A. Cecchinato, L. Degano, and P. Carnier (2010). Effects of β-κ-casein (CSN2-CSN3) haplotypes and β-lactoglobulin (BLG) genotypes on milk production traits and detailed protein composition of individual milk of Simmental cows. *J. Dairy Sci.* 93.8, pp. 3797–3808.

Bovenhuis, H., M. Visker, H. van Valenberg, A. Buitenhuis, and J. van Arendonk (2015). Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. *J. Dairy Sci.* 98.9, pp. 6572–6582.

Browning, S. R. and B. L. Browning (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81.5, pp. 1084–1097.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen (2016). Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genet.* 17.

Caroli, A. M., S. Chessa, and G. J. Erhardt (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92.11, pp. 5335–5352.

De Kruif, C. and C. Holt (2003). "Casein micelle structure, functions and interactions". In: *Adv. Dairy Chem. Proteins*. Ed. by P. F. Fox and P. L. H. McSweeney. 3rd. Vol. 1. JANUARY 2003. Springer US. Chap. 5, pp. 233–276.

Fang, Z. H., M. H. P. W. Visker, G. Miranda, A. Delacroix-Buchet, H. Bovenhuis, and P. Martin (2016). The relationships among bovine αS-casein phosphorylation isoforms suggest different phosphorylation pathways. *J. Dairy Sci.* 99.10, pp. 8168–8177.

Frederiksen, P. D., K. K. Andersen, M. Hammershøj, H. D. Poulsen, J. Sørensen, M. Bakman, K. B. Qvist, and L. B. Larsen (2011). Composition and effect of blending of noncoagulating, poorly coagulating, and well-coagulating bovine milk from individual Danish Holstein cows. *J. Dairy Sci.* 94.10, pp. 4787–4799.

Ganai, N. A., H. Bovenhuis, J. A. M. van Arendonk, and M. H. P. W. Visker (2009). Novel polymorphisms in the bovine beta-lactoglobulin gene and their effects on beta-lactoglobulin protein concentration in milk. *Anim. Genet.* 40.2, pp. 127–133.

Gautier, M., A. Capitan, S. Fritz, A. Eggen, D. Boichard, and T. Druet (2007). Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphisms in French dairy cattle. *J. Dairy Sci.* 90.6, pp. 2980–2988.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, and D. Butler (2009). ASReml user guide release 3.0. *VSN Int. Ltd, Hemel Hempstead, UK*.

Graml, R. and F. Pirchner (2003). Effects of milk protein loci on content of their proteins. *Arch. fur Tierzucht* 46, pp. 331–340.

Gustavsson, F., A. J. Buitenhuis, M. Johansson, H. P. Bertelsen, M. Glantz, N. A. Poulsen, H. L. Mansson, H. Stalhammar, L. B. Larsen, C. Bendixen, M. Paulsson, and A. Andren (2014). Effects of breed and casein genetic variants on protein profile in milk from Swedish Red, Danish Holstein, and Danish Jersey cows. *J. Dairy Sci.* 97.6, pp. 3866–3877.

Hallén, E., A Wedholm, A Andrén, A Lundén, A. Wedholm, A. Andrén, and A. Lundén (2008). Effect of beta-casein, kappa-casein and beta-lactoglobulin genotypes on concentration of milk protein variants. *J. Anim. Breed. Genet.* 125.2, pp. 119–129.

Heck, J. M. L., A. Schennink, H. J. F. Van Valenberg, H. Bovenhuis, M. H. P. W. Visker, J. A. M. Van Arendonk, and A. C. M. van Hooijdonk (2009). Effects of milk protein variants on the protein composition of bovine milk. *J. Dairy Sci.* 92.3, pp. 1192–1202.

Heck, J., C. Olieman, A. Schennink, H. van Valenberg, M. Visker, R. Meuldijk, and A. van Hooijdonk (2008). Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *Int. Dairy J.* 18.5, pp. 548–555.

Holland, J. W. and M. J. Boland (2014). "Post-translational modifications of caseins". In: *Milk proteins From Expr. to food*. Ed. by H. Singh, A. Thompson, and M. Boland. Second Edi. Elsevier Inc., pp. 141–168.

Ikonen, T., M. Ojala, and E.-L. Syväoja (1997). Effects of composite casein and beta-lactoglobulin genotypes on renneting properties and composition of bovine milk by assuming an animal model. *Agric. Food Sci.* 6.4, pp. 283–294.

Ikonen, T., M. Ojala, and O. Ruottinen (1999). Associations Between Milk Protein Polymorphism and First Lactation Milk Production Traits in Finnish Ayrshire Cows. *J. Dairy Sci.* 82.5, pp. 1026–1033.

Jensen, H. B., N. A. Poulsen, K. K. Andersen, M. Hammershøj, H. D. Poulsen, and L. B. Larsen (2012). Distinct composition of bovine milk from Jersey and Holstein-Friesian cows with good, poor, or noncoagulation properties as reflected in protein genetic variants and isoforms. *J. Dairy Sci.* 95.12, pp. 6905–6917.

Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. van Valenberg, and J. A. M. van Arendonk (2009). Genetic parameters for major milk proteins in Dutch Holstein-Friesians. *J. Dairy Sci.* 92.3, pp. 1182–1191.

Van Eenennaam, A. L. and J. F. Medrano (1991). Differences in Allelic Protein Expression in the Milk of Heterozygous κ-Casein Cows. *J. Dairy Sci.* 74.5, pp. 1491–1496.

Visker, M. H. P. W., B. W. Dibbits, S. M. Kinders, H. J. F. van Valenberg, J. A. M. van Arendonk, and H. Bovenhuis (2011). Association of bovine β-casein protein variant I with milk production and milk protein composition. *Anim. Genet.* 42.2, pp. 212–218.

Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano (2012). Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* 13.1, p. 45.

Supplementary materials

Table S3.1. Effects of β-CN genotypes on the 6 major milk proteins measured on test-day morning milk samples from 531 Montbéliarde cows[1] (SE in brackets)

β-CN genotype	α_{s1}-CN	α_{s2}-CN	κ-CN	β-CN	α-LA	β-LG
BA2 (n=169)	0	0	0	0	0	0
A1A2 (n= 38)	0.91 (0.29)	-0.19 (0.14)	0.06 (0.15)	-0.10 (0.30)	0.09 (0.10)	0.14 (0.34)
BA1 (n=38)	-0.70 (0.27)	-0.82 (0.14)	-0.12 (0.14)	3.00 (0.29)	-0.24 (0.10)	-0.97 (0.31)
A1I (n= 17)	-0.04 (0.38)	0.14 (0.19)	0.14 (0.20)	0.50 (0.40)	-0.04 (0.14)	-0.15 (0.43)
A2A2 (n=72)	2.67 (0.22)	0.04 (0.11)	0.40 (0.11)	-3.66 (0.23)	0.32 (0.08)	0.07 (0.25)
A2I (n=72)	1.09 (0.21)	0.65 (0.11)	0.30 (0.11)	-3.18 (0.23)	0.13 (0.08)	0.60 (0.25)
BB (n=50)	-2.21 (0.24)	-0.64 (0.12)	-0.65 (0.13)	2.87 (0.26)	-0.25 (0.09)	0.29 (0.28)
BI (n=64)	-1.28 (0.23)	0.34 (0.11)	0.32 (0.12)	0.39 (0.24)	0.10 (0.08)	0.10 (0.26)
II (n=10)	-0.55 (0.48)	1.03 (0.24)	0.38 (0.25)	-2.47 (0.51)	0.04 (0.18)	0.99 (0.56)
$-\text{Log}_{10}(P)$	55.4^{***}	25.8^{***}	10.5^{***}	106.1^{***}	6.0^{***}	2.6^{**}

[1] One cow with A1A1 genotype was excluded from the analysis.

* P <0.05 ** P <0.01 *** P <0.001

Table S3.2. Effects of β-CN genotypes on individual phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN, and phosphorylation degrees (PD)[1] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 531 Montbéliarde cows[2] (SE in brackets)

β-CN genotype	α_{s1}-CN			α_{s2}-CN					
	8P	9P	PD	10P	11P	12P	13P	14P	PD
BA2 (n=169)	0	0	0	0	0	0	0	0	0
A1A2(n= 38)	0.91 (0.28)	0.03 (0.17)	-0.57 (0.47)	0.07 (0.05)	-0.02 (0.09)	0.01 (0.06)	-0.18 (0.05)	-0.07 (0.02)	-1.71 (1.34)
BA1 (n=38)	-0.42 (0.26)	-0.27 (0.15)	-0.32 (0.44)	0.03 (0.05)	-0.03 (0.08)	-0.12 (0.05)	-0.09 (0.05)	-0.03 (0.02)	3.47 (1.26)
A1I (n= 17)	0.32 (0.37)	-0.36 (0.22)	-1.01 (0.62)	0.10 (0.07)	0.31 (0.12)	-0.13 (0.08)	-0.04 (0.07)	-0.02 (0.03)	-3.72 (1.79)
A2A2 (n=72)	2.29 (0.21)	0.40 (0.12)	-0.63 (0.35)	0.04 (0.04)	-0.07 (0.07)	0.15 (0.04)	-0.09 (0.04)	-0.05 (0.02)	-0.45 (1.01)
A2I (n=72)	1.08 (0.21)	0.00 (0.12)	-0.75 (0.35)	0.09 (0.04)	0.31 (0.07)	0.14 (0.04)	-0.04 (0.04)	-0.04 (0.02)	-3.75 (1.01)
BB (n=50)	-1.81 (0.24)	-0.41 (0.14)	0.36 (0.40)	-0.06 (0.05)	-0.12 (0.08)	-0.07 (0.05)	-0.01 (0.04)	0.01 (0.02)	4.03 (1.15)
BI (n=64)	-0.78 (0.22)	-0.52 (0.13)	-0.67 (0.37)	0.14 (0.04)	0.32 (0.07)	-0.08 (0.04)	-0.07 (0.04)	-0.03 (0.02)	-4.40 (1.06)
II (n=10)	-0.05 (0.47)	-0.52 (0.28)	-1.23 (0.79)	0.19 (0.09)	0.60 (0.15)	0.24 (0.10)	0.00 (0.09)	-0.03 (0.04)	-4.28 (2.29)
-$\text{Log}_{10}(P)$	43.1^{***}	6.5^{***}	1.1^{NS}	2.7^{**}	10.5^{***}	7.7^{***}	1.4^{*}	1.8^{*}	10.9^{***}

[1] α_{s1}-CN PD= (α_{s1}-CN-9P/total α_{s1}-CN) × 100; α_{s2}-CN PD = [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P)/total α_{s2}-CN] × 100.

[2] One cow with A1A1 genotype was excluded from the analysis.

* P <0.05 ** P <0.01 *** P <0.001

Table S3.3. Frequencies of α_{s1}-β-κ-CN haplotypes from 531 Montbéliarde cows

α_{s1}-β-κ-CN haplotype	Frequency
CA2A	0.063
BA2A	0.062
BA1A	0.050
BBA	0.346
BA1B	0.037
BA2B	0.272
BIB	0.160
BBB	0.003
BIA	0.003
BA1E	0.003

Chapter 4

Genetic parameters for α_{s1}- and α_{s2}-casein phosphorylation isoforms in Dutch Holstein Friesian

Z. H. Fang*†, H. Bovenhuis†, H. J. F. van Valenberg,‡, P.Martin*, T. Huppertz§ and M. H. P. W. Visker†

*Génétique Animale et Biologie Intégrative, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
†Animal Breeding and Genomics, Wageningen University & Research, PO Box 338, 6700 AH Wageningen, the Netherlands
‡ Dairy Science and Technology Group, Wageningen University & Research, PO Box 17, 6700 AA, Wageningen, the Netherlands
§ NIZO, PO Box 20, 6710 BA, Ede, the Netherlands

Journal of Dairy Science (Accepted)

Abstract

Relative concentrations of α_{s1}- and α_{s2}-casein (α_{s1}- and α_{s2}-CN) phosphorylation isoforms vary considerably among milk of individual cows. We estimated heritabilities for α_{s2}-CN phosphorylation isoforms, determined by capillary zone electrophoresis from 1857 morning milk samples, and genetic correlations among α_{s2}-CN phosphorylation isoforms in Dutch Holstein Friesian. To investigate if phosphorylation of α_{s1}- and α_{s2}-CN are due to the same genetic mechanism, we also estimated genetic correlations between α_{s1}- and α_{s2}-CN phosphorylation isoforms as well as the genetic correlations between the phosphorylation degrees (PD) of α_{s1}- and α_{s2}-CN defined as the proportion of isoforms with higher degrees of phosphorylation in total α_{s1}- and α_{s2}-CN, respectively. The intra-herd heritabilities for the relative concentrations of α_{s2}-CN phosphorylation isoforms were high and ranged from 0.54 for α_{s2}-CN-10P to 0.89 for α_{s2}-CN-12P. Furthermore, the high intra-herd heritabilities of α_{s1}-CN PD and α_{s2}-CN PD imply a strong genetic control of the phosphorylation process, which is independent of casein production. The genetic correlations between α_{s2}-CN phosphorylation isoforms are positive and moderate to high (0.33-0.90). Furthermore, the strong positive genetic correlation (0.94) between α_{s1}-CN PD and α_{s2}-CN PD suggests that the phosphorylation processes of α_{s1}- and α_{s2}-CN are related. This study shows the possibility to breed for specific α_{s1}- and α_{s2}-CN phosphorylation isoforms, and relations between the phosphorylation degrees of α_{s1}- and α_{s2}-CN and technological properties of milk need to be further investigated to identify potential benefits for the dairy industry.

Key words: milk protein composition, genetic correlation, posttranslational modification, capillary zone electrophoresis

4.1 Introduction

Phosphorylation of caseins is a crucial post-translational modification that allows caseins to interact with calcium phosphate in the assembly of colloidal structures known as casein micelles. The formation and stability of micelles depend largely on the presence of κ-CN on the micelle surface and its glycosylation, as well as the interactions between calcium phosphate nanoclusters and phosphoserine residues of α_{s1}-, α_{s2}-, and β-CN(De Kruif and Holt, 2003; De Kruif et al., 2012). Although caseins are all phosphoproteins, α_{s1}- and α_{s2}-CN are more highly phosphorylated and possess multiple phosphoserine clusters, whereas β-CN possesses only one, and κ-CN does not possess any. This suggests that α_{s1}- and α_{s2}-CN might be more relevant for stabilizing internal micellar structure (Dalgleish and Corredig, 2012; Huppertz et al., 2017).

In bovine milk, α_{s1}-CN accounts for about 35% of the total casein and has 2 common phosphorylation isoforms, i.e. α_{s1}-CN-8P as the predominant form and α_{s1}-CN-9P (Holland and Boland, 2014, where P indicates phosphate group attached;); α_{s2}-CN accounts for about 10% of the total casein and is present with various degrees of phosphorylation ranging from 10P to 14P and occasionally with 9P or 15P (Fang et al., 2016). Milk containing a greater proportion of α_{s1}- and α_{s2}-CN isoforms with lower degrees of phosphorylation (e.g. α_{s1}-CN-8P, α_{s2}-CN-10P and α_{s2}-CN -11P) has been suggested to possess better coagulation properties and to be hydrolyzed more efficiently by chymosin (Frederiksen et al., 2011; Jensen et al., 2012; Bijl et al., 2014b). Furthermore, relative concentrations of α_{s1}- and α_{s2}-CN phosphorylation isoforms in milk vary considerably between individual cows (Bijl et al., 2014a; Fang et al., 2016). Therefore, it is of great interest to investigate if genetic variation in individual α_{s1}- and α_{s2}-CN phosphorylation isoforms exists.

To date, only few studies reported heritabilities for relative concentrations of α_{s1}-CN-8P and α_{s1}-CN-9P; in Dutch Holstein Friesian (Bijl et al., 2014a), Danish Holstein and Jersey (Buitenhuis et al., 2016), and French Montbéliarde (Fang et al., 2017). Only one study in French Montbéliarde reported the hertabilities for relative concentrations of α_{s2}-CN-10P to α_{s2}-CN-14P (Fang et al., 2017). Additionally, Bijl et al. (2014a) reported a surprisingly low genetic correlation between α_{s1}-CN-8P and α_{s1}-CN-9P concentration. Heck et al. (2008) and Fang et al. (2016) reported that the proportion of α_{s1}- and α_{s2}-CN isoforms with lower degrees of phosphorylation phenotypically correlated negatively with the proportion of α_{s1}- and α_{s2}-CN isoforms with higher degrees of phosphorylation. However, genetic correlations among α_{s2}-CN phosphorylation isoforms as well as genetic correlations between α_{s1}- and α_{s2}-CN isoforms have not been reported.

Within the framework of the Dutch milk genomic initiative, detailed milk protein composition was determined by capillary zone electrophoresis (**CZE**) for approximately 2,000 cows. However, peaks assigned to α_{s2}-CN-10P and α_{s2}-CN-12P were based on predicted migration time without the use of standards of individual isoforms for confirmation, and α_{s2}-CN-13P and α_{s2}-CN-14P were not identified (Heck et al., 2008). The aim of this study was first to identify the peaks corresponding to individual α_{s2}-CN phosphorylation isoforms on CZE electropherograms for quantifying isoforms in the Dutch milk genomics initiative population, and second to estimate heritabilities for α_{s2}-CN phosphorylation isoforms and genetic correlations among

α_{s2}-CN phosphorylation isoforms. To investigate if phosphorylation of α_{s1}- and α_{s2}-CN are due to the same genetic mechanism, we also estimated genetic correlations between α_{s1}- and α_{s2}-CN phosphorylation isoforms as well as the genetic correlations between the phosphorylation degrees (**PD**) of α_{s1}- and α_{s2}-CN defined as the proportion of isoforms with higher degrees of phosphorylation in total α_{s1}- and α_{s2}-CN, respectively. We also investigated the relations of α_{s1}- and α_{s2}-CN phosphorylation isoforms with milk protein composition, phosphorus content and with routinely collected milk production traits.

4.2 Material and Methods

4.2.1 Animals

Test-day morning milk samples were collected from approximately 2,000 primiparous Dutch Holstein Friesian cows as part of the Dutch Milk Genomic Initiative. Cows were located on 398 herds in the Netherlands, and at least 3 cows per herd were sampled. The pedigree of the cows was supplied by cattle improvement organization CRV (Arnhem, the Netherlands). Detailed description of the experimental design is provided in Schopen et al. (2009). In this study, we analyzed milk production traits and milk protein composition from 1,857 milk samples collected in winter.

4.2.2 Phenotypes

Milk production traits. Protein, fat, and lactose percentages were determined by infrared spectroscopy using a MilkoScan FT 6000 (Foss Electric, Hillerød, Denmark) at the milk control station laboratory (Qlip, Zutphen, the Netherlands). phosphorus concentration was determined by inductively coupled plasma-atomic emission spectrometry (Vista Axial, Varian, Australia) of whole milk as described in (van Hulzen et al., 2009). Test-day morning milk yield was available for 1,721 cows and was obtained from CRV. Yields of protein, fat, lactose and phosphorus were calculated by multiplying the respective content traits by the observed milk yield.

Milk protein composition. The relative concentrations (% wt/wt) of individual milk proteins and their isoforms were determined by CZE by Heck et al. (2008). The κ-CN was measured as only non-glycosylated and mono-phosphorylated form. The α_{s1}-CN phosphorylation isoforms were identified by using αs-CN standards as described by Heck et al. (2008). The relative concentration of total α_{s1}-CN was the sum of relative concentrations of α_{s1}-CN-8P and α_{s1}-CN -9P and also the peak that migrates just before the α_{s1}-CN-8P peak and is predicted to be α_{s1}-CN-7P according to Heck et al. (2008). Standards for the individual isoforms were not available for α_{s2}-CN isoforms. Therefore, to identify peaks corresponding to individual α_{s2}-CN phosphorylation isoforms on electropherograms from CZE analysis, we purified α_{s2}-CN phosphorylation isoforms from milk samples. For this purpose, 10 fresh milk samples were analyzed by liquid chromatography/electrospray ionization-mass spectrometry (**LC/ESI-MS**) for screening their α_{s2}-CN phosphorylation profiles using the method of Fang et al. (2016). Fractions containing individual α_{s2}-CN phosphorylation isoforms were separated and collected by LC with

the same method using milk that contained highest concentrations of α_{s2}-CN-10P to α_{s2}-CN-14P, respectively. Collected fractions were analyzed by LC-ESI/MS to verify the components. The fractions were then analyzed by CZE using the method of Heck et al. (2008) to confirm the peaks corresponding to individual α_{s2}-CN phosphorylation isoforms in CZE electropherograms. Yields (g) of individual milk proteins and their isoforms were calculated by multiplying relative concentrations (% wt/wt) by protein yield (g).

Relative concentrations of the α_{s1}- and α_{s2}-CN phosphorylation isoforms are the result of two distinct processes: the production of α_{s1}- and α_{s2}-CN, and the posttranslational modification of both caseins. To characterize the phosphorylation process that is independent of casein production, we defined the phosphorylation degrees (PD) of α_{s1}- and α_{s2}-CN as the proportion of isoforms with higher degrees of phosphorylation, which were calculated as

$$\frac{\alpha_{s1}\text{—CN-9P}}{\alpha_{s1}\text{—CN-8P}+\alpha_{s1}\text{—CN-9P}} \times 100 \tag{4.1}$$

$$\frac{\alpha_{s2}\text{—CN-12P}}{\alpha_{s2}\text{—CN-10P}+\alpha_{s2}\text{—CN-11P}+\alpha_{s2}\text{—CN-12P}} \times 100 \tag{4.2}$$

The α_{s2}-CN-13P and α_{s2}-CN-14P were not included in 4.2 due to low reproducibility of their quantification.

Output of casein phosphorus. To investigate if the phosphorus availability would be a limiting factor for phosphorylation of all caseins when more caseins are produced, we also estimated the phenotypic and genetic correlations between the total amount of phosphate groups attached to the caseins in milk (i.e. output of casein phosphorus, **Pcas**), and the milk protein and milk phosphorus yields. To derive Pcas, we first calculated the molar concentration ($\mathbf{C}_{molar}$) of each casein (α_{s1}-, α_{s2}-, κ- and β-CN) as its concentration in milk [g/L; as protein percentage × 10 × (% wt/wt)] divided by its respective molecular weight (Da). As κ-CN carried 1 phosphate group, and β-CN carried 5 in our milk samples, C_{molar} of phosphate groups attached to κ-CN and β-CN were approximated by multiplying C_{molar} of κ-CN and β-CN by 1 and 5, respectively. The C_{molar} of phosphate groups attached to α_{s1}-CN was the sum of C_{molar} of α_{s1}-CN-8P multiplied by 8 and C_{molar} of α_{s1}-CN-9P multiplied by 9. The C_{molar} of phosphate groups attached to α_{s2}-CN was the sum of C_{molar} of α_{s2}-CN-10P multiplied by 10, C_{molar} of α_{s2}-CN-11P multiplied by 11, and C_{molar} of α_{s2}-CN-12P multiplied by 12. Therefore, the total C_{molar} of the phosphate groups attached to the caseins was the sum of C_{molar} of the phosphate groups attached to the individual caseins. Subsequently, Pcas was approximated by multiplying the total C_{molar} of phosphate groups attached to the caseins by milk yield.

4.2.3 Statistical analyses

To estimate variance components and genetic parameters, the following animal model was used:

$$\begin{aligned} y_{klmn} = {} & \mu + \beta_1 \text{dim}_{klmn} + \beta_2 \text{e}^{-0.05*\text{dim}_{klmn}} + \beta_3 \text{ca}_{klmn} + \beta_4 \text{ca}^2_{klmn} + \text{season}_k \\ & + \text{scode}_l + \text{animal}_m + \text{herd}_n + \text{e}_{klmn}; \end{aligned} \tag{4.3}$$

where y_{klmn} is the observation of the trait of interest; μ is the overall mean of the trait; dim_{klmn} is a covariate describing the effect of days in lactation, modeled with a Wilmink curve (Wilmink, 1987); ca_{klmn} is a covariate describing the effect of age at first calving; $season_k$ is the fixed effect with 3 classes of calving season (June-August 2004, September-November 2004, and December 2004-February 2005); $scode_l$ is the fixed effect accounting for possible differences in genetic level between proven bull daughters and young bull daughters; $animal_m$ is the random additive genetic effect assumed to be distributed as $N(\mathbf{0},\mathbf{A}\sigma_a^2)$, where **A** is the additive genetic relationships matrix consisting of 26,300 animals, and σ_a^2 is the additive genetic variance; $herd_n$ is the random herd effect assumed to be distributed as $N(\mathbf{0},\mathbf{I}\sigma_{herd}^2)$, where I is the identity matrix, and σ_{herd}^2 is the herd variance; e_{klmn} is the random residual effect assumed to be distributed as $N(\mathbf{0},\mathbf{I}\sigma_e^2)$, where I is the identity matrix, and σ_e^2 is the residual variance. The intra-herd heritability was defined as

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2},$$

The proportion of phenotypic variance explained by herd was defined as

$$h_{herd} = \frac{\sigma_{herd}^2}{\sigma_{herd}^2 + \sigma_a^2 + \sigma_e^2},$$

Phenotypic and genetic correlations between traits were estimated by bivariate analyses using model 4.3. All statistical analyses were performed using ASReml 4.1 (Gilmour et al., 2015).

4.3 Results and Discussion

4.3.1 Characterization of α_{s2}-CN phosphorylation profile by CZE

In this study, we first characterized the α_{s2}-CN phosphorylation profile in CZE electropherograms. To identify the peaks corresponding to individual α_{s2}-CN phosphorylation isoforms in a milk sample on the CZE electropherograms, the CZE electropherograms of the purified and identified fractions of individual α_{s2}-CN phosphorylation isoforms were aligned with the CZE electropherogram of a milk sample (Figure 4.1). The peaks corresponding to α_{s2}-CN-10P, -11P and -12P matched the migration times predicted by Heck et al. (2008). Therefore, the reproducibilities of their measurements, calculated as the coefficient of variation from repeated measurements of a reference milk sample, were 15.7% for α_{s2}-CN-10P, 5.7% for α_{s2}-CN-11P and 5.8 for α_{s2}-CN-12P (Heck et al., 2008). The peak between the α_{s2}-CN-11P peak and the α_{s2}-CN-12P peak was suggested to be products of casein hydrolysis by Recio et al. (1997). The peak corresponding to α_{s2}-CN-13P was difficult to quantify accurately using CZE due to co-migration with other components in milk, which resulted in poor reproducibility of its quantification. The peak corresponding to α_{s2}-CN-14P overlapped with the α_{s1}-CN-8P peak. Therefore, we only estimated the relative concentrations of α_{s2}-CN-10P, -11P, and -12P based on the applied CZE method. Furthermore, α_{s2}-CN-14P accounts for about 0.4% of total protein in milk

according to Fang et al. (2016). Therefore, we do not expect the measurement of α_{s1}-CN-8P to be strongly affected by the overlap with α_{s2}-CN-14P. Relative concentrations of α_{s1}-CN-8P and -9P had been determined previously based on the same CZE electropherograms (Heck et al., 2008).

