Phosphorylation of α_{S1} -casein is regulated by different genes

E. Bijl,* H. J. F. van Valenberg,* T. Huppertz,† A. C. M. van Hooijdonk,* and H. Bovenhuis‡¹

*Dairy Science and Technology Group, Wageningen University, PO Box 17, 6700 AA, Wageningen, the Netherlands

†NIZO Food Research, PO Box 20, 6710 BA, Ede, the Netherlands

‡Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH, Wageningen, the Netherlands

ABSTRACT

Casein phosphorylation is a posttranslational modification catalyzed by kinase enzymes that attach phosphate groups to specific AA in the protein sequence. This modification is one of the key factors responsible for the stabilization of calcium phosphate nanoclusters in casein micelles and for the internal structure of the casein micelles. α_{S1} -Casein (α_{S1} -CN) is of special interest because it constitutes up to 40% of the total casein fraction in milk, and it has 2 common phosphorylation states, with 8 (α_{S1} -CN-8P) and 9 (α_{S1} -CN-9P) phosphorylated serine residues. Factors affecting this variation in the degree of phosphorylation are not currently known. The objective of this research was to determine the genetic background of α_{S1} -CN-8P and α_{S1} -CN-9P. The genetic and phenotypic correlation between α_{S1} -CN-8P and α_{S1} -CN-9P was low (0.18 and 0.19, respectively). This low genetic correlation suggests a different genetic background. These differences were further investigated by means of a genome-wide association study, which showed that both α_{S1} -CN-8P and α_{S1} -CN-9P were affected by a region on *Bos taurus* autosome (BTA) 6, but only α_{S1} -CN-8P was affected by a region on BTA11 that contains the gene that encodes for β -lactoglobulin (β -LG), and only α_{S1} -CN-9P was affected by a region on BTA14 that contains the diacylglycerol acyltransferase 1 (*DGAT1*) gene. Estimated effects of β -LG protein genotypes showed that only α_{S1} -CN-8P was associated with the β -LG A/B polymorphism (g.1772G>A and g.3054C>T); the AA genotype of β -LG was associated with a lower concentration of α_{S1} -CN-8P (−0.32% wt/wt) than the BB genotype (+0.41% wt/wt). Estimated effects of *DGAT1* K232A genotypes showed that only α_{S1} -CN-9P was associated with the *DGAT1* gene polymorphism; *DGAT1* AA genotype was associated with a higher α_{S1} -CN-9P concentration (+0.53% wt/wt) than the *DGAT1* KK genotype (−0.44% wt/wt). The results give insight in phosphorylation of α_{S1} -CN-8P and α_{S1} -

CN-9P, which seem to be regulated by a different set of genes.

Key words: milk, casein, phosphorylation, protein variant

INTRODUCTION

Phosphorylation of caseins is catalyzed by kinase enzymes that attach phosphate groups to specific AA residues Ser and Thr (Mercier, 1981). This posttranslational modification occurs in the Golgi apparatus of the mammary epithelial cells after formation of the polypeptide chain (Bingham and Farrell, 1977) and is influenced by factors such as protein sequence, gene expression of kinase enzymes, substrate availability, and phosphorylation site accessibility (Holland, 2009).

In general, phosphorylation and dephosphorylation of a protein can regulate almost every aspect of protein function; for example, its biological activity, stabilization, and initiation in protein–protein interactions (Cohen, 2002). Particularly in caseins, phosphorylation is one of the key factors responsible for the stabilization of calcium phosphate nanoclusters in casein micelles and the internal structure of the casein micelles (De Kruif and Holt, 2003; Huppertz, 2013). This unique micellar structure allows milk to deliver large amounts of calcium and phosphate to the neonate without enhanced risk of biocalcification in the mammary gland (Neville, 2005; Holt and Carver, 2012).

Phosphorylation state varies greatly among the caseins. Although α_{S1} -CN, α_{S2} -CN, β -CN, and κ -CN are all phosphoproteins, the number of phosphorylated groups (n×P) they carry varies widely, from 1P to 3P on κ -CN, 4P to 5P on β -CN, 8P to 9P on α_{S1} -CN, to 10P to 13P on α_{S2} -CN (Farrell et al., 2004; Holland, 2009). The causes for variation in phosphorylated states of caseins and the effects on internal structure of casein micelles are not known and might differ by type of casein. Differences in phosphorylation of α_{S1} -CN are of special interest because total α_{S1} -CN constitutes up to 40% of the total casein fraction in milk, and it has 2 common phosphorylated states: α_{S1} -CN-8P and α_{S1} -CN-9P. The α_{S1} -CN-9P fraction contains an additional phosphorylated serine residue at position 56 of the pre-

Received February 17, 2014.

