Short communication: A new bovine milk-protein variant: α -Lactalbumin variant D

M. H. P. W. Visker,^{*1} J. M. L. Heck,[†] H. J. F. van Valenberg,[‡] J. A. M. van Arendonk,^{*} and H. Bovenhuis^{*}

^{*}Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

[†]FrieslandCampina Research, PO Box 87, 7400 AB Deventer, the Netherlands

[‡]Dairy Science and Technology Group, Wageningen University, PO Box 8129, 6700 EV Wageningen, the Netherlands

ABSTRACT

Capillary zone electrophoresis of 1,948 Holstein-Friesian cows suggested the presence of an unknown protein variant of α -lactalbumin (α -LA) in the milk of 1 cow. Sequencing genomic DNA of this cow showed a polymorphism in the α -LA gene (*LAA*) that appears to be responsible for this protein variant. This single nucleotide polymorphism g.600G > T was located in exon 2 of *LAA* and causes the amino acid change 65Gln > His in the α -LA protein. This α -LA protein variant is a new protein variant and should be called α -LA protein variant D. This amino acid change is not expected to affect protein function. Genomic DNA of 156 bulls of various dairy cattle breeds was screened to examine the presence of the new α -LA protein variant D. Single nucleotide polymorphism g.600G > T, responsible for α -LA protein variant D, was not found in any of the 156 bulls. However, 10 other polymorphisms in the coding and promoter regions of *LAA* were detected that were used to construct haplotypes.

Key words: α -lactalbumin protein variant D, dairy cow, single nucleotide polymorphism, whey

Short Communication

Variants of bovine milk proteins are of interest because of their associations with milk production traits and milk protein composition. Many variants have been described for most of the major milk proteins (for a review, see Caroli et al., 2009), albeit associations have been investigated for only a limited number of variants. Best known are associations of β -LG protein variants A and B, of β -CN variants A¹, A², B, and I, and of κ -CN variants A, B, and E (e.g., Ikonen et al., 1999; Heck et al., 2009; Bonfatti et al., 2010).

For α -LA, only 3 protein variants have been described; variant B is regarded as the reference protein

and is predominant in *Bos taurus* cattle, both variants A and B occur in *Bos indicus* cattle, whereas variant C is found only in *Bos javanicus* cattle (Farrell et al., 2004; Caroli et al., 2009). Mature α -LA and its signal peptide contain 123 and 19 AA, respectively. The protein is encoded by the α -LA gene (*LAA*), which is composed of 4 exons and 3 introns (Vilotte et al., 1987) and has been assigned to bovine chromosome 5 (Threadgill and Womack, 1990). Protein variant A is characterized by AA change 10Arg > Gln, resulting from SNP g.113G > A in exon 1 of *LAA*. Variant C is characterized by either AA change Asp > Asn or Glu > Gln at an unknown position of the α -LA protein (Farrell et al., 2004; Caroli et al., 2009). The relatively small number of protein variants of α -LA could be due to natural selection against such variants because of the essential physiological role of α -LA in regulating the synthesis of lactose (Brew et al., 1968; Fitzgerald et al., 1970), which is the major osmotic constituent in milk. The aim of this study was to identify an unknown protein variant of α -LA and examine its presence in various dairy cattle breeds by means of screening genomic DNA of 1 cow and 156 bulls of different cattle breeds.

The milk protein composition of all 1,948 Holstein-Friesian cows of the Dutch Milk Genomics Initiative has been determined using capillary zone electrophoresis (CZE; Heck et al., 2008). All cows were homozygous for α -LA protein variant B (Heck et al., 2009), with the exception of 1 animal. The milk protein of this 1 cow showed an additional peak in the electropherogram, which appeared to be an additional α -LA protein variant (Figure 1). To assess the AA composition of this additional α -LA protein variant, genomic DNA obtained from the blood of this cow was screened for polymorphisms in *LAA* by sequencing. All 4 exons of *LAA* plus their flanking parts of intron sequences were amplified by PCR and sequenced on both strands, as described by Ganai et al. (2009). Primers for PCR and sequencing were designed on GenBank *LAA* genomic DNA sequence X06366 (corresponding to α -LA protein variant B; Vilotte et al., 1987) and are given in Supplementary Table 1.

Received August 2, 2011.

Accepted December 1, 2011.

¹Corresponding author: Marleen.Visker@wur.nl

Figure 1. Electropherogram of raw bovine milk from the 1 cow with an additional α -LA protein peak; β -CN A1 and β -CN A2 are β -CN protein variants A1 and A2, respectively. The cow is heterozygous for α -LA protein variant B (α -LA B) and the new α -LA protein variant D (α -LA D).