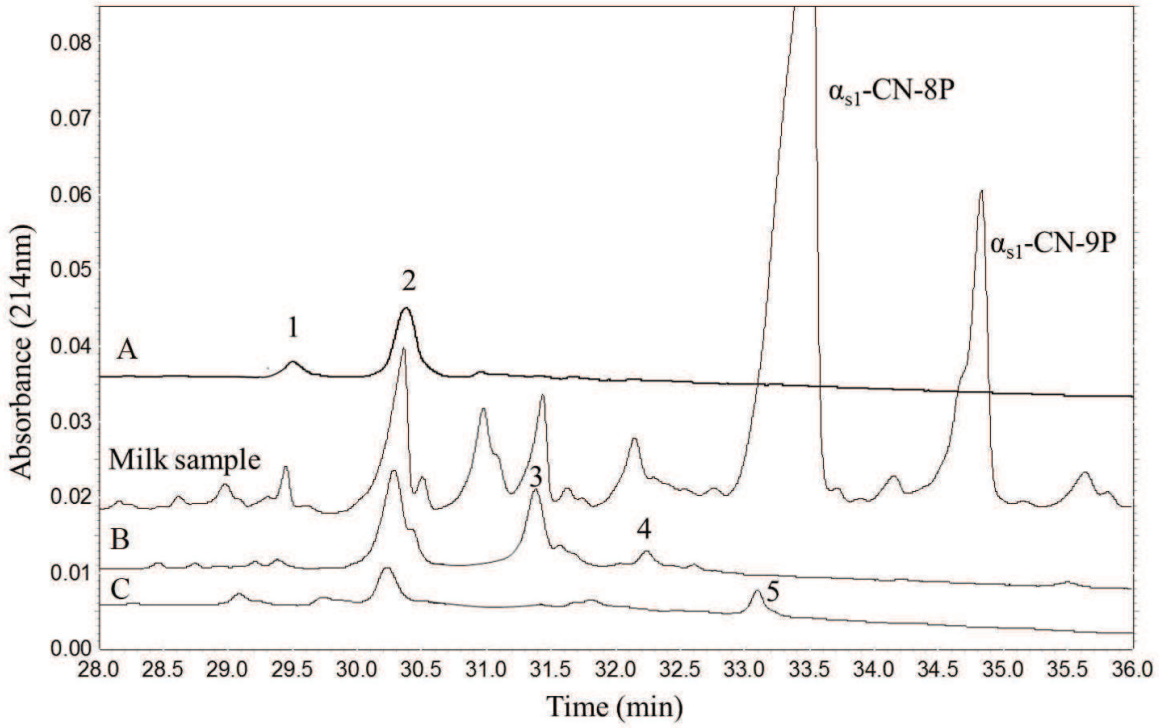

Figure 4.1. Capillary zone electropherograms of one milk sample and three purified fractions of α_{s2}-CN phosphorylation isoforms showing α_{s1}- and α_{s2}-CN phosphorylation isoforms. (A) Purified fraction containing α_{s2}-CN-10P (peak 1) and α_{s2}-CN-11P (peak 2). (B) Purified fraction containing α_{s2}-CN-11P, α_{s2}-CN-12P (peak 3) and α_{s2}-CN-13P (peak 4). (C) Purified fraction containing α_{s2}-CN-11P and α_{s2}-CN-14P (peak 5).

4.3.2 Genetic parameters

Table 4.1 shows the descriptive statistics, heritabilities and proportions of variance explained by herd for α_{s1}- and α_{s2}-CN phosphorylation isoforms and for the phosphorylation degrees of α_{s1}-CN (α_{s1}-CN PD) and α_{s2}-CN (α_{s2}-CN PD). To disentangle the casein production and the phosphorylation process, we defined the phosphorylation degree as the proportion of isoforms with higher degrees of phosphorylation in the sum of α_{s1}-CN-8P and α_{s1}-CN-9P and in the sum of α_{s2}-CN-10P, α_{s2}-CN-11P and α_{s2}-CN-12P, respectively. The most abundant α_{s2}-CN phosphorylation isoform was α_{s2}-CN-11P. The average α_{s1}-CN PD was around 26%, meaning that 26% of α_{s1}-CN is phosphorylated at the highest level (i.e. 9P), and the average α_{s2}-CN PD was around 34%, meaning that 34% of α_{s2}-CN is phosphorylated at the highest level among the detected α_{s2}-CN isoforms (i.e. 12P in the current study). The average relative concentrations of α_{s1}-CN-8P, -9P, α_{s2}-CN-10P, -11P and -12P, and the average α_{s1}-CN PD are in the range of the values

reported for French Montbéliarde (Fang et al., 2017). The α_{s2}-CN-13P and -14P were not included in the current study, but they account for about 24% of the total α_{s2}-CN concentration in French Montbéliarde (Fang et al., 2016). Because α_{s2}-CN-13P and -14P were not included, α_{s2}-CN PD is defined differently in the current study and therefore cannot be compared directly with results for French Montbéliarde (Fang et al., 2016). The average α_{s2}-CN PD is similar in both studies when using the definition of the current study to calculate the α_{s2}-CN PD in French Montbéliarde.

Table 4.1. Mean, standard deviation (SD), phenotypic variance[a] (σ_p^2), intra-herd heritability estimates (h^2), and proportion of phenotypic variance explained by herd (h_{herd}) for relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and for the phosphorylation degrees (PD)[b] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses)

Trait (% wt/wt)	Mean	SD	σ_p^2	h^2	h_{herd}
α_{s1}-CN[c]	33.64	1.66	2.80	0.52 (0.11)	0.11 (0.02)
α_{s1}-CN-8P[c]	21.26	1.13	1.32	0.48 (0.10)	0.12 (0.02)
α_{s1}-CN-9P[c]	7.42	1.07	1.18	0.76 (0.12)	0.08 (0.02)
α_{s2}-CN	6.67	0.95	0.98	0.94 (0.12)	0.08 (0.02)
α_{s2}-CN-10P	0.99	0.39	0.16	0.54 (0.11)	0.10 (0.02)
α_{s2}-CN-11P	3.44	0.57	0.33	0.89 (0.12)	0.08 (0.02)
α_{s2}-CN-12P	2.24	0.22	0.05	0.71 (0.12)	0.07 (0.02)
Phosphorylation degree					
α_{s1}-CN PD[b]	25.79	2.72	7.66	0.78 (0.12)	0.08 (0.02)
α_{s2}-CN PD[b]	34.01	4.24	18.18	0.64 (0.11)	0.09 (0.02)

[a] Phenotypic variance after adjusting for systematic effects.
[b] α_{s1}-CN PD= α_{s1}-CN-9P/ (α_{s1}-CN-8P+α_{s1}-CN-9P) × 100%; α_{s2}-CN PD = [α_{s2}-CN-12P/(α_{s2}-CN-10P + α_{s2}-CN-11P+α_{s2}-CN-12P)] × 100%.
[c] Adopted from Bijl et al. (2014a).

The intra-herd heritabilities were high to very high for relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms (0.48 to 0.89), for α_{s1}-CN PD (0.78) and for α_{s2}-CN PD (0.64). Differences between herds contributed approximately 10% to the phenotypic variation in relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD. Our results suggest that housing, management and nutritional differences between farms have little effect on α_{s1}- and α_{s2}-CN phosphorylation under Dutch farming conditions. The high intra-herd heritabilities of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms suggest an important role of genetic factors in Dutch Holstein Friesian. Furthermore, the high intra-herd heritabilities of α_{s1}-CN (0.52), α_{s2}-CN (0.94), α_{s1}-CN PD, and α_{s2}-CN PD imply a strong genetic control not only of the α_{s1}- and α_{s2}-CN synthesis but also of the phosphorylation process. It should be noted that we reported lower α_{s2}-CN concentration and higher intra-herd heritability of α_{s2}-CN compared to values reported by Schopen et al. (2009), who analyzed approximately the same set of samples. In the current study, we accounted only the α_{s2}-CN isoforms that we identified on CZE electropherograms as described, whereas Schopen et al.

(2009) accounted also multiple peaks around the α_{s2}-CN-11P peak. Therefore, α_{s2}-CN concentration is defined differently in the current study, which explains the differences between the two studies.

Few studies reported on the genetic background of individual α_{s2}-CN phosphorylation isoforms. Only a single study has reported genetic parameters for the relative concentrations of individual α_{s2}-CN phosphorylation isoforms and for α_{s2}-CN PD measured by LC-ESI/MS method in the milk of French Montbéliarde cows (Fang et al., 2017). The data did not allow Fang et al. (2017) to adjust for differences between herds. To allow comparison with that study, we also estimated inter-herd heritabilities for the current study. The inter-herd heritabilities for all the traits in this study (0.41 to 0.82) were higher than the inter-herd heritabilities from the French Montbéliarde (0.07 to 0.37; Fang et al., 2017). Discrepancies between both studies might be due to genetic differences between the breeds (Dutch Holstein Friesian vs. French Montbéliarde), limited sample size in the study of Fang et al. (2017) which resulted in high standard errors of the estimates, and the use of different analytical methods (CZE vs. LC/ESI-MS). For measurement of α_{s1}-CN isoforms and α_{s1}-CN PD, the difference between CZE and LC/ESI-MS was trivial when we compared measurements from the same 10 samples analyzed with both methods (data not shown). Therefore, differences between estimated genetic parameters for these traits might be mainly due to breed differences, e.g. allele frequencies might differ between breeds. The β-LG genotypes affect α_{s1}-CN-8P concentration, and the *DGAT1* K232A polymorphism affects α_{s1}-CN-9P concentration in Dutch Holstein Friesian (Bijl et al., 2014a). However, the roles of β-LG and DGAT1 in regulating specific phosphorylation isoforms remain unclear. Nonetheless, such established associations might vary between breeds due to differences in allele frequencies. The frequency of β-LG A and B alleles were 0.52 and 0.48 in Dutch Holstein Friesian, respectively (Bijl et al., 2014a), whereas the frequencies of β-LG A and B alleles were 0.44 and 0.55 in French Montbéliarde, respectively (Fang et al., 2016). Furthermore, the frequency of the *DGAT1* A allele was 0.65 in Dutch Holstein Friesian (Bijl et al., 2014a), whereas *DGAT1* is nearly fixed at the A allele in French Montbéliarde (0.96, Gautier et al., 2007). The *DGAT1* K232A polymorphism accounts for about 13% of the additive genetic variance of α_{s1}-CN-9P concentration according to Bijl et al. (2014a). After adjusting for the additive genetic variance attributed to the *DGAT1* K232A polymorphism and accounting for differences in the statistical model (herd differences), the inter-herd heritability of α_{s1}-CN-9P concentration in Dutch Holstein Friesian (0.60) is in range with that in French Montbéliarde (0.56). The genetic background of α_{s2}-CN isoforms is currently unknown, but a similar explanation might apply to the discrepancy between genetic parameters for individual α_{s2}-CN isoforms.

4.3.3 Correlations among phosphorylation isoforms of α_{s1}-CN and α_{s2}-CN

Phenotypic correlations among individual α_s-CN phosphorylation isoforms were similar to the genetic ones (Table 4.2). Regarding genetic correlations, α_{s1}-CN-8P correlated weakly with the other α_s-CN phosphorylation isoforms (-0.10 to 0.17), suggesting that production of α_{s1}-CN-8P is rather independent from production of other α_s-CN phosphorylation isoforms. The phe-

notypic correlations (-0.31 to 0.20) of α_{s1}-CN-8P with other α_{s1}- and α_{s2}-CN phosphorylation isoforms in this study are in range with those (-0.12 to 0.43) reported by Fang et al. (2016) for French Montbéliarde. Both studies show weak to moderate correlations of α_{s1}-CN-8P with other isoforms, but the directions of the correlations are opposite. Taken together our results, results reported by Bijl et al. (2014a) and Fang et al. (2016), α_{s1}-CN-8P seems to be regulated differently than other α_s-CN phosphorylation isoforms. Regarding genetic correlations among α_{s2}-CN isoforms, α_{s2}-CN-12P correlated weakly with the other α_s-CN phosphorylation isoforms (-0.20 to 0.33) except moderately with α_{s2}-CN-11P (0.67); α_{s2}-CN-10P correlated strongly with α_{s2}-CN-11P (0.90). The phenotypic correlation between α_{s2}-CN-10P and α_{s2}-CN-11P (0.60) is in line with the one reported for French Montbéliarde (0.62), whereas the phenotypic correlation between α_{s2}-CN-11P and α_{s2}-CN-12P (0.50) is considerably higher than the estimate for French Montbéliarde (-0.28; Fang et al., 2016). Discrepancies between the two studies could be explained by differences between studied breeds as mentioned above. Regarding genetic correlations among α_{s1}- and α_{s2}-CN isoforms, α_{s1}-CN-9P correlated strongly negatively with α_{s2}-CN-10P (-0.90) and α_{s2}-CN-11P (-0.80), suggesting an antagonistic regulation between α_{s1}-CN phosphorylated at a higher level and α_{s2}-CN phosphorylated at a lower level. The phenotypic correlations of α_{s1}-CN-9P with α_{s2}-CN-10P and α_{s2}-CN-11P in this study agree with those reported by Fang et al. (2016). Both studies show that the proportion of α_{s1}-CN isoforms with higher levels of phosphorylation (i.e. α_{s1}-CN-9P) correlate negatively with the proportion of α_{s2}-CN isoforms with lower levels of phosphorylation (i.e. α_{s2}-CN-10P and -11P). Even though we could not estimate the correlations regarding α_{s2}-CN-13P and -14P in the current study, our results partially support the hypothesis proposed by Fang et al. (2016) that either α_{s1}- and α_{s2}-CN isoforms phosphorylated at lower levels would be present in higher concentrations in milk than the ones phosphorylated at higher levels or vice versa based on the correlations of α_{s1}-CN-9P with the other α_s-CN isoforms. Furthermore, the strong positive genetic correlation (0.94) between α_{s1}-CN PD and α_{s2}-CN PD suggests that the phosphorylation degrees of α_{s1}-CN and α_{s2}-CN are related. This is also supported by strong correlations of different α_s-CN phosphorylation isoforms with both α_{s1}-CN PD and α_{s2}-CN PD.

Table 4.2. Phenotypic (above diagonal) and genetic (below diagonal) correlations among individual α_{s1}- and α_{s2}-CN phosphorylation isoforms (% wt/wt) and the phosphorylation degrees (PD)^a of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses)

Trait (%wt/wt)	α_{s1}-CN-8P	α_{s1}-CN-9P	α_{s2}-CN-10P	α_{s2}-CN-11P	α_{s2}-CN-12P	α_{s1}-CN PD	α_{s2}-CN PD
α_{s1}-CN-8P		0.20 (0.03)	-0.31 (0.03)	-0.04 (0.03)	-0.00 (0.03)	-0.05 (0.03)	0.28 (0.03)
α_{s1}-CN-9P	0.17 (0.15)		-0.72 (0.02)	-0.69 (0.02)	0.00 (0.03)	0.96 (0.00)	0.75 (0.01)
α_{s2}-CN-10P	0.07 (0.17)	-0.90 (0.04)		0.60 (0.02)	0.06 (0.03)	-0.99 (0.02)	-0.82 (0.01)
α_{s2}-CN-11P	-0.06 (0.15)	-0.80 (0.05)	0.90 (0.04)		0.50 (0.03)	-0.80 (0.05)	-0.61 (0.02)
α_{s2}-CN-12P	-0.10 (0.16)	-0.20 (0.14)	0.33 (0.14)	0.67 (0.08)		-0.02 (0.04)	0.19 (0.03)
α_{s1}-CN PD^a	-0.02 (0.16)	0.98 (0.01)	-0.68 (0.02)	-0.72 (0.02)	-0.02 (0.04)		0.71 (0.02)
α_{s2}-CN PD^a	-0.02 (0.17)	0.91 (0.04)	-0.95 (0.03)	-0.80 (0.07)	-0.07 (0.16)	0.94 (0.03)	

[a] α_{s1}-CN PD= α_{s1}-CN-9P/ (α_{s1}-CN-8P+α_{s1}-CN-9P) × 100%; α_{s2}-CN PD = [α_{s2}-CN-12P/(α_{s2}-CN-10P + α_{s2}-CN-11P+α_{s2}-CN-12P)] × 100%.

4.3.4 Correlations of α_{s1}-CN and α_{s2}-CN phosphorylation isoforms with major milk proteins

Phenotypic correlations among individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and the major milk proteins were similar to the genetic ones (Table 4.3). Regarding genetic correlations, α_{s1}-CN correlated positively with both α_{s1}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD (0.59 to 0.79) but negatively with α_{s2}-CN-10P (-0.57) and α_{s2}-CN-11P (-0.54). The α_{s2}-CN correlated positively with all α_{s2}-CN phosphorylation isoforms (0.74 to 0.99) but negatively with α_{s1}-CN-9P (-0.75), α_{s1}-CN PD (-0.77) and α_{s2}-CN PD (-0.74). The decrease of α_{s2}-CN PD while total α_{s2}-CN increases can be explained by an increase in the proportion of isoforms with lower degrees of phosphorylation (i.e. α_{s2}-CN-10P and α_{s2}-CN-11P). This suggests that forming isoforms with lower degrees of phosphorylation is favored when production of α_{s2}-CN increases, whereas the opposite is observed when total α_{s1}-CN increases. Furthermore, Schopen et al. (2009) reported negative genetic correlation between α_{s1}-CN and α_{s2}-CN (-0.49). Here, we show that higher α_{s1}-CN concentration is associated with lower α_{s2}-CN-10P and α_{s2}-CN-11P concentrations, thus results in higher degree of phosphorylation of α_{s2}-CN, whereas higher α_{s2}-CN concentration is associated with lower α_{s1}-CN-9P concentration, thus results in lower degree of phosphorylation of α_{s1}-CN.

The genetic correlations between the relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and those of the other 4 major milk proteins (κ-CN, β-CN, α-LA and β-LG) were moderate to weak (-0.40 to 0.28). It should be noted that the relative concentration of κ-CN reported in the currently study was measured as its non-glycosylated and mono-phosphorylated form. Phenotypic correlations were in line with the ones reported by Fang et al. (2016) for French Montbéliarde, and the genetic correlations of α_{s1}-CN-8P concentration with the other 4 major milk proteins were in line with those reported by Gebreyesus et al. (2016) for Danish Holstein.

The genetic correlations among α_{s1}-CN PD, α_{s2}-CN PD and the other 4 major milk proteins were weak (-0.33 to 0.26). These moderate to weak correlations of α_{s1}- and α_{s2}-CN isoforms, α_{s1}-CN PD and α_{s2}-CN PD with the other 4 major milk proteins suggest that phosphorylation of α_{s1}- and α_{s2}-CN might be regulated differently from phosphorylation of β-CN and κ-CN. This is supported by the finding that β-CN might be phosphorylated in a different compartments of the Golgi apparatus in mice and rabbit than α_s-CN (Turner et al., 1993; Boisgard and Chanat, 2000). Phosphorylation of α_s-CN occurs in the Golgi cisternae during the transportation of milk proteins, whereas the phosphorylation of β-CN seems to occur later in the trans-Golgi network.

Table 4.3. Phenotypic (r_p) and genetic (r_g) correlations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms (% wt/wt) and the phosphorylation degrees (PD)[a] of α_{s1}-CN and α_{s2}-CN with the 6 major milk proteins (% wt/wt) measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses)

Trait	α_{s1}-CN-8P		α_{s1}-CN-9P		α_{s1}-CN PD		α_{s2}-CN-10P		α_{s2}-CN-11P		α_{s2}-CN-12P		α_{s2}-CN PD	
(%wt/wt)	r_p	r_g	r_p	r_g	r_p	r_g	r_p	r_g	r_p	r_g	r_p	r_g	r_p	r_g
α_{s1}-CN	0.75	0.73	0.74	0.79	-0.59	-0.57	-0.37	-0.54	0.09	-0.09	0.52	0.65	0.58	0.59
	(0.01)	(0.08)	(0.02)	(0.06)	(0.02)	(0.11)	(0.03)	(0.11)	(0.03)	(0.16)	(0.02)	(0.09)	(0.02)	(0.10)
α_{s2}-CN	-0.28	-0.09	-0.63	-0.75	0.77	0.84	0.86	0.99	0.54	0.74	-0.61	-0.77	-0.68	-0.74
	(0.03)	(0.15)	(0.03)	(0.06)	(0.01)	(0.05)	(0.01)	(0.01)	(0.02)	(0.07)	(0.02)	(0.06)	(0.02)	(0.07)
κ-CN[b]	-0.21	-0.33	-0.42	-0.40	0.26	0.28	0.23	0.14	-0.18	-0.20	-0.37	-0.33	-0.38	-0.33
	(0.03)	(0.15)	(0.03)	(0.12)	(0.03)	(0.15)	(0.03)	(0.14)	(0.03)	(0.14)	(0.03)	(0.13)	(0.03)	(0.13)
β-CN	0.12	-0.24	-0.03	0.21	-0.26	-0.36	-0.17	-0.31	-0.14	-0.25	0.00	0.24	0.16	0.26
	(0.03)	(0.19)	(0.03)	(0.17)	(0.03)	(0.16)	(0.03)	(0.15)	(0.03)	(0.17)	(0.03)	(0.17)	(0.03)	(0.17)
α-LA	0.22	0.27	-0.05	0.11	-0.07	0.09	0.17	0.13	0.04	0.24	-0.10	0.04	-0.08	-0.03
	(0.03)	(0.15)	(0.03)	(0.15)	(0.03)	(0.16)	(0.03)	(0.14)	(0.03)	(0.14)	(0.03)	(0.15)	(0.03)	(0.15)
β-LG	-0.12	-0.18	0.04	0.09	-0.14	-0.31	-0.22	-0.30	-0.19	-0.26	0.12	0.16	0.13	0.25
	(0.03)	(0.15)	(0.04)	(0.14)	(0.03)	(0.14)	(0.04)	(0.12)	(0.04)	(0.13)	(0.04)	(0.13)	(0.03)	(0.13)

[a] α_{s1}-CN PD= α_{s1}-CN-9P/ (α_{s1}-CN-8P+α_{s1}-CN-9P) × 100; α_{s2}-CN PD = [α_{s2}-CN-12P/(α_{s2}-CN-10P + α_{s2}-CN-11P+α_{s2}-CN-12P)] × 100.

[b] κ-CN was measured as nonglycosylated mono-phosphorylated form.

4.3.5 Genetic correlations of α_s-CN phosphorylation isoforms with milk Production traits and phosphorus

Generally, genetic correlations between individual α_{s1}- and α_{s2}-CN phosphorylation isoforms (% wt/wt) and milk production traits were weak to moderate, ranging from -0.59 to 0.53 (Table 4.4). All α_{s1}- and α_{s2}-CN phosphorylation isoforms except α_{s2}-CN-12P were either uncorrelated or correlated weakly with yields of protein, fat, lactose and milk (-0.30 to 0.36). The α_{s2}-CN-12P correlated moderately positive with yields of protein (0.53), lactose (0.52) and milk (0.48). These correlations suggest that cows producing larger quantities of protein, lactose and milk produce milk protein with a higher proportion of α_{s2}-CN-12P. Moreover, cows producing milk with higher protein and fat percentages tend to have milk protein with a higher proportion of α_{s2}-CN-10P and a lower proportion of α_{s1}-CN-9P, which results in lower α_{s1}-CN PD and α_{s2}-CN PD as shown by the genetic correlations of α_{s1}-CN PD and α_{s2}-CN PD with the milk production traits (-0.53 to -0.43).

In the past decades, selection for higher milk yield has been the main focus of most dairy cattle breeding schemes (Miglior et al., 2005). Based on the weak to moderate correlations between milk yield and α_{s1}- and α_{s2}-CN phosphorylation isoforms, α_{s1}- and α_{s2}-CN phosphorylation profile might have changed as selection for milk yield is expected to result in an increase of α_{s1}-CN-9P and α_{s2}-CN-12P concentrations while having little effect on the rest of the isoforms. Some countries include protein and fat percentages in their breeding goal, such as France, Italy and Switzerland (Miglior et al., 2005). Selection for higher protein and fat percentages will decrease α_{s1}-CN-9P concentration and increase α_{s2}-CN-10P and α_{s2}-CN-11P concentrations, which leads to lower α_{s1}-CN PD and α_{s2}-CN PD. This is also observed when comparing the α_{s1}- and α_{s2}-CN phosphorylation profiles between Danish Holstein and Danish Jersey. Milk from Danish Jersey, which has higher protein and fat percentages compared to Danish Holstein, contains higher proportions of isoforms with lower degrees of phosphorylation (Buitenhuis et al., 2016, α_{s1}-CN-8P and α_{s2}-CN-11P;). Furthermore, this agrees with the effects of *DGAT1* K232A genotypes; the K allele is associated with lower milk yield, higher protein and fat percentages, and lower α_{s1}-CN-9P concentration (Grisart et al., 2002; Winter et al., 2002; Bijl et al., 2014a).

The correlations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms with milk production traits also allow us to explore possible impacts of selection on technological properties of milk. To date, a few studies have investigated the relations between relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and milk technological properties. For dairy products favoring efficient α_{s1}-CN hydrolysis by chymosin, milk containing higher α_{s1}-CN-8P concentration and lower α_{s1}-CN-9P concentration might be preferred(Bijl et al., 2014b). In addition, to improve milk coagulation properties, milk containing a greater proportion of α_{s1}- and α_{s2}-CN isoforms with lower degrees of phosphorylation (e.g. α_{s1}-CN-8P, α_{s2}-CN-10P and α_{s2}-CN-11P) might be preferred (Frederiksen et al., 2011; Jensen et al., 2012; Ketto et al., 2017). Selecting for milk production traits shows little impact on α_{s1}-CN-8P concentration. However, selecting for higher protein and fat percentages would increase α_{s2}-CN-10P concentration and decrease α_{s1}-CN-9P concentration, which might be desirable for cheese production.