Accepted July 10, 2014.

¹Corresponding author: Henk.Bovenhuis@wur.nl

protein, including the signal peptide, and was previously known as α_{S0} -CN (Manson et al., 1977). Analysis of variation in casein phosphorylation on the milk samples of almost 2,000 individual Holstein-Friesian cows showed that the α_{S1} -CN-8P fraction occurred in a 3-fold excess over the α_{S1} -CN-9P fraction (Heck et al., 2008). The heritability of total α_{S1} -CN concentration determined for the same set of samples was 0.47 (Schopen et al., 2009). However, the genetic background for the individual phosphorylated fractions α_{S1} -CN-8P and α_{S1} -CN-9P is not known.

The objective of this research was to determine the genetic background of the α_{S1} -CN-8P and α_{S1} -CN-9P fractions in milk. The results for 1,912 cows taken during winter were used to estimate genetic parameters and correlations and to perform a genome-wide association study (GWAS) to identify genomic regions associated with phosphorylation of α_{S1} -CN-8P and α_{S1} -CN-9P.

MATERIALS AND METHODS

Phenotypes

Milk samples of almost 2,000 Holstein-Friesian cows in their first lactation from 398 commercial herds in the Netherlands were collected as part of the Dutch Milk Genomics Initiative (Schennink et al., 2007; Schopen et al., 2009; Bouwman et al., 2011). The pedigrees of the cows were supplied by the CRV (Cooperative Cattle Improvement Organization, Arnhem, the Netherlands). In the current study of 1,857 milk samples taken in winter, milk protein composition was determined by capillary zone electrophoresis (Heck et al., 2008). Using this method, it was possible to estimate relative protein concentration (%) on the basis of peak area of the protein fraction and total peak area in the electropherogram. The concentrations of α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN, α -LA, and β -LG were determined. In addition to total α_{S1} -CN, it was possible to determine the concentration of the individual fractions of α_{S1} -CN-8P and α_{S1} -CN-9P.

Genotypes

DNA was isolated from blood samples of 1,736 cows for genotyping. As described in detail by Schopen et al. (2011), a 50K (~50,000) SNP chip developed by CRV (Cooperative Cattle Improvement Organization) was used to genotype cows, as well as the sires of those cows, using 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). In total, 589 SNP were located on the X chromosome and 775 SNP could not be mapped to a

Bos taurus chromosome and were therefore assigned to BTA 0. The SNP on BTA 0 and the X chromosome were included in the study.

The data set used in the association study consisted of 1,667 animals with both α_{S1} -CN-8P and α_{S1} -CN-9P phenotypes and genotypes. Polymorphisms responsible for the known protein variants of β -CN (A^1 , A^2 , A^3 , B, and I) and κ -CN (A, B, and E) were genotyped using a SNaPshot assay, as described by Visker et al. (2011). Based on the protein variants, β - κ -CN haplotypes were derived (Visker et al., 2011). In total, 1,606 cows were available with β - κ -CN haplotypes and α_{S1} -CN-8P and α_{S1} -CN-9P phenotypes. Protein variants A and B for β -LG were genotyped as described by Ganai et al. (2009). In total, 1,671 cows were available for estimating the effect of β -LG protein variants on α_{S1} -CN-8P and α_{S1} -CN-9P. Genotypes for the diacylglycerol acyltransferase 1 (*DGAT1*) K232A polymorphism were obtained as described by Schennink et al. (2007). The effect of the *DGAT1* K232A polymorphism on α_{S1} -CN-8P and α_{S1} -CN-9P was estimated based on 1,704 cows.