Limiting factors of casein phosphorylation? Phosphorylation of caseins is a process of

Table 4.4. Genetic correlations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms (% wt/wt) and the phosphorylation degrees (PD)[a] of α_{s1}-CN and α_{s2}-CN with milk production traits measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses)

Trait	Percentage (%)			Yield (kg)			
	Protein	Fat	Lactose	Protein	Fat	Lactose	Milk
(% wt/wt)							
α_{s1}-CN-8P	-0.15	-0.20	0.02	0.13	-0.04	0.20	0.21
	(0.17)	(0.17)	(0.16)	(0.22)	(0.19)	(0.18)	(0.19)
α_{s1}-CN-9P	-0.56	-0.59	0.19	-0.01	-0.29	0.36	0.32
	(0.11)	(0.12)	(0.14)	(0.21)	(0.16)	(0.16)	(0.16)
α_{s2}-CN-10P	0.47	0.51	0.02	0.01	0.33	-0.30	-0.28
	(0.13)	(0.14)	(0.16)	(0.22)	(0.15)	(0.17)	(0.18)
α_{s2}-CN-11P	0.37	0.34	-0.03	0.20	0.33	-0.05	-0.05
	(0.12)	(0.13)	(0.14)	(0.19)	(0.15)	(0.16)	(0.16)
α_{s2}-CN-12P	-0.10	-0.20	0.18	0.53	0.35	0.52	0.48
	(0.15)	(0.15)	(0.14)	(0.17)	(0.16)	(0.14)	(0.15)
α_{s1}-CN PD[a]	-0.52	-0.53	0.17	-0.07	-0.31	0.28	0.25
	(0.10)	(0.12)	(0.14)	(0.20)	(0.16)	(0.16)	(0.17)
α_{s2}-CN PD[a]	-0.43	-0.45	0.24	0.16	-0.07	0.41	0.38
	(0.14)	(0.15)	(0.15)	(0.21)	(0.19)	(0.16)	(0.17)

[a] α_{s1}-CN PD= α_{s1}-CN-9P/ (α_{s1}-CN-8P+α_{s1}-CN-9P) × 100; α_{s2}-CN PD = [α_{s2}-CN-12P/ (α_{s2}-CN-10P + α_{s2}-CN-11P+α_{s2}-CN-12P)] × 100.

phosphate addition to the polypeptide chains. Therefore, producing individual α_{s1}- and α_{s2}-CN phosphorylation isoforms is a function of both α_{s1}- and α_{s2}-CN synthesis and their subsequent phosphorylation. To investigate the biological relations between production of α_{s1}- and α_{s2}-CN isoforms, milk production, and phosphorus in milk, besides correlations with milk protein composition (Table 4.4), we also estimated the phenotypic and genetic correlations between yields of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, milk production traits and phosphorus yield (Table 4.5). The phenotypic and genetic correlations between yields of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and the protein yield were moderate to high (0.39 to 0.96). These correlations are expected as all isoforms undergo common regulation as milk protein synthesis. Furthermore, the weak phenotypic and genetic correlations (-0.23 to 0.16) among α_{s1}-CN PD, α_{s2}-CN PD and the yields of protein and phosphorus suggest that the phosphorylation is not an important rate limiting step; when more α_{s1}- and α_{s2}-CN are produced, the proportion of isoforms with higher degrees of phosphorylation is hardly affected.

We also investigated if phosphorus availability is a limiting factor for phosphorylation of all caseins when the total amount of casein is higher, including the phosphoproteins β-CN and κ-CN in addition to α_{s1}- and α_{s2}-CN. A strong positive phenotypic (0.91) and genetic (0.75) correlation between the total amount of phosphate groups attached to the caseins (Pcas) and the total amount of phosphorus in milk suggests little competition between Pcas and other

fractions of phosphorus (mainly inorganic phosphorus) for the total amount of phosphorus available in milk. Furthermore, our results are in line with those reported by Bijl et al. (2013), who showed that micellar phosphorus correlated positively with protein content and suggested that increase in total phosphorus mainly results from an increase in Pcas and micellar inorganic phosphorus. Results are supported by the fact that dietary phosphorus in cow's feed usually exceeds the requirements of the recommendation in the Netherlands (Valk et al., 2002). It is currently unknown if phosphorus distribution in milk will differ if cows suffer from phosphorus deficiency.

Table 4.5. Phenotypic and genetic correlations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms expressed in yields (g), the phosphorylation degrees (PD)[a] of α_{s1}-CN and α_{s2}-CN and the total amount of phosphate groups attached to the caseins in milk (Pcas)[b] with milk production traits and Phosphorus (P) yield measured on test-day morning milk samples from 1,721 Dutch Holstein Friesian cows (SE in parentheses)

Trait	Protein		Fat		Lactose		Milk		P	
(g)	r_p	r_g	r_p	r_g	r_p	r_g	r_p	r_g	r_p	r_g
α_{s1}-CN-8P	0.96 (0.00)	0.92 (0.04)	0.73 (0.01)	0.48 (0.16)	0.87 (0.01)	0.76 (0.09)	0.88 (0.01)	0.83 (0.07)	0.90 (0.01)	0.76 (0.10)
α_{s1}-CN-9P	0.77 (0.01)	0.53 (0.15)	0.55 (0.02)	0.06 (0.19)	0.83 (0.01)	0.69 (0.09)	0.83 (0.01)	0.72 (0.09)	0.72 (0.02)	0.30 (0.20)
α_{s2}-CN-10P	0.49 (0.02)	0.39 (0.20)	0.37 (0.03)	0.41 (0.16)	0.28 (0.03)	-0.06 (0.19)	0.30 (0.03)	-0.02 (0.19)	0.46 (0.02)	0.29 (0.21)
α_{s2}-CN-11P	0.80 (0.01)	0.59 (0.13)	0.62 (0.02)	0.50 (0.14)	0.62 (0.02)	0.29 (0.16)	0.62 (0.02)	0.31 (0.16)	0.76 (0.01)	0.60 (0.14)
α_{s2}-CN-12P	0.89 (0.01)	0.86 (0.06)	0.65 (0.02)	0.48 (0.14)	0.83 (0.01)	0.73 (0.08)	0.83 (0.01)	0.75 (0.08)	0.83 (0.01)	0.78 (0.09)
α_{s1}-CN PD[a]	-0.16 (0.03)	-0.07 (0.20)	-0.17 (0.03)	-0.31 (0.16)	0.07 (0.03)	0.28 (0.16)	0.05 (0.13)	0.25 (0.17)	-0.17 (0.03)	-0.23 (0.20)
α_{s2}-CN PD[a]	-0.14 (0.03)	0.16 (0.21)	-0.09 (0.03)	-0.07 (0.19)	0.10 (0.03)	0.41 (0.16)	0.07 (0.03)	0.38 (0.17)	-0.12 (0.03)	0.06 (0.22)
Pcas[b]	0.97 (0.00)	0.93 (0.03)	0.72 (0.01)	0.50 (0.16)	0.86 (0.01)	0.65 (0.12)	0.87 (0.01)	0.72 (0.10)	0.91 (0.01)	0.75 (0.11)

[a] α_{s1}-CN PD= α_{s1}-CN-9P/ (α_{s1}-CN-8P+α_{s1}-CN-9P) × 100; α_{s2}-CN PD = [α_{s2}-CN-12P/ (α_{s2}-CN-10P + α_{s2}-CN-11P+α_{s2}-CN-12P)] × 100.

[b] Pcas = $\sum[\frac{\text{individual casein fraction in milk (g/L)}}{\text{molecular weight (Da) of respective individual casein fraction}} \times$ number of phosphate groups attached to the attached to the respective casein fraction] × milk yield.

4.4 Conclusions

The high intra-herd heritabilites for α_{s1}- and α_{s2}-CN phosphorylation isoforms and for the phosphorylation degrees of α_{s1}- and α_{s2}-CN suggest high relevance of genetic factors and the possibility of altering α_{s1}- and α_{s2}-CN phosphorylation profiles by selective breeding in Dutch Holstein Friesian cows. Furthermore, the strong positive genetic correlation between α_{s1}-CN PD and α_{s2}-CN PD suggests that the phosphorylation degrees of α_{s1}-CN and α_{s2}-CN are related. Selecting for higher milk yield is expected to increase α_{s1}-CN-9P and α_{s2}-CN-12P concentrations. Selecting for higher protein and fat percentages is expected to increase α_{s2}-CN-10P concentration while decreasing α_{s1}-CN-9P concentration. Currently, knowledge regarding the effects of the phosphorylation degrees of α_{s1}- and α_{s2}-CN on technological properties of milk is limited and requires further investigation.

Acknowledgments

This study is part of the Dutch Milk Genomics Initiative, funded by Wageningen University (Wageningen, the Netherlands), the Dutch Dairy Association NZO (Zoetermeer, the Netherlands), Cooperative Cattle Improvement Organization CRV (Arnhem, the Netherlands), and the Dutch Technology Foundation STW (Utrecht, the Netherlands). The first author benefited from an Erasmus-Mundus fellowship and a grant by APIS-GENE, within the framework of the European Graduate School in Animal Breeding and Genetics (EGS-ABG, Paris, France). We gratefully acknowledge Dr. Guy Miranda (INRA, GABI-LGS, Jouy-en-Josas, France) and Guido Staring (NIZO food research, Ede, the Netherlands) for the excellent technical assistance of LC-ESI/MS and CZE analyses.

References

Bijl, E., H. J. F. van Valenberg, T. Huppertz, and A. C. M. van Hooijdonk (2013). Protein, casein, and micellar salts in milk: current content and historical perspectives. *J. Dairy Sci.* 96.9, pp. 5455–5464.

Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk, and H. Bovenhuis (2014a). Phosphorylation of αS1-casein is regulated by different genes. *J. Dairy Sci.* 97.11, pp. 7240–7246.

Bijl, E., H. van Valenberg, S. Sikkes, S. Jumelet, G. Sala, K. Olieman, T. van Hooijdonk, and T. Huppertz (2014b). Chymosin-induced hydrolysis of caseins: Influence of degree of phosphorylation of alpha-s1-casein and genetic variants of beta-casein. *Int. Dairy J.* 39.2, pp. 215–221.

Boisgard, R. and E. Chanat (2000). Phospholipase D-dependent and -independent mechanisms are involved in milk protein secretion in rabbit mammary epithelial cells. *Biochim. Biophys. Acta* 1495.3, pp. 281–296.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen (2016). Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genet.* 17.

Dalgleish, D. G. and M. Corredig (2012). The structure of the casein micelle of milk and its changes during processing. *Annu. Rev. Food Sci. Technol.* 3.1, pp. 449–467.

De Kruif, C. and C. Holt (2003). "Casein micelle structure, functions and interactions". In: *Adv. Dairy Chem. Proteins*. Ed. by P. F. Fox and P. L. H. McSweeney. 3rd. Vol. 1. JANUARY 2003. Springer US. Chap. 5, pp. 233–276.

De Kruif, C. G., T. Huppertz, V. S. Urban, and A. V. Petukhov (2012). Casein micelles and their internal structure. *Adv. Colloid Interface Sci.* 171-172, pp. 36–52.

Fang, Z. H., H. Bovenhuis, A. Delacroix-Buchet, G. Miranda, D. Boichard, M. H. P. W. Visker, and P. Martin (2017). Genetic and nongenetic factors contributing to differences in αS-casein phosphorylation isoforms and other major milk proteins. *J. Dairy Sci.* 100.7, pp. 5564–5577.

Fang, Z. H., M. H. P. W. Visker, G. Miranda, A. Delacroix-Buchet, H. Bovenhuis, and P. Martin (2016). The relationships among bovine αS-casein phosphorylation isoforms suggest different phosphorylation pathways. *J. Dairy Sci.* 99.10, pp. 8168–8177.

Frederiksen, P. D., K. K. Andersen, M. Hammershøj, H. D. Poulsen, J. Sørensen, M. Bakman, K. B. Qvist, and L. B. Larsen (2011). Composition and effect of blending of noncoagulating, poorly coagulating, and well-coagulating bovine milk from individual Danish Holstein cows. *J. Dairy Sci.* 94.10, pp. 4787–4799.

Gautier, M., A. Capitan, S. Fritz, A. Eggen, D. Boichard, and T. Druet (2007). Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphisms in French dairy cattle. *J. Dairy Sci.* 90.6, pp. 2980–2988.

Gebreyesus, G., M. S. Lund, L. Janss, N. A. Poulsen, L. B. Larsen, H. Bovenhuis, and A. J. Buitenhuis (2016). Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the Danish Holstein. *J. Dairy Sci.* 99.4, pp. 2863–2866.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson (2015). *ASReml user guide release 4.1 structural specification*.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res.* 12.2, pp. 222–231.

Heck, J., C. Olieman, A. Schennink, H. van Valenberg, M. Visker, R. Meuldijk, and A. van Hooijdonk (2008). Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *Int. Dairy J.* 18.5, pp. 548–555.

Holland, J. W. and M. J. Boland (2014). "Post-translational modifications of caseins". In: *Milk proteins From Expr. to food*. Ed. by H. Singh, A. Thompson, and M. Boland. Second Edi. Elsevier Inc., pp. 141–168.

Huppertz, T., I. Gazi, H. Luyten, H. Nieuwenhuijse, A. Alting, and E. Schokker (2017). Hydration of casein micelles and caseinates: Implications for casein micelle structure. *Int. Dairy J.* 74, pp. 1–11.

Jensen, H. B., N. A. Poulsen, K. K. Andersen, M. Hammershøj, H. D. Poulsen, and L. B. Larsen (2012). Distinct composition of bovine milk from Jersey and Holstein-Friesian cows with

good, poor, or noncoagulation properties as reflected in protein genetic variants and isoforms. *J. Dairy Sci.* 95.12, pp. 6905–6917.

Ketto, I. A., T. M. Knutsen, J. Øyaas, B. Heringstad, T. Ådnøy, T. G. Devold, and S. B. Skeie (2017). Effects of milk protein polymorphism and composition, casein micelle size and salt distribution on the milk coagulation properties in Norwegian Red cattle. *Int. Dairy J.* 70, pp. 55–64.

Miglior, F., B. L. Muir, and B. J. Van Doormaal (2005). Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88.3, pp. 1255–1263.

Recio, I., M. L. Pérez-Rodríguez, M. Ramos, and L. Amigo (1997). Capillary electrophoretic analysis of genetic variants of milk proteins from different species. *J. Chromatogr. A* 768.1, pp. 47–56.

Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. van Valenberg, and J. A. M. van Arendonk (2009). Genetic parameters for major milk proteins in Dutch Holstein-Friesians. *J. Dairy Sci.* 92.3, pp. 1182–1191.

Turner, M. D., S. E. Handel, C. J. Wilde, and R. D. Burgoyne (1993). Differential effect of brefeldin A on phosphorylation of the caseins in lactating mouse mammary epithelial cells. *J. Cell Sci.* 106, pp. 1221–1226.

Valk, H., L. Šebek, and A. Beynen (2002). Influence of Phosphorus Intake on Excretion and Blood Plasma and Saliva Concentrations of Phosphorus in Dairy Cows. *J. Dairy Sci.* 85.10, pp. 2642–2649.

Van Hulzen, K. J. E., R. C. Sprong, R. van der Meer, and J. A. M. van Arendonk (2009). Genetic and nongenetic variation in concentration of selenium, calcium, potassium, zinc, magnesium, and phosphorus in milk of Dutch Holstein-Friesian cows. *J. Dairy Sci.* 92.11, pp. 5754–5759.

Wilmink, J. B. M. (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16.4, pp. 335–348.

Winter, A., W. Krämer, F. a. O. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G. Thaller, and R. Fries (2002). Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc. Natl. Acad. Sci. U. S. A.* 99.14, pp. 9300–9305.

Chapter 5

Genome-wide association study for α_{s1}- and α_{s2}-casein phosphorylation in Dutch Holstein Friesian

Z. H. Fang[*†], H. Bovenhuis[†], H. J. F. van Valenberg,[‡], P.Martin[*], S. I. Duchemin[#], T. Huppertz[§] and M. H. P. W. Visker[†]

[*]Génétique Animale et Biologie Intégrative, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[†]Animal Breeding and Genomics, Wageningen University & Research, PO Box 338, 6700 AH Wageningen, the Netherlands
[‡] Dairy Science and Technology Group, Wageningen University & Research, PO Box 17, 6700 AA, Wageningen, the Netherlands
[§] NIZO, PO Box 20, 6710 BA, Ede, the Netherlands
[#] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden

Manuscript in preparation

Abstract

Phosphorylation of caseins (CN) is a crucial post-translational modification that allows caseins to form colloid particles called casein micelles. The α_{s1}- and α_{s2}-CN, showing varying degrees of phosphorylation (isoforms) in cows' milk, might be more relevant for stabilizing internal micellar structure than β- and κ-CN. However, little is known about the genetic backgrounds of individual α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN (α_{s1}-CN PD and α_{s2}-CN PD) defined as the proportion of isoforms with higher degrees of phosphorylation in total α_{s1}- and α_{s2}-CN, respectively. We aimed to identify genomic regions associated with these traits using 50K SNP for 1,857 Dutch Holstein Friesian cows. A total of 10 quantitative trait loci (QTL) regions were identified for all studied traits on 10 *Bos taurus* autosomes (BTA)1, 2, 6, 9, 11, 14, 15, 18, 24 and 28). Regions associated with multiple traits were found on BTA 1, 6, 11, and 14. We showed two QTL regions on BTA1: one affects α_{s2}-CN production, and the other one affects α_{s1}-CN PD and α_{s2}-CN PD, and harbors the *SLC37A1* gene that encodes for a protein that functions as a phosphorus antiporter. The QTL on BTA6 harbors the casein gene cluster and affects individual α_{s2}-CN phosphorylation isoforms. The QTL on BTA11 harbors the gene that encodes for β-lactoglobulin (β-LG) and affects relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. The QTL on BTA14 harbors the *DGAT1* gene and affects relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. Our results suggest that effects of identified genomic regions on α_{s1}-CN PD and α_{s2}-CN PD are probably due to changes in milk synthesis and phosphorus secretion in milk. The actual roles of SLC37A1, β-LG and DGAT1 in α_{s1}- and α_{s2}-CN phosphorylation in Dutch Holstein Friesian require further investigation.

Key words: posttranslational modification, milk protein composition, single nucleotide polymorphism, quantitative trait loci

5.1 Introduction

Protein phosphorylation regulates nearly every aspect of cell life, including disease states, by altering the structural confirmation of proteins to either activate, deactivate or modify their function. Caseins from cows' milk are the most well-studied group of phosphoproteins and play an important role in human nutrition as well as affect manufacturing properties of dairy products, especially cheese (Wedholm et al., 2006; Hallén et al., 2008; Caroli et al., 2009). Phosphorylation of caseins is a crucial post-translational modification that affects the formation and stability of casein micelles as the structure of micelles partly relies on the interactions between calcium phosphate nanoclusters and phosphoserine residues of α_{s1}-, α_{s2}-, and β-casein (**CN**) (De Kruif and Holt, 2003; De Kruif et al., 2012). As a result, large amounts of calcium and phosphorus can be delivered efficiently to the neonate without damaging the mammary gland of the mother by evoking either pathological calcification or amyloidosis (Holt et al., 2013).

Although α_{s1}-, α_{s2}-, β-, and κ-CN are all phosphorylated, α_{s1}- and α_{s2}-CN, showing various degrees of phosphorylation (isoforms) in cows' milk, are more heavily phosphorylated, possess multiple phosphoserine clusters. This suggests that α_{s1}- and α_{s2}-CN might be more relevant for stabilizing internal micellar structure than β- and κ-CN (Dalgleish and Corredig, 2012; Huppertz et al., 2017). α_{s1}-CN has been observed to carry 8 to 9 phosphate groups, and α_{s1}-CN-8P is the predominant isoform (Holland and Boland, 2014). α_{s2}-CN has been observed to carry 9 to 15 phosphate groups, and α_{s2}-CN-11P is the predominant isoform (Fang et al., 2016).

Relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms vary considerably among milk of individual cows (Bijl et al., 2014b; Fang et al., 2016), and exploitable genetic variation for these isoforms exists in French Montbéliarde (Fang et al., 2017), Danish Holstein and Jersey (Buitenhuis et al., 2016), and Dutch Holstein Friesian (Bijl et al., 2014b; Fang et al., accepted). Furthermore, the phosphorylation degrees (**PD**) of α_{s1}-CN and α_{s2}-CN, defined as the proportion of isoforms with higher degrees of phosphorylation, are heritable in French Montbéliarde (Fang et al., 2017) and highly heritable in Dutch Holstein Friesian (Fang et al., accepted). This indicates that the difference in the phosphorylation process is to a great extent determined by genetic factors. Additionally, Bijl et al. (2014a) showed that α_{s1}-CN-8P and α_{s1}-CN-9P are largely regulated by different sets of genes. Our recent work also suggests that α_{s1}-and α_{s2}-CN phosphorylated at lower degrees is regulated differently from α_{s1}- and α_{s2}-CN phosphorylated at higher degrees (Fang et al., 2016; Fang et al., 2017). To date, little is known about the genetic backgrounds of individual α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD. Therefore, this study aimed to identify genomic regions associated with these traits.

5.2 Materials and Methods

5.2.1 Animals

Test-day morning milk samples were collected from in total approximately 2,000 primiparous Dutch Holstein Friesian cows as part of the Dutch Milk Genomic Initiative. Cows were located

on 398 herds in the Netherlands, and at least 3 cows per herd were sampled. The pedigrees of the cows were supplied by cattle improvement organization CRV (Arnhem, the Netherlands). Detailed description of the experimental design is provided by Schopen et al. (2009).

5.2.2 Phenotypes

Milk production traits, phosphorus and milk protein composition from 1,857 milk samples collected in winter (February and March 2005) were available for the current study.

Milk production traits. Protein percentage was determined by infrared spectroscopy using MilkoScan FT 6000 (Foss Electric, Hillerød, Denmark) at the milk control station laboratory (Qlip, Zutphen, the Netherlands). phosphorus concentration was determined by inductively coupled plasma-atomic emission spectrometry (Vista Axial, Varian, Australia) from whole milk as described in van Hulzen et al. (2009). Test-day morning milk yield was available for 1,721 cows and was obtained from CRV. Yields of protein and phosphorus were calculated by multiplying the respective content traits by the observed test-day milk yield.

Milk protein composition. Relative concentrations (% wt/wt) of individual milk proteins and their isoforms were determined by capillary zone electrophoresis (**CZE**) by Heck et al. (2008) and Fang et al. (accepted). Yields (in grams) of individual milk proteins and their isoforms were calculated by multiplying relative concentrations (% wt/wt) by protein yield (in grams). Relative concentrations of α_{s1}- and α_{s2}-CN phosphorylation isoforms are the result of two distinct processes: the production of α_{s1}- and α_{s2}-CN and the posttranslational modification of both caseins. To characterize the phosphorylation process, we defined the phosphorylation degrees of α_{s1}-CN and α_{s2}-CN as the proportion of isoforms with higher degrees of phosphorylation (Fang et al., accepted), which were calculated as

$$\frac{\alpha_{s1}\text{—CN-9P}}{\alpha_{s1}\text{—CN-8P}+\alpha_{s1}\text{—CN-9P}} \times 100$$

$$\frac{\alpha_{s2}\text{—CN-12P}}{\alpha_{s2}\text{—CN-10P}+\alpha_{s2}\text{—CN-11P}+\alpha_{s2}\text{—CN-12P}} \times 100$$

Casein phosphorus. The phosphorus distribution in milk was quantified by estimating the content of phosphate groups attached to caseins in milk (i.e. molar concentration of casein phosphorus, $\mathbf{P}_{cas}$) and the total amount of phosphate groups attached to caseins in milk (i.e. output of casein phosphorus, P_{cas} yield). To derive P_{cas}, we first calculated the molar concentration ($\mathbf{C}_{molar}$) of each casein (α_{s1}-, α_{s2}-, κ- and β-CN) as its concentration in milk (g/L), calculated as protein percentage ×10 ×(% wt/wt), divided by its respective molecular weight (Da). As κ-CN carried 1 phosphate group and β-CN carried 5 in our milk samples, C_{molar} of phosphate groups attached to κ-CN and β-CN were approximated by multiplying C_{molar} of κ-CN and β-CN by 1 and 5, respectively. The C_{molar} of phosphate groups attached to α_{s1}-CN was the sum of C_{molar} of α_{s1}-CN-8P multiplied by 8 and C_{molar} of α_{s1}-CN-9P multiplied by 9. The C_{molar} of phosphate groups attached to α_{s2}-CN was the sum of C_{molar} of α_{s2}-CN-10P multiplied by 10, C_{molar} of α_{s2}-CN-11P multiplied by 11, and C_{molar} of α_{s2}-CN-12P multiplied by 12. Therefore, P_{cas} was the sum of C_{molar} of α_{s1}-, α_{s2}-, κ- and β-CN. Subsequently, P_{cas} yield was

approximated by multiplying P_{cas} by test-day milk yield.

5.2.3 Genotypes

DNA was isolated from blood samples of 1,868 cows for genotyping. As described in detail by Schopen et al. (2011), a 50K (50,000) SNP chip developed by CRV was used to genotype cows with the Infinium assay technology (Illumina Inc., San Diego, CA). The map positions of the SNP were based on bovine genome assembly BTAU 4.0 (Liu et al., 2009). Monomorphic SNP, SNP with a genotyping rate <80%, and SNP with less than 10 observations for one of the genotype classes were discarded (SNP with only two genotype classes instead of three were kept in the final marker set). After filtering, 44,669 SNP were retained for the genome-wide association study (*GWAS*). The data set used in the association study consisted of 1,667 animals with both phenotypes and genotypes. Protein variants A and B for β-lactoglobulin (β-LG) were genotyped for 1,671 cows as described by (Ganai et al., 2009). Genotypes for the diacylglycerol acyltransferase 1 (DGAT1) K232A polymorphism were obtained for 1,702 cows as described by Schennink et al. (2007).

5.2.4 Statistical analyses

GWAS. Single-SNP associations were analyzed using the following animal model:

$$y_{klmno} = \mu + \beta_1 \text{dim}_{klmno} + \beta_2 e^{-0.05*\text{dim}_{klmno}} + \beta_3 \text{ca}_{klmno} + \beta_4 \text{ca}^2_{klmno} + \text{season}_k + \text{scode}_l + \text{animal}_m + \text{herd}_n + \text{SNP}_o + e_{klmno}; \quad (5.1)$$

where y_{klmno} is the observation of the trait of interest; μ is the overall mean of the trait; dim_{klmno} is a covariate describing the effect of days in lactation, modeled with a Wilmink curve (Wilmink, 1987); ca_{klmno} is a covariate describing the effect of age at first calving; season_k is the fixed effect for calving season (June-August 2004, September-November 2004, and December 2004-February 2005); scode_l is the fixed effect accounting for possible differences in genetic level between proven bull daughters and young bull daughters; animal_m is the random additive genetic effect assumed to be distributed as $N(\mathbf{0},\mathbf{A}\sigma_a^2)$, where **A** is the additive genetic relationships matrix consisting of 26,300 animals, and σ_a^2 is the additive genetic variance; herd_n is the random herd effect assumed to be distributed as $N(\mathbf{0},\mathbf{I}\sigma_{herd}^2)$, where **I** is the identity matrix, and σ_{herd}^2 is the herd variance; SNP_o is the fixed effect of the SNP modelled as a class variable; e_{klmno} is the random residual effect assumed to be distributed as $N(\mathbf{0},\mathbf{I}\sigma_e^2)$, where **I** is the identity matrix, and σ_e^2 is the residual variance. The variance components were fixed to estimates obtained from 5.1 without the SNP effect. The effects of β-LG protein variants and DGAT1 genotypes were estimated using 5.1 by replacing the SNP effect by protein variant and genotype effects, respectively. All statistical analyses were performed using ASReml 4.1 (Gilmour et al., 2015).

Significance Thresholds. The genome-wide false discovery rate (**FDR**) was calculated based on the P-values obtained from single-SNP analyses using the R package qvalue (Dabney et al., 2010; R Core Team, 2015). The FDR was calculated for each trait individually. Associations with

an FDR <0.01 were considered significant. Obtained results were shown as Manhattan plots constructed by qqman R package (Turner, 2014).

QTL regions. Because of strong linkage disequilibrium between neighboring SNP in cattle genome, significant SNP located close to each other might be associated with the same causal variant. Therefore, we grouped those SNP into a single QTL region. A QTL region started with the first significant SNP on a chromosome that was followed by an additional significant SNP within 10 Mega-base pairs (**Mbp**), extended as long as another significant SNP occurred within 10 Mbp from the previous one, and ended at the last significant SNP that was not followed by another significant SNP within the next 10 Mbp. In this way, single, isolated and significant SNP were excluded because they have a higher risk of representing false positives.

5.3 Results and Discussion

In this study, we explored the genetic backgrounds of individual α_{s2}-CN phosphorylation isoforms (% wt/wt), and the phosphorylation degrees of α_{s1}-CN (α_{s1}-CN PD) and α_{s2}-CN (α_{s2}-CN PD). Phenotypic means, standard deviations and heritability estimates of all studied traits are given in Table 5.1. For α_{s2}-CN, the predominant isoform was α_{s2}-CN-11P. The proportion of isoforms with higher degrees of phosphorylation was 26% for α_{s1}-CN and 34% for α_{s2}-CN. Heritability estimates were moderate to high for all traits. Results have been discussed in detail by Fang et al. (accepted).

Table 5.1. Mean, standard deviation (SD), phenotypic variance[a] (σ_p^2), intra-herd heritability estimates (h^2), and proportion of phenotypic variance explained by herd (h_{herd}) for relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and for the phosphorylation degrees (PD)[b] of α_{s1}-CN and α_{s2}-CN measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses)

Trait (% wt/wt)	Mean	SD	σ_p^2	h^2	h_{herd}
α_{s1}-CN[c]	33.64	1.66	2.80	0.52 (0.11)	0.11 (0.02)
α_{s1}-CN-8P[c]	21.26	1.13	1.32	0.48 (0.10)	0.12 (0.02)
α_{s1}-CN-9P[c]	7.42	1.07	1.18	0.76 (0.12)	0.08 (0.02)
α_{s2}-CN[d]	6.67	0.95	0.98	0.94 (0.12)	0.08 (0.02)
α_{s2}-CN-10P[d]	0.99	0.39	0.16	0.54 (0.11)	0.10 (0.02)
α_{s2}-CN-11P[d]	3.44	0.57	0.33	0.89 (0.12)	0.08 (0.02)
α_{s2}-CN-12P[d]	2.24	0.22	0.05	0.71 (0.12)	0.07 (0.02)
Phosphorylation degree					
α_{s1}-CN PD[b]	25.79	2.72	7.66	0.78 (0.12)	0.08 (0.02)
α_{s2}-CN PD[b]	34.01	4.24	18.18	0.64 (0.11)	0.09 (0.02)

[a] Phenotypic variance after adjusting for systematic effects.
[b] α_{s1}-CN PD= α_{s1}-CN-9P/ (α_{s1}-CN-8P+α_{s1}-CN-9P) × 100%; α_{s2}-CN PD = [α_{s2}-CN-12P/(α_{s2}-CN-10P + α_{s2}-CN-11P+α_{s2}-CN-12P)] × 100%.
[c] Adopted from Bijl et al. (2014b).
[d] Adopted from Fang et al. (accepted).

The genome-wide association study showed significant associations for all studied traits, and a total of 10 QTL regions were identified (FDR <0.01) on 10 chromosomes (BTA1, 2, 6, 9, 11, 14, 15, 18, 24 and 28). Figure 5.1 shows the Manhattan plots for all traits analyzed with 44,669 SNP. Regions associated with multiple traits were found on BTA1, 6, 11, and 14, and their effects will be discussed in detail. Previously, Bijl et al. (2014a) conducted GWAS for α_{s1}-CN phosphorylation isoforms in the same Holstein Friesian population. They reported significant associations on BTA6 and BTA11 for α_{s1}-CN-8P, and on BTA6 and BTA14 for α_{s1}-CN-9P. Buitenhuis et al. (2016) conducted GWAS using BovineHD SNP for individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD in Danish Holstein and Jersey. In Danish Holstein, they detected 244 significant SNP distributed across BTA3, BTA5, BTA6, BTA9, BTA11, BTA12, BTA19, and BTA28 for α_{s2}-CN-11P, 2 significant SNP for α_{s2}-CN-12P, but no significant associations for α_{s1}-CN-8P and -9P, α_{s1}-CN PD and α_{s2}-CN PD. The majority of significant SNP were located on BTA6 where the casein gene cluster is. In Danish Jersey, they detected 25 significant SNP distributed across BTA2, 6 and 9 for α_{s2}-CN-11P, 9 significant SNP on BTA12 for α_{s1}-CN PD, but no significant associations for α_{s1}-CN-8P and -9P, α_{s2}-CN-12P and α_{s2}-CN PD. The reason for not finding significant associations for most of the traits might be their relatively small sample sizes (371 Danish Holstein and 321 Danish Jersey), which results in a lower detection power.

Fam20C was recently discovered as the genuine kinase that phosphorylates secretory pathway proteins with S-X-E/pS motifs (X represents any amino acid residue, and p indicates phosphorylation) including the caseins found in milk as well as several other proteins implicated in biomineralization (Tagliabracci et al., 2012). The *FAM20C* gene is located between 43.86 to 43.90 Mbp (BTAU 4.0) on BTA25. We did not detect a QTL signal for individual α_{s2}-CN phosphorylation isoforms, nor for α_{s1}-CN PD and α_{s2}-CN PD at that position. This is in line with results reported by Bijl et al. (2014a) and Buitenhuis et al. (2016), and suggests that no *FAM20C* variants are segregating in the Dutch Holstein population or in the Danish Holstein and Jersey populations.

Producing casein phosphorylation isoforms is a function of casein synthesis and their subsequent phosphorylation. Little is known about genes regulating the phosphorylation process. This process might be interlinked with different pathways of milk production, including milk protein synthesis and phosphorus secretion in milk. Therefore, to investigate if the detected QTL have direct effects on phosphorylation, we extended the analyses for the QTL associated with multiple traits on BTA1, 6, 11 and 14. Specifically, genotype effects of the most significantly associated (lead) SNP in each QTL were estimated for relative concentrations of α_{s1}- and α_{s2}-CN phosphorylation isoforms and for α_{s1}-CN PD and α_{s2}-CN PD, but also for yields of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, milk production traits, and content and yield of phosphorus and phosphate groups attached to caseins (P_{cas} and P_{cas} yield; Table 5.2).

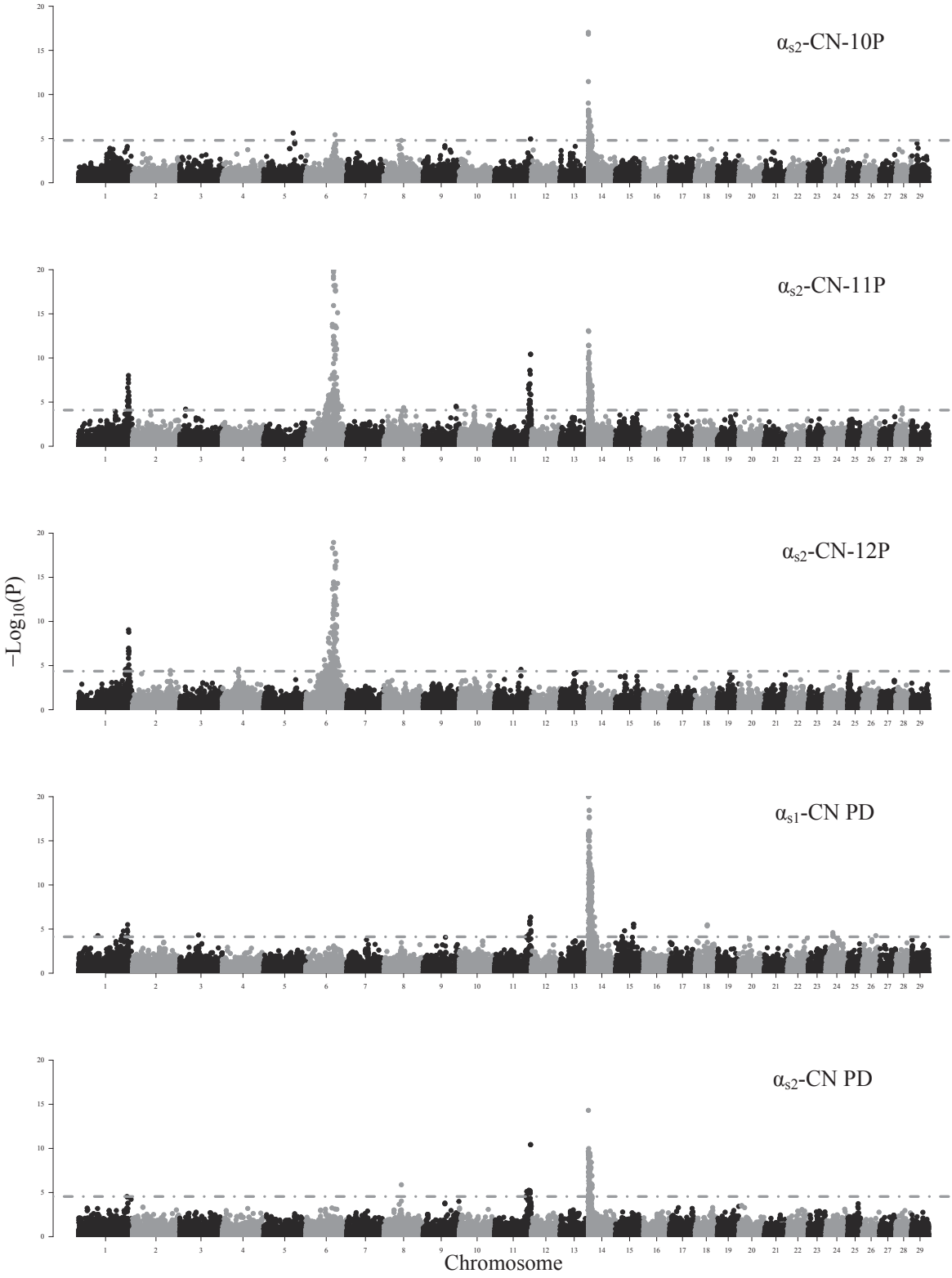

Figure 5.1. Significance [-Log$_{10}$(*P*)] of associaitons of 44,669 genome wide SNP located on 29 *Bos taurus* autosomes and the X chromosome with individual α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}-CN (α_{s1}-CN PD) and α_{s2}-CN (α_{s2}-CN PD). α_{s1}-CN PD = α_{s1}-CN-9P / (α_{s1}-CN-8P + α_{s1}-CN-9P) × 100; α_{s2}-CN PD = α_{s2}-CN-12P / (α_{s2}-CN-10P + α_{s2}-CN-11P + α_{s2}-CN-12P) × 100. The horizontal line represents a false discovery rate of 1%. The y-axes are cut off at -Log$_{10}$(*P*) =20.

5.3.1 BTA1

The QTL region between 145.55 and 152.18 Mbp on BTA1 was significantly associated with relative concentrations of α_{s2}-CN-11P and α_{s2}-CN-12P, and with α_{s1}-CN PD. However, the lead SNP differed between traits: ARS-BFGL-NGS-8140 at 149.19 Mbp was the lead SNP for α_{s2}-CN-11P concentration [-Log$_{10}$(*P*) = 8.01], ARS-BFGL-NGS-91705 (rs43282015) at 149.65 Mbp was the lead SNP for α_{s2}-CN-12P concentration [-Log$_{10}$(*P*) = 9.05] and ARS-BFGL-NGS-24811 at 146.63 Mbp was the lead SNP for α_{s1}-CN PD [-Log$_{10}$(*P*) = 5.50].

To investigate if this region could harbor multiple QTL (see Figure 5.2A for associations of α_{s2}-CN-12P and α_{s1}-CN PD as examples), associations were reanalyzed after adjusting for the lead SNP of α_{s2}-CN-12P concentration (ARS-BFGL-NGS-91705) for all studied traits (see Figure 5.2B for α_{s2}-CN-12P and α_{s1}-CN PD as examples). This analysis resulted in no significant associations between 145.55 and 152.18 Mbp for α_{s2}-CN-11P and α_{s2}-CN-12P except for one isolated SNP for α_{s2}-CN-12P. However, significant associations between 145.58 and 146.67 Mbp remained for α_{s1}-CN PD. Interestingly, associations just below the significance threshold between 144.41 and 147.37 Mbp on BTA1 were found for all studied traits at FDR <0.05 except for α_{s1}-CN-8P concentration. To characterize the region between 144.41 and 147.37 Mbp on BTA1, associations were reanalyzed after adjusting for the lead SNP of α_{s1}-CN-9P for all studied traits, and the same analyses were repeated with the respective lead SNP of α_{s2}-CN-10P, α_{s1}-CN PD and α_{s2}-CN PD (the lead SNP for each studied traits are shown in Figure 5.2C). The analysis of adjusting for the lead SNP of α_{s2}-CN-10P concentration (BTB-00068200) removed the QTL signal between 144.41 and 147.37 Mbp for all studied traits, but the QTL signal between 145.55 and 152.18 Mbp remained for α_{s2}-CN-11P and α_{s2}-CN-12P (see Figure 5.2D for α_{s2}-CN-12P as example). These results suggest that BTA1 harbors two QTL: QTL1 as the region between 145.55 and 152.18 Mbp and QTL2 as the region between 144.41 and 147.37 Mbp.

The effects of SNP ARS-BFGL-NGS-91705 (QTL1) and BTB-00068200 (QTL2) on relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD are given in Table 2. For ARS-BFGL-NGS-91705, the *G* allele was associated with lower α_{s2}-CN-11P and α_{s2}-CN-12P concentrations but not with α_{s1}-CN PD or α_{s2}-CN PD. For BTB-00068200, the *G* allele was associated with lower α_{s1}-CN-9P concentration and higher α_{s2}-CN-10P and -11P concentrations. This results in lower degrees of phosphorylation of α_{s1}-CN and α_{s2}-CN as shown by the negative association of the *GG* genotype with both α_{s1}-CN PD and α_{s2}-CN PD. Taken together, our results suggest that QTL1 between 145.55 and 152.18 Mbp on BTA1 affects α_{s2}-CN production, and QTL2 between 144.41 and 147.37 Mbp on BTA1 affects the phosphorylation degrees of α_{s1}- and α_{s2}-CN. Furthermore, combining our results and those reported by Bijl et al. (2014b) indicates that QTL1 on BTA1 is involved only in α_{s2}-CN production but not in α_{s1}-CN production, suggesting α_{s1}-CN and α_{s2}-CN are regulated differently. The lead SNP (ARS-BFGL-NGS-91705) of the QTL1 region (145.55-152.18 Mbp) is an intergenic variant. The gene closest to the lead SNP is F-box protein 25 (*FBXO25*) that is located at 149.56-149.59 Mbp on BTA1. In cattle, FBXO25 is involved in the pathway of post-translational protein ubiquitination as protein modification according to UniProt (http://www.uniprot.org/) but has not been associated with milk characteristics. The QTL2 region (144.41-147.37 Mbp) harbors the *SLC37A1*

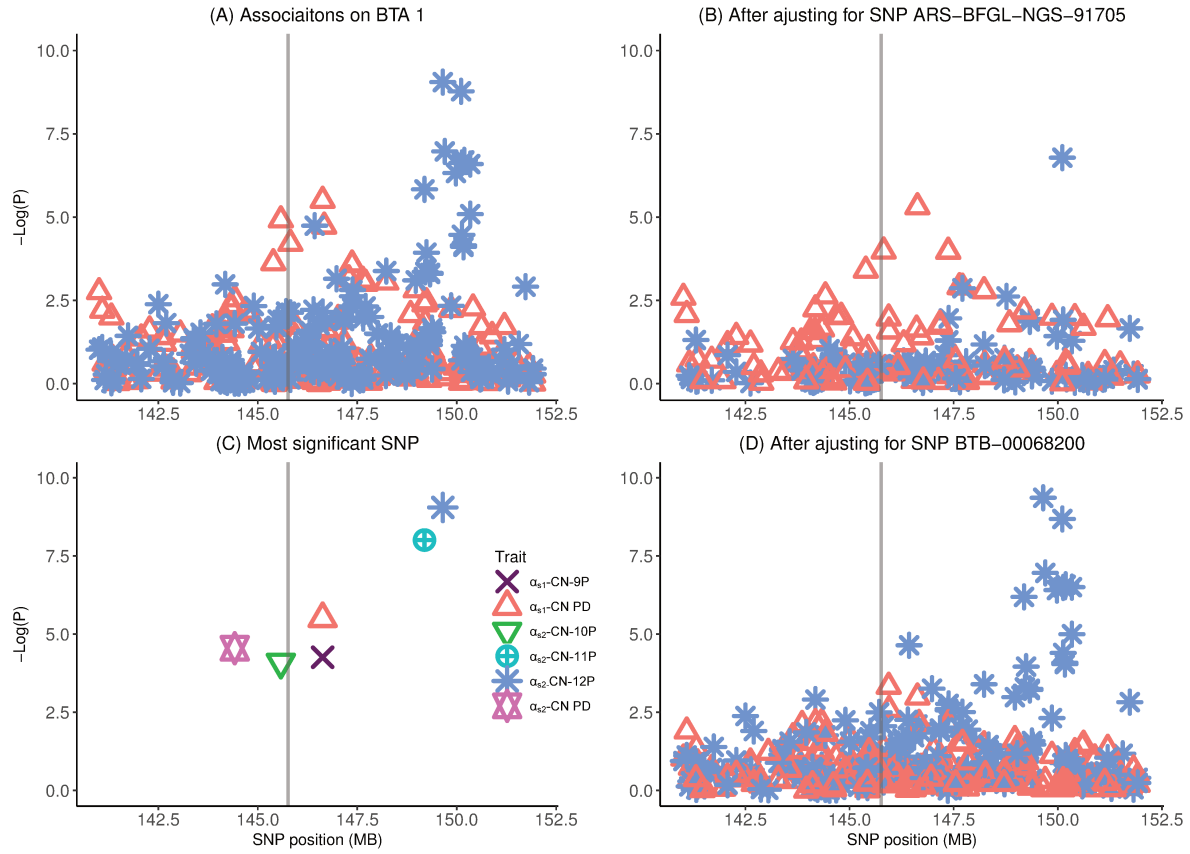

Figure 5.2. Significance [-$\text{Log}_{10}(P)$] of associations of SNP between 141 and 152.5 Mbp on BTA1 with (A) α_{s2}-CN-12P and the phosphorylation degree of α_{s1}-CN (α_{s1}-CN PD), (B) after including SNP ARS-BFGL-NGS-91705 genotypes as a fixed effect, (C) showing only the most signficant SNP for individual α_{s1}-CN and α_{s2}-CN phosphorylation isoforms, and the phosphorylation degrees of α_{s1}-CN (α_{s1}-CN PD) and α_{s2}-CN (α_{s2}-CN PD), and (D) after including SNP BTB-00068200 genotypes as a fixed effect. . α_{s1}-CN PD = α_{s1}-CN-9P / (α_{s1}-CN-8P + α_{s1}-CN-9P) × 100; α_{s2}-CN PD = α_{s2}-CN-12P / (α_{s2}-CN-10P + α_{s2}-CN-11P + α_{s2}-CN-12P) × 100. The shaded region corresponds to *SLC37A1* gene.

gene (145.72-145.80 Mbp) encoding for a protein functioning as a phosphorus antiporter that translocates inorganic phosphate in exchange of glucose-6-phosphate (Pan et al., 2011). Furthermore, a QTL associated with phosphorus concentration has been identified in this region in Danish Jersey and Australian Holstein, respectively (Buitenhuis et al., 2016; Kemper et al., 2016).

We detected significant effects of ARS-BFGL-NGS-91705 (QTL1) on the yields of α_{s1}-CN-8P, α_{s2}-CN-10P, -11P and -12P, protein, phosphorus and P_{cas} (Table 5.2). These consistent negative associations of the *G* allele with the yield traits confirm that QTL1 might affect only the production of α_{s2}-CN. This is supported by the fact that we detected the significant effect of QTL1 on test-day morning protein yield but did not detect significant effects on α_{s1}-CN PD and α_{s2}-CN PD in the current study. The effect on protein yield is relatively small, but α_{s2}-CN contributes only about 10 % to the total milk protein. Furthermore, this QTL has been reported to be associated with protein yield in Chinese Holstein (Jiang et al., 2010).

For BTB-00068200 (QTL2), we detected significant effects on yields of α_{s2}-CN-10P and -11P and phosphorus as well as phosphorus content. The *G* allele was associated with higher yields of α_{s2}-CN-10P and α_{s2}-CN -11P as well as higher content and yield of phosphorus. Furthermore, we did not detect significant effects of BTB-00068200 on yields of milk and protein. Therefore, the highly significant effect of QTL2 on phosphorus content [-Log$_{10}$(*P*)= 17.40] might be mainly due to the change of total phosphorus output in milk rather than a change of milk volume. Similarly, significant effects of QTL2 on relative concentrations of α_{s2}-CN-10P and -11P is probably mainly due to the change of yields of α_{s2}-CN-10P and α_{s2}-CN -11P rather than a change of protein yield. Taken together, these associations suggest that QTL2 have direct effects on phosphorylation, which might be related to the regulation of phosphorus output in milk. This is also supported by significant associations of BTB-00068200 with α_{s1}-CN PD and α_{s2}-CN PD. Furthermore, the *SLC37A1* gene located in this region plays a role in translocating inorganic phosphate (Pan et al., 2011), and it has been associated with the phosphorus content in cows' milk as mentioned above. Here, we show that this gene might have a direct effect on total phosphorus output in milk, especially on the inorganic phosphorus because we detected relatively small effects of BTB-00068200 on the content of phosphate groups attached to caseins (P_{cas}) and no significant effect on the total amount of phosphate groups attached to caseins (P_{cas} yield). We, therefore, hypothesize that QTL2 on BTA1 affects total phosphorus output in milk, which has an impact on the phosphorylation degrees of α_{s1}- and α_{s2}-CN.

5.3.2 BTA6

The QTL region between 46.52 and 103.18 Mbp on BTA 6 was significantly associated with relative concentrations of α_{s2}-CN-10P, α_{s2}-CN-11P, and α_{s2}-CN-12P. This region harbors the casein gene cluster (around 87 Mb). The SNP ARS-BFGL-NGS-94898 at 87.66 Mbp was the lead SNP for α_{s2}-CN-10P concentration [-Log$_{10}$(*P*) = 5.44]. The SNP ULGR_BTC-053514 at 83.57 Mbp was the lead SNP for both α_{s2}-CN-11P concentration [-Log$_{10}$(*P*) = 35.5] and α_{s2}-CN-12P concentration [-Log$_{10}$(*P*) = 42.3]. This SNP was also previously reported as the lead SNP for α_{s1}-CN-9P concentration (Bijl et al., 2014b) and for total α_{s2}-CN concentration (Schopen et al., 2011). No significant association on BTA6 was found with α_{s1}-CN PD and α_{s2}-CN PD (Figure 5.1), suggesting this region is only involved in casein production but not in the phosphorylation process. As shown by Fang et al. (accepted), the proportion of isoforms with higher degrees of phosphorylation is hardly affected when more α_{s1}- and α_{s2}-CN are produced, indicating that phosphorylation might not be an important rate limiting step.

The estimated effects of the lead SNP for α_{s2}-CN-12P concentration on relative concentrations of individual α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD show that the G allele was associated with lower concentrations of individual α_{s2}-CN phosphorylation isoforms (Table 5.2). Highly significant effects on yields of individual α_{s2}-CN phosphorylation isoforms confirm that this QTL affects α_{s2}-CN production. Note that this SNP did not pass the genome-wide significance threshold for α_{s1}-CN PD and α_{s2}-CN PD.

Table 5.2. Effects of SNP ARS-BFGL-NGS-91705 (BTA1, QTL1), BTB-00068200 (BTA1, QTL2), ULGR_BTC-053514 (BTA6), β-LG (BTA11) and diacylglycerol acyltransferase 1 (DGAT1, BTA14) genotypes on relative concentrations and yields of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, the phosphorylation degrees (PD)[a] of α_{s1}- and α_{s2}-CN, contents of protein, phosphorus (P) and P_{cas}[d], and yields of milk, protein, phosphorus and P_{cas}[e] measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses)

	BTA1 ARS-BFGL-NGS-91705[b]			**BTA1 BTB-00068200**[c]			**BTA6 ULGR_BTC-053514**			**BTA11 β-LG**			**BTA14 DGAT1**		
Trait	AA	GG	-Log$_{10}$(P)	AA	GG	-Log$_{10}$(P)	AA	GG	-Log$_{10}$(P)	AA	BB	-Log$_{10}$(P)	AA	KK	-Log$_{10}$(P)
(% wt/wt)	n=821	n=131		n=50	n=1121		n=637	n=945		n=539	n=262		n=628	n=276	
α_{s1}-CN-8P	0.00 (0.05)	0.05 (0.10)	0.06^{NS}	-0.19 (0.15)	0.02 (0.06)	0.39^{NS}	-0.42 (0.11)	0.49 (0.06)	23.80***	-0.32 (0.06)	0.41 (0.07)	17.20***	-0.06 (0.06)	0.01 (0.07)	0.27^{NS}
α_{s1}-CN-9P	-0.03 (0.05)	0.06 (0.10)	0.19 ^{NS}	0.06 (0.14)	-0.20 (0.06)	2.94**	-0.12 (0.11)	0.33 (0.05)	9.39***	0.07 (0.05)	-0.12 (0.07)	1.14^{NS}	0.53 (0.05)	-0.44 (0.06)	43.16***
α_{s2}-CN-10P	0.02 (0.02)	-0.06 (0.04)	0.87 ^{NS}	-0.03 (0.05)	0.09 (0.02)	4.13***	0.03 (0.04)	-0.08 (0.02)	4.42***	-0.06 (0.02)	0.09 (0.03)	5.33***	-0.12 (0.02)	0.13 (0.02)	19.47***
α_{s2}-CN-11P	0.08 (0.03)	-0.17 (0.05)	5.71***	-0.14 (0.08)	0.11 (0.03)	4.59***	0.20 (0.05)	-0.33 (0.03)	38.49***	-0.12 (0.03)	0.16 (0.04)	9.87***	-0.23 (0.03)	0.14 (0.03)	23.97***
α_{s2}-CN-12P	0.05 (0.01)	-0.07 (0.02)	8.51***	-0.03 (0.03)	-0.00 (0.01)	0.16^{NS}	0.13 (0.02)	-0.14 (0.01)	46.04***	-0.00 (0.01)	-0.00 (0.02)	0.07^{NS}	0.01 (0.01)	-0.04 (0.01)	1.50*
α_{s1}-CN PD[a]	-0.11 (0.12)	0.20 (0.22)	0.39 ^{NS}	0.40 (0.33)	-0.55 (0.13)	4.57***	-0.10 (0.26)	0.46 (0.13)	3.25 ***	0.42 (0.13)	-0.60 (0.16)	6.62***	1.32 (0.11)	-1.12 (0.15)	21.46***
α_{s2}-CN PD[a]	0.05 (0.13)	0.23 (0.24)	0.21 ^{NS}	0.12 (0.36)	-0.62 (0.14)	4.44***	-0.04 (0.28)	0.39 (0.13)	2.22**	0.49 (0.13)	-0.89 (0.17)	10.73***	0.94 (0.13)	-0.88 (0.16)	23.45***
Yield (g)	n=760	n=119		n=45	n=1045		n=81	n=868		n=492	n=245		n=578	n=250	
α_{s1}-CN-8P	0.21 (0.93)	-4.53 (1.76)	1.55*	-0.12 (2.69)	1.66 (1.01)	0.60^{NS}	-1.93 (2.05)	0.72 (1.00)	0.29^{NS}	-0.89 (0.98)	1.21 (1.25)	0.48^{NS}	0.18 (0.96)	-3.76 (1.24)	2.21**
α_{s1}-CN-9P	-0.05 (0.36)	-1.25 (0.68)	0.65^{NS}	0.19 (1.04)	-0.17 (0.40)	0.05^{NS}	-0.47 (0.79)	0.96 (0.38)	1.46*	0.60 (0.38)	-0.80 (0.48)	1.4*	2.68 (0.36)	-3.17 (0.46)	28.95***
α_{s2}-CN-10P	0.10 (0.11)	-0.59 (0.20)	2.31**	-0.32 (0.31)	0.42 (0.12)	3.34***	0.14 (0.23)	-0.45 (0.11)	4.28***	-0.26 (0.11)	0.35 (0.15)	2.99**	-0.60 (0.11)	0.40 (0.14)	10.66***
α_{s2}-CN-11P	0.42 (0.21)	-1.64 (0.40)	5.23***	-0.46 (0.61)	0.74 (0.23)	2.52**	0.88 (0.45)	-1.78 (0.22)	16.86***	-0.55 (0.23)	0.55 (0.29)	2.57**	-1.03 (0.22)	-0.08 (0.28)	5.05***
α_{s2}-CN-12P	0.27 (0.11)	-0.83 (0.21)	5.78***	-0.10 (0.32)	0.18 (0.12)	0.55^{NS}	0.60 (0.24)	-0.81 (0.12)	13.71***	-0.03 (0.12)	-0.14 (0.15)	0.23^{NS}	0.06 (0.11)	-0.60 (0.15)	4.24***

Continued on next page.