Statistical Analyses

Genetic Parameters. To estimate genetic parameters and variance components, the following model was used:

$$y_{klmn} = \mu + \beta_1 \text{dim}_{klmn} + \beta_2 e^{-0.05 \text{dim}_{klmn}} + \beta_3 ca_{klmn} + \beta_4 ca_{klmn}^2 + \text{season}_k + \text{scode}_l + \text{animal}_m + \text{herd}_n + e_{klmn}, \quad [1]$$

where y_{klmn} is the observation for α_{S1} -CN-8P or α_{S1} -CN-9P, μ is the overall mean of the trait, dim_{klmn} is a covariate describing the effect of days in lactation, ca_{klmn} is a covariate describing the effect of age at first calving, season_k is the fixed effect of the k th class of calving season [3 classes: summer (June–August 2004), autumn (September–November 2004), and winter (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 of animal m , herd_n is the random herd effect of the n th herd, and e_{klmn} is the random residual. The animal effects were assumed to be distributed as $N(0, \mathbf{A}\sigma_a^2)$, herd effects were assumed to be distributed as $N(0, \mathbf{I}\sigma_{\text{herd}}^2)$, and the residuals were assumed to be distributed as $N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{A}$ is the additive genetic relationships matrix, $\mathbf{I}$ is the identity matrix, σ_a^2 is the additive genetic variance, σ_{herd}^2 is the herd variance, and σ_e^2 is the residual variance.

The statistical package ASReml (Gilmour et al., 2009) was used to perform the analysis. The intraherd heritability was calculated as

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

The proportion of variance due to differences among herds was calculated as

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

For estimating genetic and phenotypic correlations between α_{SI} -CN-8P and α_{SI} -CN-9P bivariate analyses were performed using model [1].

GWAS and Single Genes. The single SNP association study was performed using model [1], extended with a fixed SNP genotype effect. In the association analysis, the variance components were fixed to estimates obtained from a model without a SNP. Similarly, to estimate effects of β -LG protein variants or the *DGAT1* K232A polymorphism, these effects were added to model [1]. Significance of genotype effects was based on the conditional Wald F Statistics (Gilmour et al., 2009).

Significance Thresholds. The genome-wide false discovery rate (FDR) was based on the *P*-values from the animal model using the qvalue package (Storey and Tibshirani, 2003) in R (R Development Core Team, 2005). The FDR was calculated for each trait individually. Associations with an FDR <0.1% were considered significant.

Haplotype Analysis. To estimate the variance explained by the β - κ -CN haplotypes, the following model was used:

$$\begin{aligned} y_{klmnop} = & \mu + \beta_1 \text{dim}_{klmnop} + \beta_2 e^{-0.05 \text{dim}_{klmnop}} + \beta_3 ca_{klmnop} \\ & + \beta_4 ca_{klmnop}^2 + \text{season}_k + \text{scode}_l + \text{haplo1}_o + \text{haplo2}_p \\ & + \text{animal}_m + \text{herd}_n + e_{klmnop}, \end{aligned} \quad [2]$$

where the variables were the same as in model [1], but extended with *haplo1_o* and *haplo2_p*, where *haplo1_o* is the effect of the first β - κ -CN haplotype and *haplo2_p* is the effect of the second β - κ -CN haplotype. The 2 haplotypes of an individual were randomly assigned to *haplo1* or *haplo2*, and the design matrices of both haplotype effects were combined to estimate one haplotype effect. Different analyses were performed with *haplo1* and *haplo2* modeled either as fixed or as random variables.

Modeling haplotypes as random effects allows estimation of the variance explained by casein haplotypes and, in that case, effects for *haplo1* and *haplo2* were assumed to be distributed as $N(0, \mathbf{I}\sigma_{\text{haplo}}^2)$, where $\mathbf{I}$ is the identity matrix and σ_{haplo}^2 is the casein haplotype variation. The proportion of variation explained by casein haplotypes was calculated as

$$h_{\text{haplo}} = \frac{\sigma_{\text{haplo}}^2}{(\sigma_{\text{haplo}}^2 + \sigma_a^2 + \sigma_e^2)}.$$