Table 5.2 (Continued) Effects of SNP ARS-BFGL-NGS-91705 (BTA1, QTL1), BTB-00068200 (BTA1, QTL2), ULGR_BTC-053514 (BTA6), β-LG (BTA11) and diacylglycerol acyltransferase 1 (DGAT1, BTA14) genotypes on relative concentrations and yields of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, the phosphorylation degrees (PD)[a] of α_{s1}- and α_{s2}-CN, contents of protein, phosphorus (P) and P_{cas}[d], and yields of milk, protein, phosphorus and P_{cas}[e] measured on test-day morning milk samples from 1,857 Dutch Holstein Friesian cows (SE in parentheses)

	BTA1 ARS-BFGL-NGS-91705[b]			**BTA1 BTB-00068200**[c]			**BTA6 ULGR_BTC-053514**			**BTA11 β-LG**			**BTA14 DGAT1**		
Trait	AA	GG	-$\text{Log}_{10}(P)$	AA	GG	-$\text{Log}_{10}(P)$	AA	GG	-$\text{Log}_{10}(P)$	AA	BB	-$\text{Log}_{10}(P)$	AA	KK	-$\text{Log}_{10}(P)$
Production and P	n=760	n=119		n=45	n=1045		n=85	n=945		n=492	n=245		n=578	n=250	
Protein content (%)	-0.01 (0.01)	-0.03 (0.03)	0.29^{NS}	-0.03 (0.04)	0.03 (0.02)	1.12^{NS}	0.02 (0.03)	-0.09 (0.01)	8.25^{***}	-0.02 (0.01)	0.02 (0.02)	0.75^{NS}	-0.15 (0.01)	0.11 (0.02)	44.41^{***}
P content (mg/kg)	4.92 (4.58)	-9.29 (8.60)	0.62^{NS}	-18.35 (12.61)	42.63 (4.96)	17.40^{***}	1.20 (9.99)	-8.90 (4.88)	0.77^{NS}	-3.93 (4.90)	2.90 (6.26)	0.22^{NS}	-47.90 (4.50)	38.66 (5.81)	41.94^{***}
P_{cas} content (mM)	-0.02 (0.04)	-0.61 (0.07)	0.18^{NS}	-0.16 (0.11)	0.08 (0.04)	1.44^{*}	0.04 (0.08)	-0.22 (0.04)	7.40^{***}	-0.16 (0.04)	0.15 (0.05)	6.49^{***}	-0.34 (0.04)	0.24 (0.05)	29.65^{***}
Milk yield (kg)	0.06 (0.13)	-0.49 (0.24)	1.05^{NS}	0.19 (0.36)	0.11 (0.14)	0.16^{NS}	-0.07 (0.28)	0.13 (0.14)	0.19^{NS}	0.15 (0.13)	-0.19 (0.17)	0.65^{NS}	0.65 (0.13)	-0.88 (0.17)	15.48^{***}
Protein yield (kg)	0.00 (0.00)	-0.02 (0.01)	1.91^{*}	0.00 (0.01)	0.01 (0.00)	0.51^{NS}	0.00 (0.01)	-0.01 (0.00)	0.57^{NS}	0.00 (0.00)	-0.00 (0.01)	0.30^{NS}	0.00 (0.00)	-0.02 (0.01)	2.65^{**}
P yield (mg)	121 (132)	-712 (248)	2.30^{**}	47 (379)	650 (143)	4.56^{***}	-23 (286)	-13 (140)	0.00^{NS}	104 (140)	-130 (177)	0.30^{NS}	-6 (137)	-448 (178)	1.46^{*}
P_{cas} yield (g)	0.20 (0.99)	-5.18 (1.00)	1.80^{*}	-0.12 (2.85)	1.77 (1.07)	0.61^{NS}	-0.21 (2.18)	-1.96 (1.06)	0.78^{NS}	-0.80 (1.05)	0.57 (1.32)	0.21^{NS}	0.36 (1.02)	-4.44 (1.32)	2.83^{**}

[a] α_{s1}-CN PD= α_{s1}-CN-9P/ (α_{s1}-CN-8P+α_{s1}-CN-9P) × 100%; α_{s2}-CN PD = [α_{s2}-CN-12P/(α_{s2}-CN-10P + α_{s2}-CN-11P+α_{s2}-CN-12P)] × 100%.

[b] ARS-BFGL-NGS-91705 (rs43282015) is the lead SNP of α_{s2}-CN-12P concentration (%wt/wt) in the QTL region between 145.55 and 152.18 Mbp on BTA1.

[c] BTB-00068200 (rs43281569) is the lead SNP of α_{s2}-CN-10P concentration (%wt/wt) in the QTL region between 144.41 and 147.37 Mbp on BTA1.

[d] $P_{cas} = \sum[\frac{\text{individual casein fraction in milk (g/L)}}{\text{molecular weight (Da) of respective individual casein fraction}} \times \text{number of phosphate groups attached to the attached to the respective casein fraction}]$.

[e] P_{cas} yield$=\sum[\frac{\text{individual casein fraction in milk (g/L)}}{\text{molecular weight (Da) of respective individual casein fraction}} \times \text{number of phosphate groups attached to the attached to the respective casein fraction}] \times \text{milk yield}$.

NS P >= 0.05, * P <0.05, ** P <0.01, *** P <0.001

5.3.3 BTA11

The QTL region between 95.06 and 109.41 Mbp on BTA11 was significantly associated with relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. This region harbors the PAEP gene encoding for β-LG. The SNP ULGR_SNP_X14710_1740 (rs41255679) at 107.2 Mbp was the lead SNP for α_{s2}-CN-10P concentration [-Log_{10}(*P*) = 4.96], α_{s2}-CN-11P concentration [-Log_{10}(*P*) = 10.42], α_{s1}-CN PD [-Log_{10}(*P*) = 6.33] and α_{s2}-CN PD [-Log_{10}(*P*) = 10.42]. This SNP was previously reported as the lead SNP on BTA11 for α_{s1}-CN-8P concentration (Bijl et al., 2014a). It is located in the promoter region of the *PAEP* gene and is in linkage disequilibrium with β-LG protein variants A and B (Ganai et al., 2009).

The estimated effects of β-LG genotypes on all studied traits (Table 5.2) show that the BB genotype was associated with higher α_{s1}-CN-8P, α_{s2}-CN-10P and α_{s2}-CN-11P concentrations (isoforms with lower degrees of phosphorylation). This results in lower degrees of phosphorylation of α_{s1}-CN and α_{s2}-CN as shown by the negative association of the BB genotype with both α_{s1}-CN PD and α_{s2}-CN PD. Buitenhuis et al. (2016) and Fang et al. (2017) did not detect significant effects of β-LG genotypes on individual α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD in Danish Holstein and Jersey, and in French Montbéliarde, respectively. Differences between studies might be due to the genetic differences between studied breeds (Holstein, Montbéliarde and Jersey) such as differences in linkage disequilibrium between β-LG genotypes and other variants that affect the traits of interest, limited sample size of Buitenhuis et al. (2016) and Fang et al. (2017), and the use of different analytical methods. Regarding differences in linkage disequilibrium between β-LG genotypes and other variants across breeds, Bijl et al. (2014b) detected a significant effect of β-LG genotypes only on α_{s1}-CN-8P concentration in Dutch Holstein Friesian, whereas Fang et al. (2017) detected significant effects of β-LG genotypes on both α_{s1}-CN-8P and α_{s1}-CN-9P concentrations in French Montbéliarde. As the reported effect of β-LG BB genotype on β-LG concentration by Fang et al. (2017) was about 1.5 times larger than the one in Dutch Holstein Friesian reported by Heck et al. (2009), the observed genotype effect in French Montbéliarde might be the combination of the effect of β-LG genotypes and the effect of one or multiple linked variants that also affect β-LG concentration. The additional effect of such a linked variant might explain the differences between the effects of β-LG genotypes on individual α_{s1}- and α_{s2}-CN isoforms, α_{s1}-CN PD and α_{s2}-CN PD in different breeds. Regarding differences in analytical methods, differences between CZE used by Bijl et al. (2014b) and LC-ESI/MS used by Fang et al. (2017) seem to be negligible for the measurement of α_{s1}-CN isoforms and α_{s1}-CN PD according to Fang et al. (accepted). However, protein fractions measured with the same analytical method, such as LC (as used by Buitenhuis et al. (2016) and by Fang et al. (2017)), may still differ because of differences in separation conditions.

For the effects of β-LG genotypes on yields of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, milk production traits, content and yield of phosphorus, P_{cas} and P_{cas} yield, we detected significant effects on yields of α_{s1}-CN-9P, α_{s2}-CN-10P and -11P, but the effects on α_{s1}-CN-9P yield were relatively small. The BB genotype was associated with higher yields of α_{s2}-CN-10P and -11P. Surprisingly, we did not detect a significant effect on the yield of α_{s1}-CN-8P, whereas we detected a highly significant effect on α_{s1}-CN-8P concentration. Furthermore, we

detected highly significant effects of β-LG genotypes on P_{cas} [-Log_{10}(*P*) = 6.49] but no effect on P_{cas} yield. Several studies have shown that the β-LG B variant decreases the proportion of β-LG, which results in increased proportions of caseins (Bobe et al., 1999; Hallén et al., 2008; Heck et al., 2009; Bonfatti et al., 2010; Fang et al., 2017). We show that the β-LG B variant increases only α_{s1}- and α_{s2}-CN isoforms phosphorylated at a lower degree and decrease α_{s1}-CN PD and α_{s2}-CN PD. Given that phosphorylation might not be an important rate limiting step when more α_{s1}- and α_{s2}-CN are produced as discussed above, it is unlikely that the effects of β-LG genotypes on α_{s1}-CN PD and α_{s2}-CN PD are due to increased casein production. Additionally, analyses showed no significant interaction between β-LG genotypes and QTL2 on BTA1 for α_{s1}-CN PD (*P* = 0.26) and α_{s2}-CN PD (*P* = 0.56), suggesting the amount of phosphorus available is not rate limiting. This is supported by the positive correlations between the total amount of phosphate groups attached to the caseins (P_{cas} yield) and the total amount of phosphorus in milk (Fang et al., accepted). Taken together, β-LG seems to play a role in regulating milk protein composition, proportion of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, and the phosphorylation process. The actual mechanism causing the effects of β-LG genotypes on the concentrations of the other milk proteins is currently unknown as well as the role of β-LG in the phosphorylation process.

5.3.4 BTA14

The QTL region between 0.2 and 19.36 Mbp on BTA14 was significantly associated with relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD, and α_{s2}-CN PD. This region harbors the *DGAT1* gene. The SNP ULGR_SNP_AJ318490_1c (rs109234250) at 0.44 Mbp was the lead SNP for α_{s2}-CN-10P concentration [-Log_{10}(*P*) = 17.04], α_{s2}-CN-11P concentration [-Log_{10}(*P*) = 21.98], α_{s1}-CN PD [-Log_{10}(*P*) = 48.79] and α_{s2}-CN PD [-Log_{10}(*P*) = 21.55]. This SNP was previously reported as the lead SNP on BTA14 for α_{s1}-CN-9P concentration (Bijl et al., 2014b), and is one of two SNPs responsible for the DGAT1 K232A polymorphism.

The effects of DGAT1 genotypes on all studied traits (Table 5.2) show that the *K* allele was associated with higher α_{s2}-CN-10P and α_{s2}-CN-11P concentrations (isoforms with lower degrees of phosphorylation) and lower concentration of α_{s1}-CN-9P (isoform with higher degrees of phosphorylation). This results in lower degrees of phosphorylation of α_{s1}- and α_{s2}-CN as shown by the negative association of the *K* allele with both α_{s1}-CN PD and α_{s2}-CN PD. Furthermore, DGAT1 does not affect α_{s1}-CN-8P concentration (Bijl et al., 2014b) and α_{s2}-CN-12P concentration (this study) at genome wide significance threshold. Bovenhuis et al. (2016) showed that DGAT1 affects α_{s2}-CN concentration in Dutch Holstein Friesian and Danish Holstein Friesian. The *K* allele was associated with higher α_{s2}-CN concentration. Here, we show that the increase of α_{s2}-CN concentration is due to the increase of α_{s2}-CN-10P and α_{s2}-CN-11P concentrations in Dutch Holstein Friesian.

For the effects of DGAT1 genotypes on the yields of individual isoforms and milk production traits, we detected significant effects on yields of all α_{s1}- and α_{s2}-CN phosphorylation isoforms, milk, protein, phosphorus and P_{cas} as well as contents of protein, phosphorus and P_{cas}. The effects of the DGAT1 genotypes on the yields of α_{s1}-CN-8P and α_{s2}-CN-12P were relatively

small, and might be due to the change in the protein yield as the genotype effects on the yields of α_{s1}-CN-8P, α_{s2}-CN-12P and protein were in the same direction and of similar magnitude. The highly significant effects on the yields of α_{s1}-CN-9P, α_{s2}-CN-10P and α_{s2}-CN-11P suggest direct effects of DGAT1 on these isoforms as the direction and magnitude of effects on their relative concentration in milk and yields are similar. The biological relation between DGAT1, content and yield of fat, and fatty acid composition are easier to comprehend as the DAGT1 enzyme is involved in biosynthesis of triacylglycerols (Coleman and Lee, 2004), whereas the biological relation of DGAT1 and phosphorylation of caseins is still unclear. It seems that the contribution of DGAT1 to the variation in specific isoforms in Dutch Holstein Friesian might rather be linked to the correlated response of change in milk production, such as milk volume. The DGAT1 K232 polymorphism does not segregate in French Montebéliarde (Gautier et al., 2007), so other genes rather than DGAT1 contribute to the variation in specific isoforms and phosphorylation degrees of α_{s1}- and α_{s2}-CN in this breed.

5.3.5 Additional regions

In addition to the four QTL regions shared by multiple traits as described above, we also detected trait-specific QTL on BTA2, 9, 15, 18, 24 and 28. On BTA2, a QTL region located between 113.63 and 113.67 Mbp was associated uniquely with α_{s2}-CN-12P concentration. The gene closest to this QTL region is ephrin type-A receptor 4 precursor (*EPHA4*) that is located at 114.15-114.20 Mbp on BTA2. In cattle, EPHA4 is an uncharacterized protein, whereas in human, it is a kinase phosphorylating tyrosine and is involved in cell adhesion and neurogenesis according to UniProt. Two QTL regions were associated uniquely with α_{s2}-CN-11P concentration. A QTL region located between 98.45 and 99.32 Mbp on BTA9 was associated uniquely. The lead SNP ARS-BFGL-NGS-102803 (rs109099768) is an intron variant located in the serine active site containing 1 (*SERAC1*) gene. In human, the SERAC1 protein plays an important role in mediating phospholipid exchange that is essential for both mitochondrial functioning and intracellular cholesterol trafficking according to UniProt. The other QTL region located between 18.55 and 19.13 Mbp on BTA28 harbors the receptor accessory protein 3 (*REEP3*) gene. A total of four unique QTL regions were associated with α_{s1}-CN PD, which were located at 54.61 Mbp on BTA15, between 35.68 and 36.09 Mbp on BTA18, between 20.49 and 21.11 Mbp on BTA24, respectively. On BTA15, the gene closest to the QTL region is microtubule affinity-regulating kinase 1 (*MARK1*). On BTA18, the QTL region harbors the proteasome 26S subunit, non-ATPase 7 (*PSMD7*) gene (36.00-36.02 Mbp). On BTA24, the QTL region harbors the CUGBP Elav-like family member 4 (*CELF4*) gene. None of the genes mentioned above has been associated with milk characteristics. Thus, no clear candidate genes could be identified for those trait-specific QTL.

5.4 Conclusions

We detected a total of 10 QTL regions for relative concentrations of individual α_{s2}-CN phosphorylation isoforms, and the phosphorylation degrees of α_{s1}- and α_{s2}-CN (α_{s1}-CN PD and α_{s2}-CN

PD) on chromosomes 1, 2, 6, 9, 11, 14, 15, 18, 24 and 28. Regions associated with multiple traits were found on BTA 1, 6, 11, and 14. We showed two QTL regions on BTA1: one affects α_{s2}-CN production, and the other encodes SLC37A1 and affects α_{s1}-CN PD and α_{s2}-CN PD. The QTL region on BTA6 harbors the casein gene cluster and affects individual α_{s2}-CN phosphorylation isoforms. The QTL region on BTA11 harbors the gene encoding β-LG and affects relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD The QTL region on BTA14 harbors DGAT1 and affects relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. Elucidation of the actual roles of SLC37A1, β-LG and DGAT1 in α_{s1}- and α_{s2}-CN phosphorylation in Dutch Holstein Friesian requires further investigation.

Acknowledgments

This study is part of the Dutch Milk Genomics Initiative, funded by Wageningen University (Wageningen, the Netherlands), the Dutch Dairy Association NZO (Zoetermeer, the Netherlands), Cooperative Cattle Improvement Organization CRV (Arnhem, the Netherlands), and the Dutch Technology Foundation STW (Utrecht, the Netherlands). The first author benefited from an Erasmus-Mundus fellowship and a grant by APIS-GENE, within the framework of the European Graduate School in Animal Breeding and Genetics (EGS-ABG, Paris, France).

References

Bijl, E., R. de Vries, H. van Valenberg, T. Huppertz, and T. van Hooijdonk (2014a). Factors influencing casein micelle size in milk of individual cows: Genetic variants and glycosylation of κ-casein. *Int. Dairy J.* 34.1, pp. 135–141.

Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk, and H. Bovenhuis (2014b). Phosphorylation of αS1-casein is regulated by different genes. *J. Dairy Sci.* 97.11, pp. 7240–7246.

Bobe, G., D. Beitz, A. Freeman, and G. Lindberg (1999). Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. *J. Dairy Sci.* 82.12, pp. 2797–2804.

Bonfatti, V., G. Di Martino, A. Cecchinato, L. Degano, and P. Carnier (2010). Effects of β-κ-casein (CSN2-CSN3) haplotypes and β-lactoglobulin (BLG) genotypes on milk production traits and detailed protein composition of individual milk of Simmental cows. *J. Dairy Sci.* 93.8, pp. 3797–3808.

Bovenhuis, H., M. Visker, N. Poulsen, J. Sehested, H. van Valenberg, J. van Arendonk, L. Larsen, and A. Buitenhuis (2016). Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. *J. Dairy Sci.* 99.4, pp. 3113–3123.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen (2016). Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genet.* 17.

Caroli, A. M., S. Chessa, and G. J. Erhardt (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92.11, pp. 5335–5352.

Coleman, R. A. and D. P. Lee (2004). Enzymes of triacylglycerol synthesis and their regulation. *Prog. Lipid Res.* 43.2, pp. 134–176.

Dabney, A., J. D. Storey, and G. R. Warnes (2010). *qvalue: Q-value estimation for false discovery rate control*.

Dalgleish, D. G. and M. Corredig (2012). The structure of the casein micelle of milk and its changes during processing. *Annu. Rev. Food Sci. Technol.* 3.1, pp. 449–467.

De Kruif, C. and C. Holt (2003). "Casein micelle structure, functions and interactions". In: *Adv. Dairy Chem. Proteins*. Ed. by P. F. Fox and P. L. H. McSweeney. 3rd. Vol. 1. JANUARY 2003. Springer US. Chap. 5, pp. 233–276.

De Kruif, C. G., T. Huppertz, V. S. Urban, and A. V. Petukhov (2012). Casein micelles and their internal structure. *Adv. Colloid Interface Sci.* 171-172, pp. 36–52.

Fang, Z. H., H. Bovenhuis, A. Delacroix-Buchet, G. Miranda, D. Boichard, M. H. P. W. Visker, and P. Martin (2017). Genetic and nongenetic factors contributing to differences in αS-casein phosphorylation isoforms and other major milk proteins. *J. Dairy Sci.* 100.7, pp. 5564–5577.

Fang, Z. H., H. Bovenhuis, H. J. F. van Valenberg, P. Martin, T. Huppertz, and M. H. P. W. Visker (accepted). Genetic parameters for αS1- and αS2-casein phosphorylation isoforms in Dutch Holstein Friesian. *J. Dairy Sci.*

Fang, Z. H., M. H. P. W. Visker, G. Miranda, A. Delacroix-Buchet, H. Bovenhuis, and P. Martin (2016). The relationships among bovine αS-casein phosphorylation isoforms suggest different phosphorylation pathways. *J. Dairy Sci.* 99.10, pp. 8168–8177.

Ganai, N. A., H. Bovenhuis, J. A. M. van Arendonk, and M. H. P. W. Visker (2009). Novel polymorphisms in the bovine beta-lactoglobulin gene and their effects on beta-lactoglobulin protein concentration in milk. *Anim. Genet.* 40.2, pp. 127–133.

Gautier, M., A. Capitan, S. Fritz, A. Eggen, D. Boichard, and T. Druet (2007). Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphisms in French dairy cattle. *J. Dairy Sci.* 90.6, pp. 2980–2988.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson (2015). *ASReml user guide release 4.1 structural specification*.

Hallén, E., A Wedholm, A Andrén, A Lundén, A. Wedholm, A. Andrén, and A. Lundén (2008). Effect of beta-casein, kappa-casein and beta-lactoglobulin genotypes on concentration of milk protein variants. *J. Anim. Breed. Genet.* 125.2, pp. 119–129.

Heck, J. M. L., A. Schennink, H. J. F. Van Valenberg, H. Bovenhuis, M. H. P. W. Visker, J. A. M. Van Arendonk, and A. C. M. van Hooijdonk (2009). Effects of milk protein variants on the protein composition of bovine milk. *J. Dairy Sci.* 92.3, pp. 1192–1202.

Heck, J., C. Olieman, A. Schennink, H. van Valenberg, M. Visker, R. Meuldijk, and A. van Hooijdonk (2008). Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *Int. Dairy J.* 18.5, pp. 548–555.

Holland, J. W. and M. J. Boland (2014). "Post-translational modifications of caseins". In: *Milk proteins From Expr. to food*. Ed. by H. Singh, A. Thompson, and M. Boland. Second Edi. Elsevier Inc., pp. 141–168.

Holt, C., J. A. Carver, H. Ecroyd, and D. C. Thorn (2013). Invited review: Caseins and the casein micelle: Their biological functions, structures, and behavior in foods. *Joural Dairy Sci.* 96.10, pp. 6127–6146.

Huppertz, T., I. Gazi, H. Luyten, H. Nieuwenhuijse, A. Alting, and E. Schokker (2017). Hydration of casein micelles and caseinates: Implications for casein micelle structure. *Int. Dairy J.* 74, pp. 1–11.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang (2010). Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One* 5.10, e13661.

Kemper, K. E., M. D. Littlejohn, T. Lopdell, B. J. Hayes, P. M. Visscher, M. Carrick, and M. E. Goddard (2016). Leveraging genetically simpler traits to identify small-effect variants for complex phenotypes. *BMC Genomics*, pp. 1–9.

Liu, Y., X. Qin, X. Z. Song, H. Jiang, Y. Shen, K. J. Durbin, S. Lien, M. P. Kent, M. Sodeland, Y. Ren, L. Zhang, E. Sodergren, P. Havlak, K. C. Worley, G. M. Weinstock, and R. A. Gibbs (2009). Bos taurus genome assembly. *BMC Genomics* 10.1, p. 180.

Pan, C. J., S. Y. Chen, H. S. Jun, S. R. Lin, B. C. Mansfield, and J. Y. Chou (2011). SLC37A1 and SLC37A2 are phosphate-linked, glucose-6-phosphate antiporters. *PLoS One* 6.9, pp. 1–8.

R Core Team (2015). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*.

Schennink, A., W. M. Stoop, M. H. P. W. Visker, J. M. L. Heck, H. Bovenhuis, J. J. Van Der Poel, H. J. F. Van Valenberg, and J. A. M. Van Arendonk (2007). DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. *Anim. Genet.* 38.5, pp. 467–473.

Schopen, G. C. B., P. D. Koks, J. A. M. van Arendonk, H. Bovenhuis, and M. H. P. W. Visker (2009). Whole genome scan to detect quantitative trait loci for bovine milk protein composition. *Anim. Genet.* 40.4, pp. 524–37.

Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and H. Bovenhuis (2011). Whole-genome association study for milk protein composition in dairy cattle. *J. Dairy Sci.* 94.6, pp. 3148–3158.