RESULTS AND DISCUSSION

Genetic Parameters

The means, standard deviation (SD), coefficient of variation (CV), and genetic parameters for α_{SI} -CN fractions are given in Table 1. α_{SI} -Casein-8P accounted for 21.3% of total protein in milk and α_{SI} -CN-9P 7.4%. The CV for α_{SI} -CN-9P concentration was greater (14%) than that for α_{SI} -CN-8P (5%). The proportion of variation due to differences between herds was similar for concentrations of α_{SI} -CN-8P and α_{SI} -CN-9P and was relatively low (~ 0.10). In contrast, the intraherd heritability for α_{SI} -CN-9P concentration was high (0.76). The heritability for α_{SI} -CN-8P concentration was lower (0.48). The correlations between total α_{SI} -CN concentration and both underlying traits (α_{SI} -CN-8P and α_{SI} -CN-9P) were high. However, the correlations among α_{SI} -CN-8P and α_{SI} -CN-9P concentrations, which account for a large part of total α_{SI} -CN, were surprisingly low: the phenotypic correlation was 0.19 ± 0.04 and the genetic correlation was 0.18 ± 0.15 , which is not significantly different from zero. The low genetic correlations indicate that α_{SI} -CN-8P and α_{SI} -CN-9P concentration are, to a large extent, genetically different traits.

GWAS

The genetic differences between α_{SI} -CN-8P and α_{SI} -CN-9P concentration were further explored by means of GWAS, the results of which are displayed in Figure 1. Both components were affected by a region on BTA 6, but only α_{SI} -CN-8P concentration was affected by a region on BTA 11 and only α_{SI} -CN-9P concentration was affected by a region on BTA 14. This is the first time that a different genetic origin has been shown for α_{SI} -CN-8P and α_{SI} -CN-9P fractions in milk. The effects on BTA 6, 11, and 14 will be further discussed in the following sections.

BTA 6. The SNP with the most significant association on BTA 6 for the concentration of α_{SI} -CN-8P

Table 1. Means (% wt/wt of total protein in milk), standard deviations (SD), coefficients of variation (CV), phenotypic variance (σ_p^2), intraherd heritability (h^2), proportion of variance explained by herd (h_{herd}), and correlations (genetic below diagonal and phenotypic above diagonal) between the concentrations of α_{S1} -CN, α_{S1} -CN-8P, and α_{S1} -CN-9P of 1,857 Holstein-Friesian cows (SE in parentheses)

Trait ¹	Mean (%)	SD (%)	CV (%)	σ_p^2	h^2	h_{herd}	Correlation		
							α_{S1} -CN	α_{S1} -CN-8P	α_{S1} -CN-9P
α_{S1} -CN	33.6	1.7	5	2.80	0.52 (0.11)	0.11 (0.02)	—	0.75 (0.02)	0.73 (0.02)
α_{S1} -CN-8P	21.3	1.1	5	1.32	0.48 (0.10)	0.12 (0.02)	0.73 (0.08)	—	0.19 (0.04)
α_{S1} -CN-9P	7.4	1.1	14	1.18	0.76 (0.12)	0.08 (0.02)	0.79 (0.06)	0.18 (0.15)	—

¹ α_{S1} -CN-8P and α_{S1} -CN-9P = 2 common phosphorylated states of α_{S1} -CN, with 8 and 9 phosphorylated serine residues, respectively.

was ULGR_BTC-043582 [$-\log_{10}(P) = 34.5$] at position 88.1 Mb. This SNP was also significant for α_{S1} -CN-9P concentration [$-\log_{10}(P) = 7.8$] and was earlier demonstrated to be the lead SNP on BTA 6 for total α_{S1} -CN concentration by Schopen et al. (2011). The SNP ULGR_BTC-043582 is located at 88,091,127 bp (Btau 4.6.1), in an intron of SULT1D1. This is close to the α_{S1} -CN gene, which is located at 88,241,049 to 88,258,584 bp (Btau 4.6.1). The SNP ULGR_BTC-043582 is in strong linkage disequilibrium with the β -CN protein variant I ($r^2 = 0.74$) and κ -CN B ($r^2 = 0.53$). The SNP with the most significant association on BTA 6 for α_{S1} -CN-9P concentration was ULGR_BTC-

053514 [$-\log_{10}(P) = 9.0$] at position 83,503,006 bp (Btau 4.6.1). This SNP was also significant for α_{S1} -CN-8P concentration [$-\log_{10}(P) = 22.1$]. Interestingly, this SNP was earlier demonstrated to be the lead SNP for total α_{S2} -CN concentration by Schopen et al. (2011). Single-nucleotide polymorphisms ULGR_BTC-053514 and ULGR_BTC-043582 are in linkage disequilibrium ($r^2 = 0.56$).