Tagliabracci, V. S., J. L. Engel, J. Wen, S. E. Wiley, C. A. Worby, L. N. Kinch, J. Xiao, N. V. Grishin, and J. E. Dixon (2012). Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* 336.6085, pp. 1150–3.

Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *BioRxiv*, p. 5165.

Van Hulzen, K. J. E., R. C. Sprong, R. van der Meer, and J. A. M. van Arendonk (2009). Genetic and nongenetic variation in concentration of selenium, calcium, potassium, zinc, magnesium, and phosphorus in milk of Dutch Holstein-Friesian cows. *J. Dairy Sci.* 92.11, pp. 5754–5759.

Wedholm, A., L. B. Larsen, H. Lindmark-Månsson, A. H. Karlsson, and A. Andrén (2006). Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *J. Dairy Sci.* 89.9, pp. 3296–305.

Wilmink, J. B. M. (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16.4, pp. 335–348.

Chapter 6

General Discussion

In this thesis, genetic backgrounds of bovine α_{s1}- and α_{s2}-casein (α_{s1}- and α_{s2}-CN) phosphorylation were investigated in French Montbéliarde and Dutch Holstein Friesian. We characterized phosphorylation profiles of α_{s1}- and α_{s2}-CN by liquid chromatography/electrospray ionization-mass spectrometry (LC/ESI-MS) and propose at least two regulatory systems responsible for phosphorylation of α_{s1}- and α_{s2}-CN in French Montbéliarde based on phenotypic correlations between traits (**Chapter 2**). Furthermore, the presence of substantial genetic variation in α_{s1}- and α_{s2}-CN phosphorylation isoforms and their phosphorylation degrees (PD), and the effects of genetic variants of major milk proteins on these traits in French Montbéliarde are described (**Chapter 3**). We also characterized phosphorylation profiles of α_{s2}-CN by capillary zone electrophoresis (CZE) in Dutch Holstein Friesian, and results show substantial genetic variation in α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN (α_{s1}-CN PD and α_{s2}-CN PD) also in this breed (**Chapter 4**). Furthermore, we identified 4 genomic regions (BTA1, 6, 11, and 14) associated with α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN in Dutch Holstein Friesian (**Chapter 5**).

In this general discussion, I will discuss the possible causes for the differences between studies dealing with milk protein composition, phosphorylation mechanism and prospects for future research.

6.1 What are the causes for the discrepancies between studies of milk protein composition?

In this thesis, we carried out two studies for exploring the genetic backgrounds of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and for their phosphorylation degrees. In Chapter 2 and 3, we used LC-ESI/MS to quantify the six major milk proteins and α_{s1}- and α_{s2}-CN phosphorylation isoforms and estimated their genetic parameters in French Montbéliarde. In Chapter 4 and 5, we used CZE to quantify α_{s1}- and α_{s2}-CN phosphorylation isoforms and estimated their genetic parameters in Dutch Holstein Friesian. Therefore, we used two analytical methods to quantify α_{s1}- and α_{s2}-CN phosphorylation isoforms, and two trait definitions for defining the phosphorylation degrees of α_{s1}- and α_{s2}-CN in the two breeds. These two data sets allow us to explore how results can vary due to these differences. Differences between analytical methods, trait definitions and breeds coincide with the differences between the two studies. I will discuss in more detail how these factors contribute to the differences between our two studies, such as analytical methods, trait definitions, and breeds on the estimation of genetic parameters and correlations using the two data sets in this thesis. Such arguments could be extended to the comparison across studies dealing with milk protein composition.

Analytical methods. In Chapter 4, to identify peaks corresponding to individual α_{s2}-CN phosphorylation isoforms, we screened the phosphorylation profiles of α_{s2}-CN from 10 fresh milk samples using the LC/ESI-MS method implemented in Chapter 2 and 3. Besides purifying individual α_{s2}-CN phosphorylation isoforms, we also analyzed the same set of samples by CZE to check if the results from the two methods were comparable. The correlations between the results obtained from the two methods show that the quantification of α_{s1}-CN-8P and -9P, α_{s2}-

CN-10P, -11P and -12P was comparable (Figure 6.1). The correlations between α_{s1}-CN-8P and -9P, α_{s2}-CN-10P, -11P and -12P measured by LC-ESI/MS and CZE were 0.98, 0.76, 0.76, 0.81, and 0.67, respectively. However, α_{s2}-CN-13P and -14P could not be quantified accurately by CZE due to co-migration with other components in milk. Subsequently, the relative concentration of total α_{s2}-CN was estimated as the sum of α_{s2}-CN-10P, -11P and -12P and, therefore, was underestimated in Dutch Holstein Friesian as α_{s2}-CN-13P and -14P account for about 24% of the total α_{s2}-CN concentration in French Montbéliarde based on the results from Chapter 2.

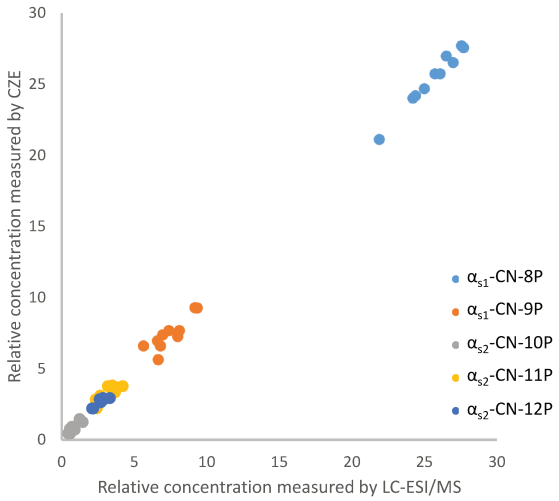

Figure 6.1. Relative concentrations of α_{s1}-CN-8P and -9P, α_{s2}-CN-10P, -11P and -12P from ten milk samples measured by LC-ESI/MS (X axis) and CZE (Y axis), respectively.

Another matter of using CZE is that the relative concentration of total κ-CN was estimated as nonglycosylated mono-phosphorylated form (Heck et al., 2008; Schopen et al., 2009). Glycosylated κ-CN and κ-CN-2P and -3P account for about 35% of total κ-CN based on French Montbéliarde (data not shown from chapter 2). As a result, the relative concentration of total κ-CN was underestimated in Dutch Holstein Friesian. Furthermore, glycosylated κ-CN and κ-CN-2P and -3P partly co-migrate with β-CN variant A1 or A2 in CZE. Consequently, the relative concentration of β-CN was overestimated for Holstein Friesian cows carrying β-CN variant A1 or A2. Therefore, reported phenotypic and genetic correlations of total κ-CN and β-CN with other milk proteins and with α_{s1}- and α_{s2}-CN phosphorylation isoforms in Dutch Holstein Friesian (e.g. in Schopen et al. (2009) and results from Chapter 4) should be interpreted with care, especially for inferring biological relations.

Both methods applied in our studies show advantages and disadvantages. Regarding LC/ESI-MS, it allows to quantify the six major milk proteins including their isoforms and identify genetic variants simultaneously. For the genetic variants or isoforms that have similar retention times on LC, we can still discriminate them based on the molecular masses measured by ESI-MS. This was applied to measure α_{s2}-CN isoforms, i.e. the mass signal intensity was introduced to calculate the proportions of individual isoforms as a fraction of α_{s2}-CN because LC could not completely separate these isoforms. This procedure is more time consuming than a standard

LC or CZE analysis. One possible solution is to optimize the elution conditions to achieve better separation of individual α_{s2}-CN phosphorylation isoforms by LC. Regarding CZE, we achieved a better separation of α_{s2}-CN-10P, -11P and -12P and α_{s1}-CN-8P and -9P compared to LC/ESI-MS. However, we could not accurately measure α_{s2}-CN-13P and -14P by CZE. One possible solution is to prolong the running time in CZE to better separate overlapping peaks.

As described above, the difference in measurements of α_{s1}-CN-8P and -9P, α_{s2}-CN-10P, -11P and -12P, and α_{s1}-CN PD between CZE and LC-ESI/MS in our two studies is negligible. Therefore, the differences in genetic parameters for these traits and correlations among them might be mainly due to the genetic differences between French Montbéliarde and Dutch Holstein Friesian. On the contrary, the differences in measurements of α_{s2}-CN PD (definition excludes α_{s2}-CN-13P and-14P in Dutch Holstein Friesian) and total α_{s2}-, κ- and β-CN between CZE and LC-ESI/MS in our two studies result in changes of definitions of these traits. Thus, the difference in analytical methods might explain a part of the differences in estimated genetic parameters for these traits and correlations among them between French Montbéliarde and Dutch Holstein Friesian. Comparison of estimated genetic parameters and correlations for milk protein composition is rather difficult across studies as well due to different analytical methods used. The differences between LC and CZE are discussed above. Additionally, LC could differ because of differences in separation conditions, especially for measurements of individual isoforms or individual genetic variants. Another difficulty is that traits were expressed in different units across studies. Bobe et al. (1999), Schopen et al. (2009), Gebreyesus et al. (2016), Buitenhuis et al. (2016) and studies in this thesis reported milk protein composition expressed as relative concentrations (as a fraction of total amount of milk protein), whereas Bonfatti et al. (2011) reported milk protein composition expressed as absolute concentrations (as grams per liter of milk) as well as casein composition (individual casein was expressed as a fraction of total amount of casein) and whey composition (β-LG was expressed as a fraction of total amount of whey). Absolute concentration in milk is a result of protein production and relative composition. Therefore, attention is needed when inferring biological relations between traits across studies as traits that expressed in different units result in different biological interpretations. A comparative analysis of measurements of milk protein composition between studies will allow adequate comparison of reported genetic parameters and correlations.

Trait definitions. The α_{s1}- and α_{s2}-CN phosphorylation isoforms are the result of the production of α_{s1}- and α_{s2}-CN and the subsequent phosphorylation. To disentangle the casein production and the phosphorylation process, we defined the phosphorylation degrees of α_{s1}- and α_{s2}-CN (α_{s1}-CN PD and α_{s2}-CN PD) as the proportion of isoforms with higher degrees of phosphorylation according to phenotypic correlations and hierarchical analysis among isoforms (described in Chapter 2). For French Montbéliarde (Chapter 2 and 3), the isoforms phosphorylated at lower degrees were α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P, and the isoforms phosphorylated at higher degrees were α_{s1}-CN-9P, α_{s2}-CN-12P, -13P and -14P. For Dutch Holstein Friesian, α_{s2}-CN PD was defined differently because α_{s2}-CN-13P and -14P were not included in these chapters due to the low reproducibility of their quantification by CZE. For Dutch Holstein Friesian (Chapter 4 and 5), the isoforms phosphorylated at lower degrees were α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P, and the isoforms phosphorylated at higher degrees were α_{s1}-CN-9P

and α_{s2}-CN-12P. Consequently, results regarding α_{s2}-CN PD, such as descriptive statistics and its correlations with other isoforms, for French Montbéliarde (Chapter 2 and 3) and Dutch Holstein Friesian (Chapter 4 and 5) cannot be compared directly. To assess the effects of these different trait definitions on descriptive statistics and correlations, I compared the results using two definitions of total α_{s2}-CN and α_{s2}-CN PD, i.e. the definitions as in Chapter 2 and 3 versus the definitions as in Chapter 4 and 5, with the data from French Montbéliarde.

For total α_{s2}-CN, the phenotypic mean and standard deviation were 8.41 and 0.72, respectively, when including α_{s2}-CN-13P and -14P, and were 6.43 and 0.75, respectively, when excluding α_{s2}-CN-13P and -14P. For α_{s2}-CN PD, the phenotypic mean and standard deviation were 57.05 and 8.36, respectively, when including α_{s2}-CN-13P and -14P, and were 42.10 and 6.39, respectively, when excluding α_{s2}-CN-13P and -14P. The averages for both traits decreased by about 25% when excluding α_{s2}-CN-13P and -14P. When α_{s2}-CN-13P and -14P were excluded from calculating total α_{s2}-CN, correlations of total α_{s2}-CN with α_{s1}-CN-9P, α_{s2}-CN-10P, α_{s2}-CN-11P and α_{s1}-CN PD became stronger (Table 6.1). Regarding the correlations with α_{s1}-CN-9P and α_{s1}-CN PD, in Chapter 2 it was shown that α_{s1}-CN-9P correlated negatively with α_{s2}-CN-10P and -11P (-0.68 and -0.67, respectively) and positively with α_{s2}-CN-13P and -14P (0.50 and 0.49, respectively) in French Montbéliarde. Consequently, when α_{s2}-CN-13P and -14P are not part of total α_{s2}-CN, this results in the major part of α_{s2}-CN (α_{s2}-CN-10P and -11P) being negatively correlated with α_{s1}-CN-9P. Thus, it explains the observed stronger negative correlation between total α_{s2}-CN and α_{s1}-CN-9P as well as the stronger negative correlation between total α_{s2}-CN and α_{s1}-CN PD. Regarding the correlations of total α_{s2}-CN with individual α_{s2}-CN isoforms, correlations increased for all traits because of stronger part-whole correlations as the number of components of total α_{s2}-CN decreased. These stronger correlations of total α_{s2}-CN are very similar to the correlations observed for Holstein Friesian using the same trait definition. When α_{s2}-CN-13P and -14P were excluded from calculating α_{s2}-CN PD, correlations of α_{s2}-CN PD with other traits were similar except for the decrease in the correlation between α_{s2}-CN PD and α_{s1}-CN-8P and the increase in the correlation between α_{s2}-CN PD and α_{s2}-CN-12P (Table 6.1). The similarity in correlations results in a similar interpretation, namely that the highly phosphorylated isoforms correlate positively with α_{s2}-CN PD and the less phosphorylated isoforms correlate negatively with α_{s2}-CN PD. Thus, both trait definitions successfully quantify the proportion of isoforms with higher degrees of phosphorylation. The decrease in the correlation between α_{s2}-CN PD and α_{s1}-CN-8P is mainly due to ignoring the moderate negative correlations of α_{s1}-CN-8P with α_{s2}-CN-13P and -14P according to results reported in Chapter 2. The increase in the correlation between α_{s2}-CN PD and α_{s2}-CN-12P is mainly due to stronger part-whole correlation as the number of components of total α_{s2}-CN decreased, which increases the contribution of the proportion of α_{s2}-CN-12P in total α_{s2}-CN. The changed correlations of α_{s2}-CN PD are still not similar to the correlations observed for Holstein Friesian using the same trait definition. Using two trait definitions in French Montbéliarde allows us to visualize how correlations change due to difference in measurements resulted from analytical methods used and to compare adequately the correlations between French Montbéliarde and Dutch Holstein Friesian. When using the same trait definitions, the phenotypic correlations among traits in French Montbéliarde became similar or closer to the phenotypic correlations

in Dutch Holstein Friesian, however, the changed correlations of α_{s2}-CN PD with total α_{s1}-CN, α_{s1}-CN-8P α_{s2}-CN-11P and -12P as well as the changed correlations of total α_{s2}-CN with total α_{s1}-CN, α_{s1}-CN-8P and α_{s2}-CN-12P are still not similar to the correlations observed for Holstein Friesian (Table 6.1). Since the measurements of individual α_{s1}-CN and α_{s2}-CN isoforms are similar between LC-ESI/MS and CZE as shown above, the differences in phenotypic correlations might be mainly due to breed differences.

In this section, I show different trait definitions resulting from different analytical methods used can complicate the interpretation of phenotypic (and genetic) correlations for inferring biological relations between traits across studies. The main conclusions for the correlations are the same for both trait definitions applied in French Montbéliarde, which is that α_{s1}-CN PD and α_{s2}-CN PD are positively correlated, the highly phosphorylated isoforms correlate positively and the less phosphorylated isoforms correlate negatively with α_{s2}-CN PD, and total α_{s2}-CN correlates negatively with α_{s2}-CN PD. However, the interpretation of some correlations may change due to different trait definitions, such as the stronger part-whole correlations and correlations between compositional traits (e.g. total α_{s2}-CN) and related components (e.g. individual isoforms). Furthermore, the observed correlations between a ratio and its components are difficult to interpret in the same way as correlations among independently determined variables. Ratio traits with different numerators and denominators (e.g. α_{s2}-CN PD) should not be compared directly across studies and might affect the interpretation of their correlations. In French Montbéliarde, α_{s2}-CN PD stands for the proportion of the isoforms phosphorylated at higher degrees (α_{s2}-CN-12P, -13P and -14P) among the isoforms detected (α_{s2}-CN-10P to α_{s2}-CN-14P), whereas in Dutch Holstein Friesian, it stands for the proportion of α_{s2}-CN-12P among the isoforms detected (α_{s2}-CN-10P, -11P and -12P). It was shown above that both trait definitions in our two studies successfully quantify the proportion of isoforms with higher degrees of phosphorylation, however, the biological interpretations of these two trait definitions are somehow different. Therefore, biological relations inferred from this type of correlations should be interpreted with care when comparing studies that use different analytical methods.

Table 6.1. Phenotypic correlations of total α_{s2}-CN (% wt/wt) and α_{s2}-CN PD with individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, total α_{s1}-CN (% wt/wt) and α_{s1}-CN PD[a] in Dutch Holstein Friesian (NL) and in French Montbéliarde (F) using different trait definition

Trait	α_{s1}-CN			α_{s2}-CN			PD	
	Total	8P	9P	10P	11P	12P	α_{s1}-CN	α_{s2}-CN
F: α_{s2}-CN[b]	0.07(0.04)	0.22(0.04)	-0.32(0.04)	0.49(0.03)	0.66(0.02)	0.17(0.04)	-0.34(0.04)	-0.56(0.03)
F: α_{s2}-CN excl[c]	-0.02(0.04)	0.17(0.04)	-0.61(0.03)	0.78(0.02)	0.84(0.01)	0.30(0.04)	-0.62(0.03)	-0.52(0.03)
NL: α_{s2}-CN excl[c]	-0.54 (0.03)	-0.28(0.03)	-0.63(0.03)	0.77(0.01)	0.86(0.01)	0.54(0.02)	-0.61(0.02)	-0.68(0.03)
F: α_{s2}-CN PD[b]	-0.11(0.04)	-0.40(0.04)	0.59(0.03)	-0.68(0.02)	-0.75(0.02)	0.34(0.04)	0.72(0.02)	
F: α_{s2}-CN PD excl[c]	0.11(0.04)	-0.17(0.04)	0.63(0.04)	-0.62(0.03)	-0.85(0.01)	0.66(0.02)	0.63(0.03)	
NL:α_{s2}-CN PD excl[c]	0.58(0.02)	0.28(0.03)	0.75(0.01)	-0.82(0.01)	-0.61(0.02)	0.19(0.03)	0.61(0.03)	

[1] α_{s1}-CN PD= α_{s1}-CN-9P /(α_{s1}-CN-8P + α_{s1}-CN-9P) x 100.
[2] α_{s2}-CN (total)= α_{s2}-CN-10P+ α_{s2}-CN-11P+ α_{s2}-CN-12P+ α_{s2}-CN-13P+ α_{s2}-CN-14P; α_{s2}-CN PD= [(α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P) / total α_{s2}-CN] x 100 in French Montbéliarde.
[3] α_{s2}-CN (total)= α_{s2}-CN-10P+ α_{s2}-CN-11P+ α_{s2}-CN-12P; α_{s2}-CN PD= (α_{s2}-CN-12P / total α_{s2}-CN) x 100 in Dutch Holstein Friesian.

Breeds. As discussed above, differences in analytical methods and trait definitions for α_{s1}-CN-8P and -9P, α_{s2}-CN-10P, -11P and -12P, and α_{s1}-CN PD between our two studies are negligible. Therefore, differences in estimated genetic parameters and correlations for these traits between French Montbéliarde (Chapter 3) and Dutch Holstein Friesian (Chapter 4) might be mainly due to breed differences. Here, I will illustrate differences in estimated genetic parameters and correlations due to breed differences using heritability estimates as an example. To allow comparison of heritability estimates also for α_{s2}-CN PD in the two breeds, I re-estimated the heritability for α_{s2}-CN PD in French Montbéliarde using the same definition as in Dutch Holstein Friesian. The heritability estimate slightly decreased from 0.23 to 0.19 in French Montbéliarde when excluding α_{s2}-CN-13P and -14P.

Breed differences can be the result of differences in allele frequencies of genes affecting the traits of interest. Take the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism as an example. Our GWAS results for individual α_{s2}-CN isoforms, α_{s1}-CN PD and α_{s2}-CN PD (Chapter 5) together with the GWAS results from Bijl et al. (2014a) show that DGAT1 K232A genotypes affect relative concentrations of α_{s1}-CN-9P, α_{s2}-CN-10P and -11P, α_{s1}-CN PD and α_{s2}-CN PD in Dutch Holstein Friesian. Since DGAT1 K232A is nearly fixed at the A allele in French Montbéliarde (0.96, (Gautier et al., 2007), the DGAT1 polymorphism does not contribute to the variation in these traits in French Montbéliarde. To assess the differences in genetic parameters due to this breed difference, I re-estimated the inter-herd heritability after adjusting for the DGAT1 genotypes for the relative concentrations of α_{s1}-CN-9P, α_{s2}-CN-10P and -11P, α_{s1}-CN PD and α_{s2}-CN PD in Dutch Holstein Friesian. The estimates of inter-herd heritability for these traits decrease from 0.69 to 0.60 for α_{s1}-CN-9P, 0.49 to 0.39 for α_{s2}-CN-10P, 0.82 to 0.77 for α_{s2}-CN-11P, 0.75 to 0.71 for α_{s1}-CN PD, and 0.41 to 0.39 for α_{s2}-CN PD in Dutch Holstein Friesian. For α_{s1}-CN-9P and α_{s2}-CN PD, the estimates of inter-herd heritability after accounting for DGAT1 genotypes in Dutch Holstein Friesian are still in range but closer to the values in French Montbéliarde [0.56 and 0.19 (using the same definition as in Dutch Holstein Friesian), respectively]. This suggests that the differences in heritability estimates between French Montbéliarde and Dutch Holstein Friesian for α_{s1}-CN-9P and α_{s2}-CN PD can largely be explained by the difference in DGAT1 allele frequencies between the two breeds. On the contrary, for α_{s2}-CN-10P and -11P and α_{s1}-CN PD, the estimates of inter-herd heritability after accounting for DGAT1 genotypes in Dutch Holstein Friesian are still much higher than those in French Montbéliarde (0.11, 0.32 and 0.37, respectively). This indicates that the differences between the two breeds for these traits can only be partly explained by the difference in DGAT1 allele frequencies. Consequently, differences in frequencies of other variants affecting these traits might also exist.

Differences in frequencies of linked variants in other genes involved might also exist between the two breeds. Our GWAS results in Chapter 5 together with results from Bijl et al. (2014a) show significant effects of β-LG genotypes on α_{s1}-CN-8P and α_{s2}-CN-11P concentrations, α_{s1}-CN PD and α_{s2}-CN PD in Dutch Holstein Friesian, whereas our results in Chapter 3 show significant effects of β-LG genotypes on both α_{s1}-CN-8P and α_{s1}-CN-9P concentrations and no significant effects on individual α_{s2}-CN isoforms, α_{s1}-CN PD and α_{s2}-CN PD in French Montbéliarde. The frequency of β-LG A and B alleles were 0.52 and 0.48 in Dutch Holstein

Friesian, respectively (Bijl et al., 2014a) , whereas the frequencies were 0.44 and 0.55 in French Montbéliarde, respectively (Chapter 3). Since α_{s1}-CN-8P is affected by β-LG genotypes in both breeds, I used α_{s1}-CN-8P as an example to assess if the small difference in the frequencies of β-LG genotypes between the breeds contributes to the difference in heritability estimates between the two breeds. After adjusting for the β-LG genotypes for estimating the inter-herd heritability in both breeds, the heritability estimate decreased from 0.84 to 0.80 in French Montbéliarde and from 0.48 to 0.43 in Dutch Holstein Friesian. These heritability estimates are still different, indicating that the difference between the two breeds for this trait can only partly be explained by differences in β-LG allele frequencies. As the reported effect of the β-LG BB genotype on β-LG concentration for French Montbéliarde (Chapter 3) was about 1.5 times larger than the one reported for Dutch Holstein Friesian (Heck et al., 2009), it might be possible that β-LG genotypes in both breeds are linked with one or multiple variants that also affect individual α_{s1}- and α_{s2}-CN isoforms and their phosphorylation degrees (Ganai et al., 2009; Bedere and Bovenhuis, 2017). Hence, the differences between the effects of β-LG genotypes on individual α_{s1}- and α_{s2}-CN isoforms, α_{s1}-CN PD and α_{s2}-CN PD in the two breeds could be due to the differences between the two breeds in the frequencies of such linked variants. Obtaining genotypes of these variants in both breeds can help to provide insight into the differences in the effects of β-LG genotypes on α_{s1}- and α_{s2}-CN phosphorylation between the two breeds. To conclude, breed differences can be the results of either differences in allele frequencies of genes affecting the traits of interest or differences in allele frequencies of one or multiple linked variants, thus, their effects might vary across breeds and populations due to differences in linkage disequilibrium. These breed differences contribute to the differences in estimated genetic parameters and can be extended to the differences in estimated correlations across studies.

6.2 Phosphorylation of α_{s1}- and α_{s2}-CN

In this thesis, we explored the relations between individual α_{s1}- and α_{s2}-CN phosphorylation isoforms in French Montbéliarde and Dutch Holstein Friesian. In Chapter 2, based on phenotypic correlations between individual α_{s1}- and α_{s2}-CN phosphorylation isoforms in French Montbéliarde, we show that at least two different regulatory systems are responsible for phosphorylation of α_{s1}- and α_{s2}-CN. One system might be involved in regulating isoforms with a lower degree of phosphorylation (α_{s1}-CN-8P, α_{s2}-CN-10P, and α_{s2}-CN-11P), whereas another system favors a higher degree of phosphorylation (α_{s1}-CN-9P, α_{s2}-CN-12P, α_{s2}-CN-13P, and α_{s2}-CN-14P). This hypothesis agrees with Bijl et al. (2014a), who showed that two different chromosomal regions are associated with α_{s1}-CN-8P and α_{s1}-CN-9P concentration, respectively. In Dutch Holstein Friesian, the negative phenotypic and genetic correlations of α_{s1}-CN-9P with α_{s2}-CN-10P and α_{s2}-CN-11P are also observed, suggesting an antagonistic regulation between α_{s1}-CN phosphorylated at a higher degree and α_{s2}-CN phosphorylated at a lower degree. (Chapter 4). Furthermore, the highly positive phenotypic and genetic correlations between α_{s1}-CN PD and α_{s2}-CN PD show that the phosphorylation of α_{s1}- and α_{s2}-CN are related (Chapter 4). I will discuss possible mechanisms for regulating the phosphorylation of α_{s1}- and α_{s2}-CN in the following section.