The estimated effects of β -LG and *DGAT1* genotypes and β - κ -CN haplotypes on α_{S1} -CN-8P and α_{S1} -CN-9P concentration are shown in Table 2. The IB β - κ -CN haplotype was associated with a lower concentration of α_{S1} -CN-8P than other β - κ -CN haplotypes and the A¹B

Figure 1. The $-\log_{10}(P)$ -values for all 48,874 SNP for all 29 BTA as well as BTA 0 (which contained the unmapped SNP) and the X chromosome for α_{S1} -CN-8P and α_{S1} -CN-9P. The false discovery rate equal to 0.1% is indicated by the dotted line.

Table 2. Effect of β -LG ($n = 1,671$ animals) and diacylglycerol acyltransferase 1 (*DGAT1*; $n = 1,704$ animals) genotypes and β - κ -CN haplotypes ($n = 1,606$ animals; Visker et al., 2011) on the concentration of α_{S1} -CN-8P and α_{S1} -CN-9P (in % wt/wt of total protein in milk)^{1,2}

Genotype/ haplotype	No.	α_{S1} -CN-8P	α_{S1} -CN-9P
<i>β-LG</i>			
AA	539	-0.32 (0.06)	0.06 (0.05)
AB	870	0.00	0.00
BB	262	0.41 (0.07)	-0.12 (0.07)
$-\log_{10}(P)$		17.2	1.2
<i>DGAT1</i>			
AA	797	-0.06 (0.06)	0.53 (0.05)
AK	628	0.00	0.00
KK	279	0.01 (0.07)	-0.44 (0.06)
$-\log_{10}(P)$		0.3	43.9
<i>β-κ-CN³</i>			
A2A	1,503	0.16 (0.06)	0.03 (0.06)
IB	623	-0.56 (0.07)	-0.30 (0.07)
A1A	428	0.00	0.00
A1E	297	0.06 (0.08)	0.05 (0.08)
A1B	178	0.48 (0.10)	0.02 (0.10)
A2B	118	-0.08 (0.10)	-0.06 (0.11)
BB	65	-0.51 (0.13)	0.15 (0.13)
$-\log_{10}(P)$		48.3	6.5

¹Standard errors are in parentheses.

² α_{S1} -CN-8P and α_{S1} -CN-9P = 2 common phosphorylated states of α_{S1} -CN, with 8 and 9 phosphorylated serine residues, respectively.

³The estimated effect is the effect of having 1 copy of that haplotype and is based on a model with haplotypes modeled as fixed effect.

haplotype with a higher concentration of α_{S1} -CN-8P. The fraction of the total genetic variance that could be explained by the casein haplotypes on α_{S1} -CN-8P was 0.32 ± 0.15 . In addition, a small effect was found of the IB haplotype on α_{S1} -CN-9P concentration. The fraction of the total genetic variance that could be explained by the casein haplotypes on α_{S1} -CN-9P was 0.02 ± 0.02 .

The fact that α_{S1} -CN-8P and α_{S1} -CN-9P were associated with BTA 6 can be well understood, because the α_{S1} -CN gene, as well as the other casein genes, is located in this region (Threadgill and Womack, 1990).

BTA 11. We detected a highly significant effect [$-\log_{10}(P) > 5.41$] of a chromosomal region on BTA 11 on α_{S1} -CN-8P concentration. This region contains the *PAEP* gene, which encodes for the β -LG protein. The SNP that showed the strongest association with α_{S1} -CN-8P concentration was rs41255679. This SNP is located at 107.2 Mb, in the promoter region of the *PAEP* gene, and is known to be in linkage disequilibrium with β -LG protein variants A and B (Ganai et al., 2009). The estimated effects of β -LG genotypes on α_{S1} -CN-8P and α_{S1} -CN-9P concentration (Table 2) showed that the AA genotype of β -LG was associated with a lower concentration of α_{S1} -CN-8P (-0.32% wt/wt) than the BB genotype ($+0.41\%$ wt/wt). For α_{S1} -CN-9P, we observed no significant effect of β -LG genotypes. Based on a GWAS and using mainly the same data as in the current study, Schopen et al. (2011) reported that the