Regulatory systems of casein phosphorylation. In Chapter 5, we showed three major genomic regions that influence the phosphorylation degrees of α_{s1}- and α_{s2}-CN on BTA1 harboring *SLC37A1* (a phosphorus antiporter), BTA11 harboring the gene for β-LG, and BTA14 harboring *DGAT1* in Dutch Holstein Friesian. The estimated genotype effects of SNP BTB-00068200 (BTA1), β-LG variants A and B, and DGAT1 K232A support two main regulatory systems of α_{s1}- and α_{s2}-CN phosphorylation (see Table 5.2). The effects of SNP BTB-00068200 and DGAT1 on α_{s1}-CN-9P are in the opposite direction of their effects on α_{s2}-CN-10P and -11P, suggesting that the regions on BTA1 and BTA14 contribute to an antagonistic regulation between α_{s1}-CN phosphorylated at a higher degree and α_{s2}-CN phosphorylated at a lower degree. Contrastingly, the effects of β-LG on α_{s1}-CN-8P are in the same direction as its effects on α_{s2}-CN-10P and -11P, but no effects were detected on α_{s1}-CN-9P and α_{s2}-CN-12P, suggesting that the region on BTA11 is only involved in the regulation of α_{s1}- and α_{s2}-CN isoforms with a lower degree of phosphorylation. Taken together, these results support the hypothesis proposed in Chapter 2 that one regulatory system is more effective than the other depending on the individual cow, meaning either α_s-CN isoforms phosphorylated at lower degrees would be more abundant than those phosphorylated at higher degrees or vice versa.

The actual roles of SLC37A1, β-LG and DGAT1 in α_{s1}- and α_{s2}-CN phosphorylation in Dutch Holstein Friesian require further investigation. Our results from Chapter 5 suggest that effects of identified genomic regions on α_{s1}-CN PD and α_{s2}-CN PD are probably due to changes in milk synthesis and phosphorus secretion in milk probably because different pathways of milk production, including milk protein synthesis and phosphorus secretion in milk, are interrelated. In support of this hypothesis, we show in Chapter 5 that SNP BTB-00068200 (BTA1) affects α_{s1}-CN PD, α_{s2}-CN PD and mainly inorganic phosphorus but not the amount of phosphate groups attached to caseins. Dietary phosphorus in cow's feed usually exceeds the requirements of the recommendation in the Netherlands (Valk et al., 2002). phosphorus-deficient cows could improve our understanding of the effect of changes in phosphorus output on α_{s1}- and α_{s2}-CN phosphorylation, as phosphorus-deficient cows show decreased phosphorus output in milk (Valk et al., 2002). Therefore, comparing milk composition of cows before phosphorus deficiency and after, especially the phosphorylation degrees of α_{s1}- and α_{s2}-CN and phosphorus distribution in milk (milk serum, casein micelles, phosphate groups attached to caseins, etc.), might provide insight into the impact of a change in phosphorus output in milk on α_{s1}- and α_{s2}-CN phosphorylation.

Casein kinases. Fam20C was recently discovered as the kinase that phosphorylates secretory pathway proteins with Ser-X-Glu/pSer motifs (X represents any amino acid residue, and p indicates phosphorylation) including caseins in milk as well as several other proteins implicated in biomineralization (Tagliabracci et al., 2012). The *FAM20C* gene is located between 43.86 to 43.90 Mbp (BTAU 4.0) on BTA25. However, our GWAS did not detect a QTL signal for individual α_{s1}- and α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD or α_{s2}-CN PD at that position, suggesting that no *FAM20C* variants segregate in the Dutch Holstein Friesian population or their effects are too small to be detected (Chapter 5). All caseins belong to the secretory calcium-binding phosphoprotein gene family, which includes bone, tooth, milk and salivary proteins that are involved in many aspects of the control of biomineralisation (Kawasaki et al., 2011), and

Fam20C is ubiquitously expressed at high levels in mineralized tissues and in mammary gland (Nalbant et al., 2005; Wang et al., 2010; Hao et al., 2007; Lizio et al., 2015). As such, Fam20C is a good candidate for the kinase that phosphorylates caseins physiologically as it has been shown to be secreted and present in milk (Tagliabracci et al., 2012). Interestingly, for α_{s2}-CN to be able to carry more than 12 phosphate groups, Fam20C would have to phosphorylate threonine residues located among the 2 Thr-X-Asp and 4 Thr-X-Glu motifs as only 12 serine residues are available in the α_{s2}-CN protein sequence (see Figure 1.2, consensus motif as Ser/Thr-X-Glu/pSer/Asp; (Mercier, 1981; Meggio et al., 1988; Lasa-Benito et al., 1996). Threonine in Thr-X-Glu motifs has been shown phosphorylated in caprine, human, and equine β-CN when they are phosphorylated at the highest level (Greenberg et al., 1984; Neveu et al., 2002; Matéos et al., 2010). However, Thr-X-Asp/Glu motifs have not been proven as substrates of Fam20C. This suggests that another kinase would be responsible for phosphorylating Thr-X-Asp/Glu motifs. Additional indication for the existence of another kinase is α_{s1}-CN-9P that carries an extra phosphorylated serine at Ser41 in the Ser-X-Asp motif of the mature peptide chain (Manson et al., 1977). Also this Ser-X-Asp motif has not been proven as substrate of Fam20C. Furthermore, results from previous studies attempting to purify kinases from the Golgi-enriched fraction of the lactating mammary gland suggest the existence of two related, but distinct, kinase activities (Brooks, 1989; Duncan et al., 2000). Taken together, the hypothesis of two kinases phosphorylating caseins can explain the two main regulatory systems proposed in this thesis. It cannot be excluded that Fam20C has broader substrate specify as shown by Tagliabracci et al. (2015). However, the new recognition motifs detected in that study exist rarely in protein sequences of caseins and do not include Thr-X-Asp/Glu motifs. Identifying phosphorylation sites of α_{s2}-CN using LC-MS/MS can provide more information about substrate specificity of casein-phosphorylating kinases. Additionally, studying the substrate specificity of Fam20C and its role in casein phosphorylation can also provide insight into the phosphorylation mechanisms of caseins. To confirm if Fam20C phosphorylates Ser in the Ser-X-Asp motif in α_{s1}-CN, one approach can be the incubation of Fam20C with recombinant (non-phosphorylated) α_{s1}-CN produced by *Escherichia coli* and subsequently examine if Ser in the Ser-X-Asp motif is phosphorylated by Fam20C after incubation. Another approach is to establish an in vitro model system suitable for studying the role of Fam20C in casein phosphorylation. An ideal in vitro model is a cell line that expresses and secretes both Fam20C and phosphorylated caseins. In this way, generating Fam20C-knock-out cells and analyzing the phosphorylation profiles of the caseins secreted from the knock-out cells can provide insight into the role of Fam20C in casein phosphorylation.

6.3 Can phosphorylation degrees of α_{s1}- and α_{s2}-CN be predicted using genotype information?

Studying the variation in the phosphorylation degrees of α_{s1}- and α_{s2}-CN may have an impact through two avenues. Firstly, it will improve our fundamental knowledge of the biological processes in the synthesis of milk proteins. Secondly, the change of the phosphorylation degrees of α_{s1}- and α_{s2}-CN might have impact on milk technological properties according to previous

studies (Frederiksen et al., 2011; Jensen et al., 2012; Bijl et al., 2014b). The evidence is currently limited, and it requires further research to identify potential benefits for the dairy industry. As α_{s1}- and α_{s2}-CN are more highly phosphorylated and possess multiple phosphoserine clusters than β- and κ-CN, they may be more relevant for stabilizing internal micellar structure (Dalgleish and Corredig, 2012; Huppertz et al., 2017). Furthermore, when comparing the micelle structure between African elephant (milk contains only β- and κ-CN) and cow, the elephant micelles seem to be less stable and more easily disturbed than the cow micelles, suggesting phosphoserine clusters of α_{s1}- and α_{s2}-CN are highly relevant for stabilizing calcium phosphate nanoclusters (Madende, 2017). Thus, investigating relations between phosphorylation degrees of α_{s1}- and α_{s2}-CN and micelle structure could provide a better understanding of the role of the phosphorylation degree in the formation of casein micelles and the potential impact on milk technological properties. To do this, the next step could be comparing micelle properties from cows that produce milk with contrasting phosphorylation degrees of α_{s1}- and α_{s2}-CN. Furthermore, comparing transcriptome profiles of cows exhibiting contrasting phosphorylation degrees of α_{s1}- and α_{s2}-CN could help to identify genes that contribute to the regulation of α_{s1}- and α_{s2}-CN phosphorylation. As described above, we detected three major genomic regions on BTA1, 11 and 14 that influence the phosphorylation degrees of α_{s1}- and α_{s2}-CN in Dutch Holstein Friesian. This information could be used to pre-select cows based on their genotypes of markers for further studies, so that it is not needed to analyze milk protein composition in a new large population for selecting contrasting phenotypes. Furthermore, costs of protein analyses and genotyping are similar, but genotypic is less time consuming compared to proteomic tools. I will show here if it is possible to pre-select Dutch Holstein Friesian cows based on their genotypes of SNP BTB-00068200, β-LG variants A and B, and DGAT1 K232A.

Phosphorylation profiles of α_{s1}- and α_{s2}-CN vary considerably between cows as shown in Chapter 2 and 4 (illustrated in Figure 6.2). The principal component analysis of phosphorylation profiles of α_{s1}- and α_{s2}-CN in Dutch Holstein Friesian shows that the first principal component explaining about 57% of the variation represents the phosphorylation degrees of α_{s1}- and α_{s2}-CN, and the second principal component explaining about 18% of the variation represents α_{s2}-CN production (Figure 6.3A). Here, I used histograms of α_{s2}-CN PD of individual cows for illustrating that distinguishing contrasting phosphorylation degrees based on a cow's genotypes of SNP BTB-00068200, β-LG variants A and B, and DGAT1 K232A is difficult (Figure 6.3B-D). This is expected as these three markers explain together 20-25% of the additive genetic variance of α_{s1}-CN PD and α_{s2}-CN PD (Chapter 5). Therefore, analyzing milk protein composition is still required for searching cows exhibiting contrasting phosphorylation degrees of α_{s1}- and α_{s2}-CN in a different Dutch Holstein Friesian population for future studies. An alternative is to study whether phosphorylation degrees can be predicted with 50K SNP or 777K SNP using genomic prediction methods as they also contain markers that explain small parts of the genetic variation (Meuwissen et al., 2001).

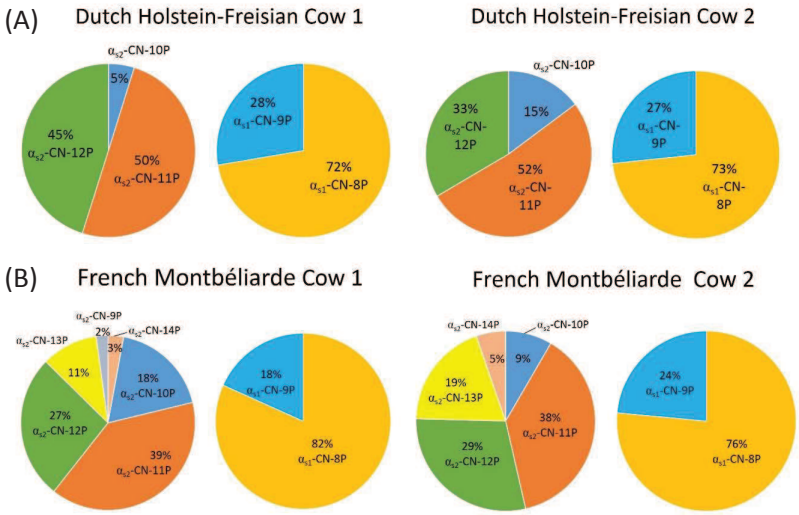

Figure 6.2. Distinct α_{s1}- and α_{s2}-CN phosphorylation profiles of two Dutch Holstein-Friesian (A) and two French Montbéliarde (B) cows. Cows correspond to the 5% and 95% quantile of the mahalanobis distance calculated based on their phosphorylation profiles.

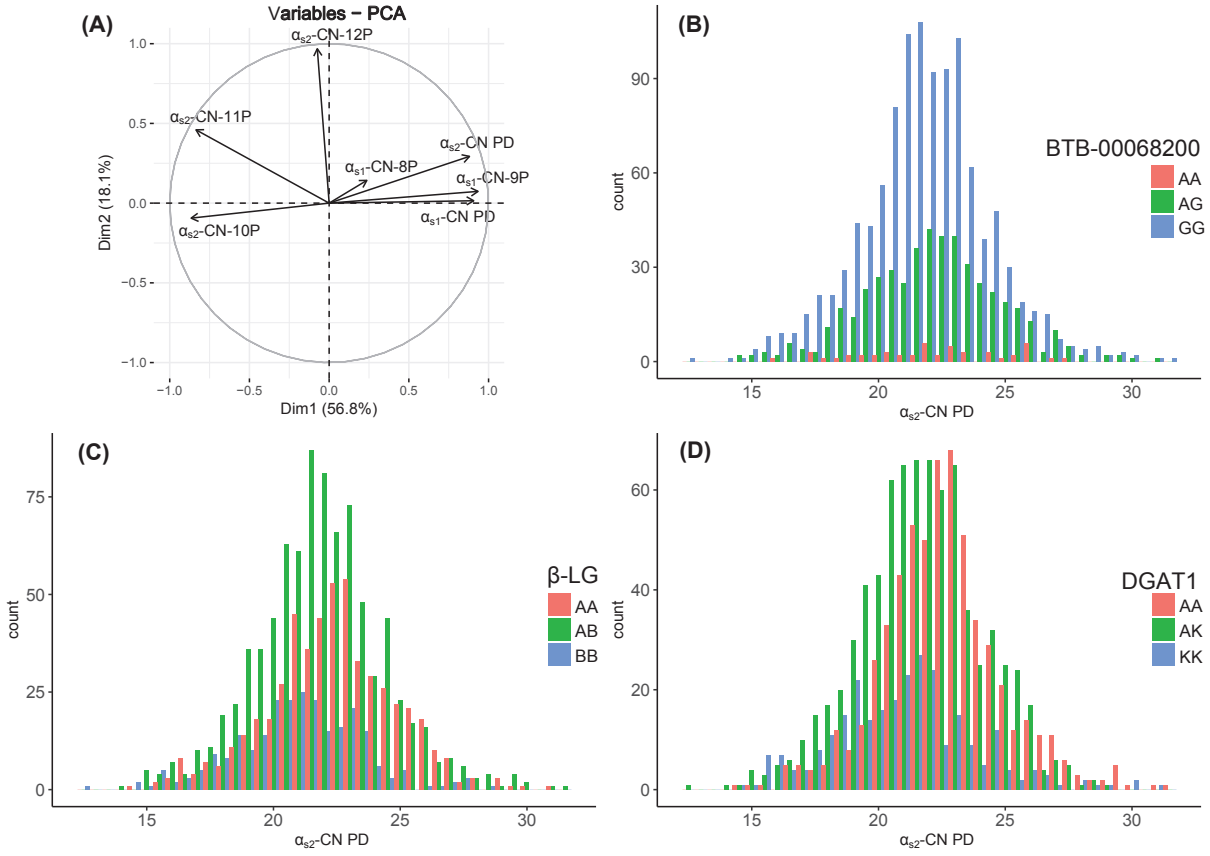

Figure 6.3. (A) Principal component analysis of phosphorylation profiles of α_{s1}- and α_{s2}-CN in Dutch Holstein Friesian. Histogram of α_{s2}-CN PD of Dutch Holstein-Friesian cows grouped by their genotypes of SNP BTB-00068200 (B), β-LG variants A and B (C) and DGAT1 K232A (D). α_{s2}-CN PD= α_{s2}-CN-12P / (α_{s2}-CN-10P+ α_{s2}-CN-11P+ α_{s2}-CN-12P) x 100.

6.4 Concluding remarks

In this thesis, I characterized the phosphorylation profiles of α_{s1}- and α_{s2}-CN and showed that considerable phenotypic and genetic variation exist in individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN in two dairy cattle breeds. The relations between individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and genomic regions detected for individual isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN suggest two main regulatory systems for phosphorylation of α_{s1}- and α_{s2}-CN, providing insight in how phosphorylation of α_{s1}- and α_{s2}-CN is regulated.

References

Bedere, N. and H. Bovenhuis (2017). Characterizing a region on BTA11 affecting β-lactoglobulin content of milk using high-density genotyping and haplotype grouping. *BMC Genet.* 18.1, p. 17.

Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk, and H. Bovenhuis (2014a). Phosphorylation of αS1-casein is regulated by different genes. *J. Dairy Sci.* 97.11, pp. 7240–7246.

Bijl, E., H. van Valenberg, S. Sikkes, S. Jumelet, G. Sala, K. Olieman, T. van Hooijdonk, and T. Huppertz (2014b). Chymosin-induced hydrolysis of caseins: Influence of degree of phosphorylation of alpha-s1-casein and genetic variants of beta-casein. *Int. Dairy J.* 39.2, pp. 215–221.

Bobe, G., D. Beitz, A. Freeman, and G. Lindberg (1999). Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. *J. Dairy Sci.* 82.12, pp. 2797–2804.

Bonfatti, V., A. Cecchinato, L. Gallo, A. Blasco, and P. Carnier (2011). Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. *J. Dairy Sci.* 94.10, pp. 5183–5193.

Brooks, C. L. (1989). Two physiological substrate-specific casein kinases are present in the bovine mammary gland. *FEBS Lett.* 243.2, pp. 385–8.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen (2016). Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genet.* 17.

Dalgleish, D. G. and M. Corredig (2012). The structure of the casein micelle of milk and its changes during processing. *Annu. Rev. Food Sci. Technol.* 3.1, pp. 449–467.

Duncan, J. S., M. C. Wilkinson, and R. D. Burgoyne (2000). Purification of Golgi casein kinase from bovine milk. *Biochem. J.* 350 Pt 2, pp. 463–8.

Frederiksen, P. D., K. K. Andersen, M. Hammershøj, H. D. Poulsen, J. Sørensen, M. Bakman, K. B. Qvist, and L. B. Larsen (2011). Composition and effect of blending of noncoagulating, poorly coagulating, and well-coagulating bovine milk from individual Danish Holstein cows. *J. Dairy Sci.* 94.10, pp. 4787–4799.

Ganai, N. A., H. Bovenhuis, J. A. M. van Arendonk, and M. H. P. W. Visker (2009). Novel polymorphisms in the bovine beta-lactoglobulin gene and their effects on beta-lactoglobulin protein concentration in milk. *Anim. Genet.* 40.2, pp. 127–133.

Gautier, M., A. Capitan, S. Fritz, A. Eggen, D. Boichard, and T. Druet (2007). Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphisms in French dairy cattle. *J. Dairy Sci.* 90.6, pp. 2980–2988.

Gebreyesus, G., M. S. Lund, L. Janss, N. A. Poulsen, L. B. Larsen, H. Bovenhuis, and A. J. Buitenhuis (2016). Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the Danish Holstein. *J. Dairy Sci.* 99.4, pp. 2863–2866.

Greenberg, R., M. L. Groves, and H. J. Dower (1984). Human beta-casein. Amino acid sequence and identification of phosphorylation sites. *J. Biol. Chem.* 259.8, pp. 5132–5138.

Hao, J., K. Narayanan, T. Muni, A. Ramachandran, and A. George (2007). Dentin matrix protein 4, a novel secretory calcium-binding protein that modulates odontoblast differentiation. *J. Biol. Chem.* 282.21, pp. 15357–15365.

Heck, J. M. L., A. Schennink, H. J. F. Van Valenberg, H. Bovenhuis, M. H. P. W. Visker, J. A. M. Van Arendonk, and A. C. M. van Hooijdonk (2009). Effects of milk protein variants on the protein composition of bovine milk. *J. Dairy Sci.* 92.3, pp. 1192–1202.

Heck, J., C. Olieman, A. Schennink, H. van Valenberg, M. Visker, R. Meuldijk, and A. van Hooijdonk (2008). Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *Int. Dairy J.* 18.5, pp. 548–555.

Huppertz, T., I. Gazi, H. Luyten, H. Nieuwenhuijse, A. Alting, and E. Schokker (2017). Hydration of casein micelles and caseinates: Implications for casein micelle structure. *Int. Dairy J.* 74, pp. 1–11.

Jensen, H. B., N. A. Poulsen, K. K. Andersen, M. Hammershøj, H. D. Poulsen, and L. B. Larsen (2012). Distinct composition of bovine milk from Jersey and Holstein-Friesian cows with good, poor, or noncoagulation properties as reflected in protein genetic variants and isoforms. *J. Dairy Sci.* 95.12, pp. 6905–6917.

Kawasaki, K., A. G. Lafont, and J. Y. Sire (2011). The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol. Biol. Evol.* 28.7, pp. 2053–2061.

Lasa-Benito, M., O. Marin, F. Meggio, and L. A. Pinna (1996). Golgi apparatus mammary gland casein kinase: monitoring by a specific peptide substrate and definition of specificity determinants. *FEBS Lett.* 382.1-2, pp. 149–52.

Lizio, M., J. Harshbarger, H. Shimoji, J. Severin, T. Kasukawa, S. Sahin, I. Abugessaisa, S. Fukuda, F. Hori, S. Ishikawa-Kato, C. J. Mungall, E. Arner, J. Baillie, N. Bertin, H. Bono, M. de Hoon, A. D. Diehl, E. Dimont, T. C. Freeman, K. Fujieda, W. Hide, R. Kaliyaperumal, T. Katayama, T. Lassmann, T. F. Meehan, K. Nishikata, H. Ono, M. Rehli, A. Sandelin, E. A. Schultes, P. 't Hoen, Z. Tatum, M. Thompson, T. Toyoda, D. W. Wright, C. O. Daub, M. Itoh, P. Carninci, Y. Hayashizaki, A. Forrest, and H. Kawaji (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 16.1, p. 22.

Madende, M. (2017). "Elucidation of African elephant beta casein phosphorylation state and casein micelle structure. PhD Thesis". PhD Thesis. University of the Free State.

Manson, W., T. Carolan, and W. D. Annan (1977). Bovine αS0 casein; a phosphorylated homologue of αS1 casein. *Eur. J. Biochem.* 417, pp. 411–417.

Matéos, A., J. Girardet, D. Mollé, C. Corbier, J. Gaillard, and L. Miclo (2010). Identification of phosphorylation sites of equine β-casein isoforms. *Rapid Commun. Mass Spectrom.* 24.11, pp. 1533–1542.

Meggio, F., A. P. Boulton, F. Marchiori, G. Borin, D. P. Lennon, A. Calderan, and L. A. Pinna (1988). Substrate-specificity determinants for a membrane-bound casein kinase of lactating mammary gland. A study with synthetic peptides. *Eur. J. Biochem.* 177.2, pp. 281–4.

Mercier, J. C. (1981). Phosphorylation of caseins, present evidence for an amino acid triplet code posttranslationally recognized by specific kinases. *Biochimie* 63.1, pp. 1–17.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157.4, pp. 1819–29.

Nalbant, D., H. Youn, S. I. Nalbant, S. Sharma, E. Cobos, E. G. Beale, Y. Du, and S. C. Williams (2005). FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. *BMC Genomics* 6.1, p. 11.

Neveu, C., D. Mollé, J. Moreno, P. Martin, and J. Léonil (2002). Heterogeneity of caprine beta-casein elucidated by RP-HPLC/MS: genetic variants and phosphorylations. *J. Protein Chem.* 21.8, pp. 557–567.

Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. van Valenberg, and J. A. M. van Arendonk (2009). Genetic parameters for major milk proteins in Dutch Holstein-Friesians. *J. Dairy Sci.* 92.3, pp. 1182–1191.

Tagliabracci, V. S., J. L. Engel, J. Wen, S. E. Wiley, C. A. Worby, L. N. Kinch, J. Xiao, N. V. Grishin, and J. E. Dixon (2012). Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* 336.6085, pp. 1150–3.

Tagliabracci, V. S., S. E. Wiley, X. Guo, L. N. Kinch, E. Durrant, J. Wen, J. Xiao, J. Cui, K. B. Nguyen, J. L. Engel, J. J. Coon, N. Grishin, L. A. Pinna, D. J. Pagliarini, and J. E. Dixon (2015). A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* 161.7, pp. 1619–1632.

Valk, H., L. Šebek, and A. Beynen (2002). Influence of Phosphorus Intake on Excretion and Blood Plasma and Saliva Concentrations of Phosphorus in Dairy Cows. *J. Dairy Sci.* 85.10, pp. 2642–2649.

Wang, X., J. Hao, Y. Xie, Y. Sun, B. Hernandez, A. K. Yamoah, M. Prasad, Q. Zhu, J. Q. Feng, and C. Qin (2010). Expression of FAM20C in the osteogenesis and odontogenesis of mouse. *J. Histochem. Cytochem.* 58.11, pp. 957–967.

Summary

Phosphorylation of caseins (CN) is a crucial post-translational modification that allows caseins to interact with calcium phosphate to form large colloidal structures called casein micelles. The formation and stability of casein micelles are highly relevant to transportation of abundant minerals to the neonate and manufacturing of dairy products. Therefore, it is of great interest to explore variation in degrees of phosphorylation of caseins, in particular α_{s1}- and α_{s2}-CN which are highly phosphorylated, and study to what extent genetic and other factors contribute to this variation. This thesis aimed to investigate the genetic background of bovine milk protein composition with a focus on the phosphorylation of α_{s1}- and α_{s2}-CN. For this purpose, two studies were conducted to quantify phosphorylation levels of α_{s1}- and α_{s2}-CN: one in French Montbéliarde using liquid chromatography coupled with electrospray ionization mass spectrometry (LC/ESI-MS; **Chapter 2 and 3**) and the other in Dutch Holstein Friesian using capillary zone electrophoresis (CZE; **Chapter 4**). Additionally, genomic regions associated with individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and their phosphorylation degrees were identified in Dutch Holstein Friesian (**Chapter 5**).