SNP rs41255679 showed the strongest association with β -LG concentration. It is known that β -LG protein variants A and B are associated with β -LG concentration (Heck et al., 2009); therefore, the observed effect on α_{S1} -CN-8P concentration might be due either to physical differences between protein variants A and B, or to differences in β -LG concentration. Additionally, interactions between β -LG protein variants and β -LG concentration might exist. Analyses showed no significant interaction ($P = 0.89$) between the effects of β -LG protein variants and β -LG concentration on α_{S1} -CN-8P concentration. However, we found highly significant effects of the β -LG protein variants [$-\log_{10}(P) = 12$] and β -LG concentration [$-\log_{10}(P) = 6$] on α_{S1} -CN-8P concentration. After adjusting for differences between β -LG protein variants, a higher β -LG concentration was associated with a higher α_{S1} -CN-8P concentration. However, when β -LG protein variants were excluded from the model, the relationship between β -LG concentration and α_{S1} -CN-8P concentration was negative. This can be explained by the higher β -LG concentration of β -LG AA milk compared with β -LG BB milk (Heck et al., 2009). β -Lactoglobulin A was associated with a lower α_{S1} -CN-8P concentration and β -LG B with a higher α_{S1} -CN-8P concentration, irrespective of accounting for β -LG concentration in the statistical analysis. The current study demonstrates a distinct effect of β -LG protein variants and β -LG concentration on α_{S1} -CN-8P concentration, and indicates that part of the negative effect of β -LG A on α_{S1} -CN-8P concentration was compensated for by the higher β -LG concentration. No other genes that could explain the association with α_{S1} -CN-8P were evident in this region.

The function of β -LG in milk has not yet been elucidated. Possible functions that have been suggested do not include protein phosphorylation processes and first need to be further studied. Therefore, the mechanism behind the established association between β -LG protein variants, as well as β -LG concentration and α_{S1} -CN-8P concentration, remains unclear.

BTA 14. We detected a highly significant effect (FDR $< 0.1\%$) of a chromosomal region on BTA 14 on α_{S1} -CN-9P concentration; this region contains the *DGAT1* gene. The SNP that showed the strongest association with α_{S1} -CN-9P concentration was ULGR_SNP_AJ318490_1c [rs109234250, $-\log_{10}(P) = 42.3$] at position 0.4 Mb followed by ULGR_SNP_AJ318490_1b [rs109326954, $-\log_{10}(P) = 41.4$], also at position 0.4 Mb. These 2 SNP are in full linkage disequilibrium with the *DGAT1* K232A polymorphism, in which Lys (K variant) at AA position 232 is replaced by Ala (A variant; Grisart et al., 2002; Winter et al., 2002). Besides the *DGAT1* gene, no other evident genes were located in this region.

Three unmapped SNP (which were assigned to BTA 0) showed highly significant effects on α_{S1} -CN-9P: ULGR_BTC-049604 [rs42213788, $-\log_{10}(P) = 10.7$], ULGR_rs29024688 [rs29024688, $-\log_{10}(P) = 16.4$] and ULGN_SNP_AJ318490_2 [ss1067343547, $-\log_{10}(P) = 41.1$]. Searching the SNP sequence of ULGR_BTC-049604 and ULGR_rs29024688 by BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) showed that, based on the UMD 3.1 map, these SNP are located on BTA14, close to the *DGAT1* region. Blasting the ULGN_SNP_AJ318490_2 sequence against the *DGAT1* sequence showed that this SNP is located in the *DGAT1* gene. Further, ULGN_SNP_AJ318490_2 is in high linkage disequilibrium ($r^2 = 0.99$) with the *DGAT1* K232A SNP (Bouwman et al., 2011).

The estimated effects of *DGAT1* genotypes on α_{S1} -CN-8P and α_{S1} -CN-9P concentration (Table 2) showed that the *DGAT1* AA genotype was associated with a higher concentration of α_{S1} -CN-9P (+0.53% wt/wt) than the KK genotype (−0.44% wt/wt). We detected no significant effect of *DGAT1* genotypes on α_{S1} -CN-8P.