The first study in French Montbéliarde consists two parts. Firstly, **Chapter 2** describes the characterization of the bovine α_{s1}-CN and α_{s2}-CN phosphorylation profile using LC-ESI/MS. Three new phosphorylation isoforms were detected: α_{s2}-CN-9P, α_{s2}-CN-14P, and α_{s2}-CN-15P, in addition to the known isoforms α_{s1}-CN-8P, α_{s1}-CN-9P, α_{s2}-CN-10P, α_{s2}-CN-11P, α_{s2}-CN-12P, and α_{s2}-CN-13P. Relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms varied considerably among cows. Phenotypic correlations show that isoforms phosphorylated at higher degrees (*i.e.* α_{s1}-CN-9P, α_{s2}-CN-12P, and α_{s2}-CN-13P and α_{s2}-CN-14P) correlated negatively with isoforms phosphorylated at lower degrees (*i.e.* α_{s1}-CN-8P, α_{s2}-CN-10P and α_{s2}-CN-11P). Secondly, **Chapter 3** investigates to what extent genetic and other factors contribute to the variation in relative concentrations of individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN (α_{s1}-CN PD and α_{s2}-CN PD) defined as the proportion of isoforms phosphorylated at higher degrees in total α_{s1}- and α_{s2}-CN, respectively. The α_{s1}- and α_{s2}-CN phosphorylation profile showed changes across parity and lactation, and exploitable genetic variation for the phosphorylation degrees of α_{s1}- and α_{s2}-CN exists in the French Montbéliarde. Furthermore, the β-CN I variant is associated with a greater proportion of isoforms phosphorylated at lower degrees (α_{s1}-CN-8P, α_{s2}-CN-10P and α_{s2}-CN-11P); the β-CN B variant is associated with a greater proportion of isoforms phosphorylated at higher degrees (α_{s1}-CN-9P, α_{s2}-CN-12P, and α_{s2}-CN-13P and α_{s2}-CN-14P). The second study in Dutch Holstein Friesian also consists of two parts. Firstly, *Chapter 4* describes the characterization of the bovine α_{s2}-CN phosphorylation profile using CZE and estimation of genetic parameters for individual α_{s2}-CN phosphorylation isoforms, α_{s1}-CN PD and α_{s2}-CN PD. Three α_{s2}-CN isoforms, namely α_{s2}-CN-10P, α_{s2}-CN-11P, α_{s2}-CN-12P, and the phosphorylation degrees of α_{s1}- and α_{s2}-CN were quantified. High intra-herd heritabilites for individual α_{s2}-CN phosphorylation isoforms (0.54 to 0.89), α_{s1}-CN PD (0.78) and α_{s2}-CN PD (0.64) suggest that genetic factors contribute substantially to observed differences in α_{s1}- and α_{s2}-CN phosphorylation profiles and the phosphorylation process. Furthermore, the strong positive genetic correlation between α_{s1}-CN PD and α_{s2}-CN PD (0.94) suggests that the phosphorylation degrees of α_{s1}-CN and α_{s2}-CN are related. Secondly, in Chapter 5, the genome-wide association study

reveals a total of 10 chromosomal regions that were associated with relative concentrations of individual α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN. These regions were distributed across chromosome 1, 2, 6, 9, 11, 14, 15, 18, 24 and 28. Regions associated with multiple traits were found on chromosome 1, 6, 11, and 14. Two quantitative trait loci (QTL) regions were detected on BTA1: one affecting α_{s2}-CN production, and the other affecting α_{s1}-CN PD and α_{s2}-CN PD. The QTL region on BTA6 affected only individual α_{s2}-CN isoforms. The QTL regions on BTA11 and 14 affected relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. Results suggested that effects of identified genomic regions on α_{s1}-CN PD and α_{s2}-CN PD are probably due to changes in milk synthesis and phosphorus secretion in milk.

Chapter 6 is the general discussion. The first part addresses the effects of the differences among studies, such as analytical methods, trait definitions, and breed differences on estimated genetic parameters and correlations using dataset from the two studies described in this thesis. Analyzing the same milk samples using LC/ESI-MS and CZE show that the measurements of α_{s1}-CN-8P and -9P, α_{s2}-CN-10P, -11P and -12P, and α_{s1}-CN PD were similar between two analytical methods. For these traits, it was assumed that differences in heritability estimates between French Montbéliarde and Dutch Holstein Friesian were due to breed differences, which might be due to differences in allele frequencies. For α_{s2}-CN PD, it was defined as (α_{s2}-CN-12P + α_{s2}-CN-13P + α_{s2}-CN-14P) / (α_{s2}-CN-10P+ α_{s2}-CN-11P+ α_{s2}-CN-12P+ α_{s2}-CN-13P+ α_{s2}-CN-14P) in French Montbéliarde and as α_{s2}-CN-12P / (α_{s2}-CN-10P+ α_{s2}-CN-11P+ α_{s2}-CN-12P) in Dutch Holstein Friesian. When comparing phenotypic correlations of α_{s2}-CN PD with other traits using both definitions in French Montbéliarde, it is shown that both trait definitions successfully quantified the proportion of isoforms phosphorylated at lower degrees because of the observed similarity in estimated correlations. Furthermore, when using the same trait definition, the phenotypic correlations among traits in French Montbéliarde became similar or closer to the phenotypic correlations in Dutch Holstein Friesian with some exceptions, which might be due to breed differences.

The second part of the general discussion concentrates on possible mechanisms regulating the phosphorylation of α_{s1}- and α_{s2}-CN. The effects of SNP BTB-00068200 and DGAT1 on α_{s1}-CN-9P are in the opposite direction of their effects on α_{s2}-CN-10P and -11P, suggesting that the regions on chromosome 1 and 14 contribute to an antagonistic regulation between α_{s1}-CN phosphorylated at a higher degree and α_{s2}-CN phosphorylated at a lower degree. Contrastingly, the effects of β-LG on α_{s1}-CN-8P are in the same direction as its effects on α_{s2}-CN-10P and -11P, but no effects were detected on α_{s1}-CN-9P and α_{s2}-CN-12P, suggesting that the region on chromosome 11 is only involved in the regulation of α_{s1}- and α_{s2}-CN isoforms phosphorylated at lower degrees. These results indicate that either α_{s1}- and α_{s2}-CN isoforms phosphorylated at lower degrees would be more abundant than those phosphorylated at higher degrees or vice versa. One possible hypothesis is a two-casein-kinase system, meaning one kinase is responsible for forming isoforms phosphorylated at lower degrees, and the other one is responsible for forming isoforms phosphorylated at higher degrees.

The last part of the general discussion is about the relevance of findings described in the thesis for further research. For a better understanding of the role of the phosphorylation de-

gree in the formation of casein micelles and identifying the potential impact on milk technological properties, comparing micelle properties from cows that produce milk with contrasting phosphorylation degrees of α_{s1}- and α_{s2}-CN can provide an insight into relations between phosphorylation degrees of α_{s1}- and α_{s2}-CN and micelle structure. To achieve this, it is shown that analyzing milk protein composition on a large scale is still required for searching cows exhibiting contrasting phosphorylation degrees of α_{s1}- and α_{s2}-CN because using three genetic markers identified in this thesis is not sufficient to predict phosphorylation degrees of α_{s1}- and α_{s2}-CN in Dutch Holstein Friesian.

Acknowledgments/Remerciements

Doing a PhD was just like hopping on infinity roller coaster. It felt like that it would never end during the ride, but in this case, you could not just close your eyes until it was over (people who went to Walibi with me know what I was like on roller coasters :D). As obvious as it is, I wouldn't have survived without all the love and support, so I would like to express my very profound gratitude to all the people who contribute to this PhD and life around it during the past four years.

I'm very grateful for EGS-ABG program including all the coordinators and secretaries for making this PhD happened. A special thanks goes to my supervisors, Patrice, Henk and Marleen for their guidance and support. Without you, this thesis wouldn't have existed. Henk, thanks a lot for all the valuable input, the intellectual challenges and the random chats that sometimes turned into philosophy lessons :) I really appreciate your take on exploring different aspects of analyzing data and science, and that your door is always open when I have questions. I hope I've grown at least 1% like you as a scientist after the last two years we've been working together. Marleen, besides the scientific support, I've learned a lot from you about how to structure a manuscript. Your modifications improved tremendously my manuscripts, and I'm really grateful for that. Thanks a lot for your help for organizing the my last two years.

A big shout out to all my collaborators! Hein, your passion for research is inspiring and contagious, and I really admire that you always take initiatives for collaborations. Without you, I wouldn't have known so many people and participated in interesting projects. Thom, I really enjoyed our collaborations, brainstorming and conversations. We could've done a lot more together if I have more time, but let's leave some for the future research :) Guido, thanks for your assistance for running CZE and LC-MS (for a very short while until the machine broke....) in NIZO. Sjef, thank you for all the help for running LC/MS-MS analysis. Jared, I enjoyed a lot our collaboration though we only shared a pretty short ride together. Your honesty and passion for science inspire me. I really appreciate the time we worked together! I hope there will be much more collaborations in the future and you're drinking G& T while reading it:) I would also like to thank Jeroen Heck and Etske Bijl for helping me to get around with Dutch milk genomic initiative data regarding phosphorylation isoforms.

I would like to thank my colleagues from INRA Jouy-en-Josas. Un grand merci à vous tous et toutes, l'équipe LGS, pour tout votre amitié au labo. Je voudrais remercier quelques personnes en particulier. Guido le magnifico, merci infiniment pour ton soutien pendent ma thèse, de m'avoir supportée au bureau et de m'avoir appris le français et LC-MS! Tu étais toujours là pour m'aider quand j'en avais besoin. Merci encore Guido! Je t'attends toujours pour visiter Taiwan et goûter des yeux de poisson! Agnès, merci beaucoup pour ton soutien surtout pour la partie du projet FROM'MIR. J'apprécie beaucoup ton efficacité et tes commentaires sur mes papiers. Je suis vraiment contente d'avoir travaillé avec toi. Zuzana et Léonardo, merci beaucoup pour votre soutien, votre amitié et les bons moments passé ensemble. Yuli, I would like to thank you for all the advise, scientific discussions and friendship for the time spent in Jouy.

I wish to thank all the EGS-ABG fellows! For some, we only saw each other in conferences and courses, but it was always tremendous fun spending time together! A special thanks to Edin and Bélen, you guys were always there for me when I was panicking and needed moral support and suggestions. I'm really grateful for that!

I would also like to thank all the colleagues from ABGC for the discussions during QDG and TLM, fun activities and drinks. I have learned a lot while having a lot of fun for my two years here! List is really long to go through every name, but I would like to mention a couple of special ones. Thank you Ada and Lisette for helping out all the administration! Amabel, thank you for all the help for programming and fun together. Mandy, the sweetest girl I've ever met, thank you for all the help, always being there for me when I needed to vent and for all the tasty beef jerky you brought whenever you went home. Thank you, Sandrine for always listening to me and helping me with analysis. Juanito, thanks a lot for being a grown-up. You're always being a good friend to me, especially being available for a sympathetic ear. Thank you for all the good (drinking) time together my party people: Pascal, Floor, Claudia, Jérémie, Shuwen, Yvonne, André, Marcos, Merina, Juanito, Amabel, Grummy, Rodrigo, Tom (I would be more grateful if you stop videotaping me all the time at parties), Maria from Spain, Gareth, Crystal, Mathijs, Coralia, Maria from Colombia and Sonia. A big shout out to the best office, Gareth, Maria, Coralia and Sonia! Thanks a lot for all the laughters, chatters, scientific discussions and sharing frustrations. The last stretch of PhD would have been tougher without you guys. Gareth, the hardcore quantitative geneticist, thanks a lot for sharing knowledge and bringing the Thirsty-Thursday spirit! I hope we can do some more together in the future. Cori, thanks a lot for all the fun time, especially girls' nights! I hope we can organize a trip to Mexico soon! Mi amor, thank you for speaking minion and always being supportive even when you're far away in Ethiopia! Sonia, ce fut une période difficile et ton aide fut plus qu'appréciable. C'est si bon de savoir que tu étais présente à mes côtés. Dans les moments où j'étais trop fatiguée, tu étais toujours là pour m'aider à programmer, rédiger mes emails, répondre à mes questions sans queue ni tête, m'écouter etc... Merci infiniment!!!!!!!

A special thanks for all my friends I gained along the way outside work, friends from EM-ABG, from home and from everywhere, you all contributed a lot of happiness to sustain me during the past four years. I would also like to thank Mu-En for all the scientific discussions. I hope one day we can combine our expertise to start a lab together :) Mathieu, merci pour ton soutien dans certains moments difficiles surtout vers la fin de thèse. Tu étais là pour me réconforter, me supporter dans les moments où j'étais pénible, m'aider à trouver des solutions, enfin organiser un peu ma vie :D Merci de tout ce que tu m'apportes. Merci d'être à mes côtés. J'ai hâte de découverir le monde avec toi. Maintenant, tu peux t'amuser à corriger ce paragraphe :D

Finally, I would like to thank my family, my dog and cats for their endless love and support even I'm far away for long.

Curriculum vitae

About the author

Zih-Hua Fang was born on 3 March in Tainan, Taiwan. She obtained her B.Sc. in Animal Science and Technology from National Taiwan University in 2010. Her interest for scientific research led her to pursue the European Master in Animal Breeding and Genetics (EM-ABG) with full scholarship from 2011. During her master, Zih-Hua spent one year in Universität für Bodenkultur Wien and did a minor thesis about the effect of selection and population structure in genome-wide association studies. She later spent her second year of master in AgroParisTech and did her major thesis in INRA Jouy-en-Josas where she worked with both proteomic and genomic approaches to characterize new alleles encoding novel milk protein variants in cattle and sheep. In 2013, she started her PhD in the framework of the European Graduate School in Animal Breeding and Genetics between AgroParisTech-INRA and Wageningen University. In her PhD, Zih-Hua investigated the genetic background of α_{s1}- and α_{s2}-casein phosphorylation in cows' milk, and the results are presented in this thesis.

List of publications

Peer-reviewed publications

Fang, Z.H., Bovenhuis, H., van Valenberg, H.J.F., Martin, P., Huppertz, T. and Visker, M.H.P.W., Accepted. Genetic parameters for α_{s1}- and α_{s2}-casein phosphorylation isoforms in Dutch Holstein Friesian. *Journal of Dairy Science*.

Fang, Z.H., Bovenhuis, H., Delacroix-Buchet, A., Miranda, G., Boichard, D., Visker, M.H.P.W. and Martin, P., 2017. Genetic and nongenetic factors contributing to differences in α_s-casein phosphorylation isoforms and other major milk proteins. *Journal of Dairy Science*, 100(7), pp.5564-5577.

Fang, Z.H., Visker, M.H.P.W., Miranda, G., Delacroix-Buchet, A., Bovenhuis, H. and Martin, P., 2016. The relationships among bovine α_s-casein phosphorylation isoforms suggest different phosphorylation pathways. *Journal of Dairy Science*, 99(10), pp.8168-8177.

Manuscript in preparation

Fang, Z.H., Bovenhuis, H., van Valenberg, H.J.F., Martin, P., Duchemin S. I., Huppertz, T. and Visker, M.H.P.W., Manuscript in preparation. Genome-wide association study for α_{s1}- and α_{s2}-casein phosphorylation in Dutch Holstein Friesian.

Conference Proceedings

Fang, Z.H., Bovenhuis, H., Visker, M.H.P.W., Delacroix-Buchet, A., Miranda, G., and Martin, P., 2016. Genetic and nongenetic factors contribute to differences in relative proportion of α_s-casein phosphorylation isoforms among Montbéliarde cows. In: *Proceedings of 12th International Meeting on Mountain Cheese, Padova, Italy, 20-22 June 2017*.

Fang, Z.H., Miranda, G. and Martin, P., 2014. Characterization of a new allele encoding α_{s2}-casein in the Lacaune dairy sheep: possible phylogeny relationship among *CSN1S2* alleles. In: *Proceedings of 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada, 17-22 August 2014*.

Individual and Training Supervision Plan

The Basic Package (7.5 credits)	
Welcome to EGS-ABG , Addis Ababa (Ethiopia)	2013
EGS-ABG research school, Addis Ababa (Ethiopia)	2013
EGS-ABG research school, Uppsala (Sweden)	2016
Research integrity & ethics and animal science, Wageningen (The Netherlands)	2016

Dissemination of Knowledge (12 credits)	
International conferences	
World Congress on Genetics Applied to Livestock Production, Vancouver (Canada)	2014
International Symposium on Milk Genomics and Human Health, Sydney (Australia)	2015
ADSA-ASAS-CSAS Joint Annual Meeting, Salt lake city (USA)	2016
67th Annual Meeting of the European Federation of Animal Science, Belfast (UK)	2016
Seminars and workshops	
Annual meeting of the doctoral candidates of the Animal Genetics Division of INRA, Jouy-en-Josas (France)	2014
Annual meeting of the doctoral candidates of the Animal Genetics Division of INRA, La Rochelle (France)	2015
Presentations	
Poster; Annual meeting of the doctoral candidates of the Animal Genetics Division of INRA, Jouy-en-Josas (France)	2014
Poster; World Congress on Genetics Applied to Livestock Production, Vancouver (Canada)	2014
Oral; Annual meeting of the doctoral candidates of the Animal Genetics Division of INRA, La Rochelle (France)	2015
Poster; International Symposium on Milk Genomics and Human Health, Sydney (Australia)	2015
Oral; ADSA-ASAS-CSAS Joint Annual Meeting, Salt lake city (USA)	2016
Oral; 67th Annual Meeting of the European Federation of Animal Science, Belfast (UK)	2016

Advanced Scientific Courses (17 credits)	
Bioinformatics approaches to identify causative sequence variants in farm animals, Uppsala (Sweden)	2014
PH525x Data analysis for genomics, Paris (France)	2014
PH525.5x Case Study: RNA-seq data analysis, Paris (France)	2014
Statistical models for genomic prediction in animals and plants, Aarhus (Denmark)	2015
Course for ASReml package, Wageningen (The Netherlands)	2016
Statistics for the life sciences, Wageningen (The Netherlands)	2016
Statistical genetics, Faro (Portugal)	2016
Quantitative Genetics Discussion Group (QDG), Wageningen (The Netherlands)	2015-2017

Professional Skills Support Courses (5 credits)	
Getting it Across in English, Paris (France)	2015
Scientific writing, Wageningen (The Netherlands)	2016
Techniques for writing and presenting a scientific paper, Wageningen (The Netherlands)	2016
The final touch: how to write the general introduction and discussion,Wageningen (The Netherlands)	2017

Research Skills Training (4 credits)	
External training period, Wageningen (The Netherlands)	2015-2017

Complied with the educational requirements set by the European Graduate School in Animal Breeding and Genetics, the Graduate School of Agriculture, Food, Biology, Environment and Health of the Agricultural, Veterinary and Forest Institute of France and the Graduate School of Animal Science of Wageningen University & Research

Colophon

The research described in this thesis was financially supported by a joint grant from the European Commission (within the framework of the Erasmus-Mundus joint doctorate ”EGS-ABG”) and APIS-GENE.

The data used in Chapter 2 and 3 was part of the French “FROM’MIR” project, funded by the French Ministry of Agriculture, Agro-food and Forest, the French Dairy Interbranch Organization (CNIEL, Paris, France), the Regional Union of Protected Designation cheeses of Franche-Comté (URFAC, Poligny, France), and the Regional Council of Bourgogne Franche-Comté (Besan- çon, France), and the data used in Chapter 4 and 5 was part of the Dutch Milk Genomics Initiative, funded by Wageningen University (Wageningen, the Netherlands), the Dutch Dairy Association NZO (Zoetermeer, the Netherlands), Cooperative Cattle Improvement Organization CRV (Arnhem, the Netherlands), and the Dutch Technology Foundation STW (Utrecht, the Netherlands).

Cover design: Wei-Ping Li
Printed by: Digiforce

Title : The genetic background of bovine α_{s1}- and α_{s2}-casein phosphorylation

Keywords : milk protein composition, posttranslational modification, heritability, quantitative trait loci

Abstract :

Phosphorylation of caseins (CN) is a crucial post-translational modification allowing caseins to aggregate as micelles. The formation and stability of casein micelles are important for transporting abundant minerals to the neonate and manufacturing of dairy products. Therefore, it is of great interest to explore variation in degrees of phosphorylation of caseins and study to what extent genetic and other factors contribute to this variation. This thesis aimed to investigate the genetic background of bovine milk protein composition with a focus on phosphorylation of α_{s1}- and α_{s2}-CN. Thus, two studies were conducted to quantify phosphorylation levels of α_{s1}- and α_{s2}-CN: one in French Montbéliarde (FM) using liquid chromatography coupled with electrospray ionization mass spectrometry and the other in Dutch Holstein Friesian (DHF) using capillary zone electrophoresis. In FM, in addition to the known isoforms α_{s1}-CN-8P and-9P and α_{s2}-CN-10P to -13P, three new phosphorylation isoforms were detected, namely α_{s2}-CN-9P, α_{s2}-CN-14P, and α_{s2}-CN-15P. Relative concentrations of the phosphorylation isoforms varied considerably among cows. Phenotypic correlations showed that isoforms phosphorylated at higher degrees (α_{s1}-CN-9P and α_{s2}-CN-12P to -14P) correlated negatively with isoforms phosphorylated at lower degrees (α_{s1}-CN-8P, α_{s2}-CN-10P, and -11P). Furthermore, it was shown that α_{s1}- and α_{s2}-CN phosphorylation profiles changed across parity and lactation, and exploitable genetic variation for the phosphorylation degrees of α_{s1}- and α_{s2}-CN (defined as the proportion of higher-degree isoforms in α_{s1}- and α_{s2}-CN, respectively) exists in FM. In DHF, three α_{s2}-CN isoforms, namely α_{s2}-CN-10P to -12P, and the phosphorylation degrees of α_{s1}- and α_{s2}-CN were quantified. High intra-herd heritabilities were estimated for individual α_{s2}-CN phosphorylation isoforms and the phosphorylation degrees of α_{s1}- and α_{s2}-CN (ranging from 0.54 to 0.89). This suggests that genetic factors contribute substantially to observed differences in α_{s1}- and α_{s2}-CN phosphorylation profiles. The correlation between the phosphorylation degrees of α_{s1}- and α_{s2}-CN was 0.94. Additionally, a total of 10 regions, distributed across *Bos taurus* autosomes (BTA) 1, 2, 6, 9, 11, 14, 15, 18, 24 and 28, were detected to be associated with individual α_{s1}- and α_{s2}-CN phosphorylation isoforms and their phosphorylation degrees in DHF. Regions on BTA1, 6, 11 and 14 were associated with multiple traits studied. Two quantitative trait loci (QTL) regions were detected on BTA1: one affecting α_{s2}-CN production, and the other affecting α_{s1}-CN PD and α_{s2}-CN PD. The QTL region on BTA6 affected only individual α_{s2}-CN isoforms. The QTL region on BTA11 and 14 affected relative concentrations of α_{s2}-CN-10P and α_{s2}-CN-11P, α_{s1}-CN PD and α_{s2}-CN PD. Results suggested that effects of identified genomic regions on α_{s1}-CN PD and α_{s2}-CN PD are probably due to changes in milk synthesis and phosphorus secretion in milk. Finally, differences among studies due to factors such as analytical methods, trait definitions, and breed on genetic parameters and correlations are discussed using the two dataset from this thesis. It is concluded that differences in heritability estimates for α_{s1}-CN-8P and -9P, α_{s2}-CN-10P, -11P and -12P, and α_{s1}-CN PD between FM and DHF were mainly due to genetic differences between breeds. As for α_{s2}-CN PD, it was defined differently in FM and DHF due to analytical methods used. It is shown that both trait definitions successfully quantified the proportion of isoforms with higher degrees of phosphorylation because of similar estimated correlations using both definitions on the FM dataset. Additionally, it is hypothesized that a two-casein-kinase system is involved in the phosphorylation of α_{s1}- and α_{s2}-CN based on results in this thesis.

Titre : The genetic background of bovine α_{s1}- and α_{s2}-casein phosphorylation

Mots-clés : lactoprotéine, modification post-traductionnelle, héritabilite, locus à Caractère Quantitatif

Résumé :

La phosphorylation des caséines (CN) est une modification post-traductionnelle qui permet l'aggrégation des caséines sous formes de micelles. La stabilité de ces structures est essentielle pour transporter les minéraux au nouveau-né et fabriquer des produits laitiers. Il est par conséquent fondamental d'explorer la variation du degré de phosphorylation (PD) des CN, et d'étudier dans quelle mesure les facteurs génétiques contribuent à cette variation. Cette thèse visait à étudier le contexte génétique de la composition des protéines du lait de vache, en mettant l'accent sur la phosphorylation des CN α_{s1} et α_{s2}. Deux études ont été menées pour analyser leur niveau de phosphorylation : la première en race Montbéliarde française (FM), en utilisant la chromatographie liquide couplée à la spectrométrie de masse à ionisation par électro-nébulisation et l'autre en race Holstein Friesian néerlandaise (DHF), en utilisant l'électrophorèse capillaire de zone. En race FM, en plus des isoformes connues : α_{s1}-CN-8P et -9P, de α_{s2}-CN-10P à -13P, trois nouvelles isoformes ont été détectées (α_{s2}-CN-9P, α_{s2}-CN-14P et α_{s2}-CN-15P). Les concentrations relatives de ces isoformes variaient considérablement entre vaches. Les corrélations phénotypiques ont montré que les isoformes phosphorylées aux degrés les plus élevés (α_{s1}-CN-9P, de α_{s2}-CN-12P à -14P) étaient corrélées négativement avec les isoformes phosphorylées à des degrés inférieurs (α_{s1}-CN-8P, α_{s2}-CN-10P et -11P). En outre, les profils de phosphorylation des CN α_{s1} et α_{s2} changent avec la parité et le stade de lactation, et il est possible d'exploiter en FM la variation génétique du PD de ces caséines (définie comme la proportion d'isoformes ayant les taux de phosphorylation les plus élevés). En DHF, nous avons quantifié 3 isoformes d'α_{s2}-CN, à savoir α_{s2}-CN-10P, -11P, et -12P et les PD des CN α_{s1} et α_{s2}. Des héritabilités intra-troupeau élevées ont été estimées pour les isoformes de phosphorylation d'α_{s2}-CN et les PD des CN α_{s1} et α_{s2} (comprise entre 0,54 et 0,89). Ceci suggère que des facteurs génétiques contribuent substantiellement aux différences observées dans les profils de phosphorylation de ces CN. La correlation entre les degrés de phosphorylation de α_{s1}-CN et d'α_{s2}-CN était de 0,94. Nous avons montré que 10 régions, réparties sur les autosomes BTA1, 2, 6, 9, 11, 14, 15, 18, 24 et 28, sont associées aux isoformes de phosphorylation des CN α_{s1} et α_{s2} et leurs PD en DHF. Les régions sur BTA1, 6, 11 et 14 étaient associées à plusieurs caractères étudiés. Deux régions quantitative trait loci (QTL) ont été détectées sur BTA1 : l'une affectant la production d'α_{s2}-CN et l'autre le PD des deux CN. La région QTL localisée sur BTA6 affectait uniquement les isoformes d'α_{s2}-CN. Les régions QTL situées sur BTA11 et BTA14 influencent les concentrations relatives en α_{s2}-CN-10P et -11P, et les PD des CN α_{s1} et α_{s2}. Ces résultats suggèrent que les effets des régions génomiques identifiées impactant les PD sont probablement dus à des modifications de biosynthèse des constituants du lait et à la sécrétion de phosphore dans le lait. Enfin, les différences observées sur les paramètres génétiques estimés et les corrélations, sont discutées sur la base de l'ensemble des données issues des deux études décrites dans cette thèse. En conclusion, les différences observées sur les estimations d'héritabilité entre FM et DHF semblent principalement dues à des composantes génétiques. Quant au PD de l'α_{s2}-CN, il a été défini différemment en FM et en DHF, en raison des méthodes analytiques. Il a été possible de quantifier avec succès la proportion d'isoformes présentant des degrés de phosphorylation élevés, en raison de la similitude dans les corrélations estimées avec les deux définitions aux données de FM. Finalement, sur la base des résultats présentés dans cette thèse, nous avons émis l'hypothèse qu'un système comportant deux caséines-kinases puisse être impliqué dans la phosphorylation des CN α_{s1} et α_{s2}.