The *DGAT1* K232A polymorphism influences milk fat percentage and composition: the K variant is associated with a higher fat content and more saturated fat compared with the A variant (Grisart et al., 2002; Winter et al., 2002; Schennink et al., 2007). Although it could be expected that the *DGAT1* polymorphism would mainly influence milk fat characteristics, ULGR_SNP_AJ318490_1b was earlier demonstrated to be the lead SNP on BTA14 for total α_{S1} -CN and α_{S2} -CN concentration and protein percentage in milk by Schopen et al. (2011). Furthermore, Lu (2013) concluded that the membrane composition of the epithelial cells of the mammary gland could be different between cows with *DGAT1* AA and KK genotypes. It seems, therefore, that the effect of *DGAT1* polymorphism is not limited to milk fat characteristics.

Substrate Specificity of Casein Kinases

The GWAS and estimated effects of β -LG and *DGAT1* genotypes showed that phosphorylation of α_{S1} -CN-8P and α_{S1} -CN-9P is not regulated by the same set of genes. Substrate specificity of casein kinases on α_{S1} -CN phosphorylation sites can provide understanding for the established differences. From the AA sequence of α_{S1} -CN-8P, we know that the 8 Ser residues that can be phosphorylated follow the S-x-E/pS motif, whereas the Ser residue at position 56, which is additionally phosphorylated in α_{S1} -CN-9P, follows the S-x-D motif. The S-x-D motif is a unique recognition site that occurs only at this position in α_{S1} -CN and not in any of the other caseins. Previously, it was concluded that mammary gland protein kinase was less active for this recogni-

tion site than for the common S-x-E/pS motif (Manson et al., 1977). However, recently, more insight has been gained from the consensus phosphorylation sites of kinase enzymes (Ubersax and Ferrell, 2007). With regard to casein kinases, Golgi-enriched fraction casein kinase (**GEF-CK**) from lactating mammary gland is thought to be responsible for phosphorylation of caseins (Moore et al., 1985). This fraction has the consensus sequence of S-x-E/pS and fails to recognize Asp, and therefore the S-x-D motif, for phosphorylation (Lasa-Benito et al., 1996). Also, Tagliabracci et al. (2012) showed that one of the candidates for the GEF-CK fraction, which has been identified as FAM20C, specifically phosphorylates S-x-E motifs. It seems, therefore, that another casein kinase in the Golgi apparatus of the lactating mammary gland could be responsible for phosphorylation of the S-x-D motif in α_{S1} -CN-9P. This enzyme has not been identified to date.

CONCLUSIONS

Considering the genetic correlations, GWAS, and effects of β -LG and *DGAT1* genotypes, we conclude that α_{S1} -CN-8P and α_{S1} -CN-9P are, to a large extent, genetically different traits and are not regulated by the same set of genes. Both traits were associated with SNP on BTA 6, but only α_{S1} -CN-8P was associated with β -LG protein variants and β -LG concentration whereas only α_{S1} -CN-9P was associated with the *DGAT1* K232 polymorphism. Substrate specificity of casein kinases on α_{S1} -CN phosphorylation sites might be an important influencing factor; the unique S-x-D motif of the α_{S1} -CN-9P fraction might be phosphorylated by a different enzyme than GEF-CK, which has been recognized to phosphorylate the 8 S-x-E/pS motifs of α_{S1} -CN-8P. The results indicate a specific purpose toward phosphorylation of both α_{S1} -CN-8P and α_{S1} -CN-9P fractions.

ACKNOWLEDGMENTS

This study is part of the Milk Genomics Initiative funded by Wageningen University (Wageningen, the Netherlands), the Dutch Dairy Association (NZO; Zoetermeer, the Netherlands), cattle cooperative CRV (Arnhem, the Netherlands), the Dutch Technology Foundation STW. The authors thank Alex Hulzebosch from Animal Breeding and Genomics Centre of Wageningen University for his technical assistance with running the statistical analyses by ASReml.

REFERENCES

- Bingham, E. W., and H. M. Farrell Jr. 1977. Phosphorylation of casein by the lactating mammary gland: A review. *J. Dairy Sci.* 60:1199–1207.

- Bouwman, A. C., H. Bovenhuis, M. Visker, and J. A. M. van Arendonk. 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. *BMC Genet.* 12:43–54.
- Cohen, P. 2002. The origins of protein phosphorylation. *Nat. Cell Biol.* 4:E127–E130.
- De Kruif, C. G., and C. Holt. 2003. Casein micelle structure, function and interactions. Pages 233–276 in *Advanced Dairy Chemistry. Vol. 1: Proteins*. P. F. Fox and P. L. H. McSweeney, ed. Kluwer Academic/Plenum, New York, NY.
- Farrell, H. M., Jr., R. Jimenez-Flores, 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.* 87:1641–1674.
- Ganai, N. A., H. Bovenhuis, J. A. M. van Arendonk, and M. Visker. 2009. Novel polymorphisms in the bovine beta-lactoglobulin gene and their effects on β -lactoglobulin protein concentration in milk. *Anim. Genet.* 40:127–133.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0. VSN International Ltd., Hemel Hempstead, UK.
- 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.* 12:222–231.
- Heck, J. M. L., C. Olieman, A. Schennink, H. J. F. van Valenberg, M. H. P. W. Visker, R. C. R. Meuldijk, and A. C. M. van Hooijdonk. 2008. Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. *Int. Dairy J.* 18:548–555.
- Heck, J. M. L., A. Schennink, H. J. F. van Valenberg, H. Bovenhuis, M. 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.* 92:1192–1202.
- Holland, J. W. 2009. Post-translational modifications of caseins. Pages 107–132 in *Milk Proteins: From Expression to Food*. A. Thompson, M. Boland, and H. Singh, ed. Elsevier, Amsterdam, the Netherlands.
- Holt, C., and J. A. Carver. 2012. Darwinian transformation of a 'scarcely nutritious fluid' into milk. *J. Evol. Biol.* 25:1253–1263.
- Huppertz, T. 2013. Chemistry of the caseins. Pages 135–160 in *Advanced Dairy Chemistry. Vol. 1A: Proteins: Basic Aspects*. P. L. H. McSweeney and P. F. Fox, ed. Springer, New York, NY.
- 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.* 382:149–152.
- Liu, Y., X. Qin, X. Z. H. Song, H. Y. Jiang, Y. F. Shen, K. J. Durbin, S. Lien, M. P. Kent, M. Sodeland, Y. R. Ren, L. Zhang, E. Sodergren, P. Havlak, K. C. Worley, G. M. Weinstock, and R. A. Gibbs. 2009. *Bos taurus* genome assembly. *BMC Genomics* 10:180–190.
- Lu, J. 2013. The biology of milk synthesis from a proteomics perspective. PhD Thesis. Wageningen University, Wageningen, the Netherlands.
- Manson, W., T. Carolan, and W. D. Annan. 1977. Bovine alpha-s-0 casein-phosphorylated homolog of alpha-s-1 casein. *Eur. J. Biochem.* 78:411–417.
- Mercier, J. C. 1981. Phosphorylation of caseins, present evidence for an amino-acid triplet code post-translationally recognized by specific kinases. *Biochimie* 63:1–17.
- 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.* 152:729–737.
- Neville, M. C. 2005. Calcium secretion into milk. *J. Mammary Gland Biol. Neoplasia* 10:119–128.
- R Development Core Team. 2005. R: A language and environment for statistical computing, reference index version 2.2.1. R Foundation for Statistical Computing, Vienna, Austria. <http://www.Rproject.org>.
- Schennink, A., W. M. Stoop, M. 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.* 38:467–473.
- Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. 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.* 92:1182–1191.
- Schopen, G. C. B., M. 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.* 94:3148–3158.
- Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genome-wide studies. *Proc. Natl. Acad. Sci. USA* 100:9440–9445.
- Tagliabracci, V. S., J. L. Engel, J. Z. Wen, S. E. Wiley, C. A. Worby, L. N. Kinch, J. Y. Xiao, N. V. Grishin, and J. E. Dixon. 2012. Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* 336:1150–1153.
- Threadgill, D. W., and J. E. Womack. 1990. Genomic analysis of the major bovine-milk protein genes. *Nucleic Acids Res.* 18:6935–6942.
- Ubersax, J. A., and J. E. Ferrell Jr. 2007. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* 8:530–541.
- Visker, M. H., B. W. Dibbitts, S. M. Kinders, H. J. F. van Valenberg, J. A. M. van Arendonk, and H. Bovenhuis. 2011. Association of bovine beta-casein protein variant I with milk production and milk protein composition. *Anim. Genet.* 42:212–218.
- Winter, A., W. Kramer, 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. USA* 99:9300–9305.