One polymorphism in *LAA* was detected that appears to be responsible for the additional α -LA protein variant. This SNP g.600G > T (ss469415640) was located in exon 2 of *LAA* and causes the AA change 65Gln > His in the α -LA protein. Histidine at position 65 of α -LA is not in accordance with the known protein variants A, B, and C of α -LA (Farrell et al., 2004). Apparently, this α -LA protein variant is a new protein variant that has not been described before. Based on the nomenclature system of milk proteins (Farrell et al., 2004) this new variant should be called α -LA protein variant D (GenBank: JN258330).

Glutamine at position 65 of α -LA seems to be conserved in different species (Pettersson et al., 2006). However, based on algorithms using sequence alignment and physiochemical properties the 65Gln > His AA change is not expected to affect protein function (SIFT: Kumar et al., 2009; ProPhyLER: Binkley et al., 2010). Illustrative is that histidine at position 65 of α -LA is suggested to be common in a group of whales (Rychel et al., 2004). Moreover, the AA change 65Gln > His is not in any of the functional sites of α -LA, as those involved in binding of calcium (Stuart et al., 1986; Anderson et al., 1997) which is important for folding and stability of α -LA, or involved in binding with β -1,4-galactosyltransferase to form the lactose synthase complex, or in binding of glucose to the lactose synthase complex (Grobler et al., 1994; Malinovskii et al., 1996; Pike et al., 1996). Consequently, effects of

this AA change on the properties of the α -LA protein are not likely. Associations of the new α -LA protein variant could not be studied, because only 1 cow had this α -LA variant D. Sequencing showed that this cow was heterozygous for SNP g.600G > T, which is in agreement with the presence of 2 peaks for α -LA in the electropherogram: 1 for α -LA protein variant B and 1 for the new protein variant D. Obvious phenotypes such as concentrations of α -LA (2.64% wt/wt of total protein) and lactose (4.75%) of this cow's milk did not deviate from the population averages (2.44 and 4.64%, respectively).

To investigate the presence of the new α -LA protein variant D in various dairy cattle breeds, *LAA* was screened for polymorphisms by sequencing genomic DNA obtained from semen of 11 Ayrshire, 20 Brown Swiss, 27 Simmental (Fleckvieh), 9 Guernsey, 72 Holstein-Friesian and 17 Jersey bulls, as described above. To check for linkage disequilibrium (LD) between the SNP responsible for α -LA protein variant D and SNP at possibly regulating positions, 1.6 kb of promoter region upstream of the transcription start site was screened additionally. Primers for PCR and sequencing of the promoter region were designed on GenBank *LAA* genomic DNA sequence M90645 (Bleck and Bremel, 1993) and are given in Supplementary Table 1.

Single nucleotide polymorphism g.600G > T, responsible for α -LA protein variant D, was not found in any of the 156 bulls. However, 10 further polymorphisms

were detected (Table 1): 4 in the promoter, 1 in the 5' untranslated region of exon 1, 3 in different introns and 2 in the 3' untranslated region of *LAA*. No further polymorphisms causing AA changes relative to α -LA reference protein variant B were found. Only SNP g.15G > A has been reported previously (Bleck and Bremel, 1993; Voelker et al., 1997) and associations of SNP g.15G > A with concentrations of α -LA and lactose have been described (Lundén and Lindersson, 1998; Schopen et al., 2011). Homozygous differences were found compared with the GenBank *LAA* genomic DNA sequences M90645 and X06366 on 9 positions in all 157 animals that were investigated (Table 1, footnote 1). This suggests that the GenBank sequences M90645 and X06366 do not represent the consensus sequence for *LAA* on these 9 positions.

All 10 polymorphisms were observed in the Holstein-Friesian bulls, with almost equal frequencies for proven and young bulls (results not shown). In each of the other breeds, 2 or more of the polymorphisms were not observed (Table 1). This could be due to rather low numbers of animals per breed combined with low frequencies of variant alleles. It may also be that these polymorphisms are indeed not present in these breeds. For example, SNP g.15G > A and g.1778T > C, which were in complete LD, had relatively high frequencies of variant alleles in Holstein-Friesian bulls, but were not observed in Brown Swiss, Simmental, Guernsey, and Jersey bulls. In contrast, SNP g.-1206C > A and g.-1000T > C were observed in all 6 breeds. Breed specificity of polymorphisms in *LAA* is also suggested by previous studies in which the variant allele of SNP g.15G > A was found only in Holstein (Bleck and Bremel, 1993; Martins et al., 2008) and in Swedish Red and White animals (Lundén and Lindersson, 1998).

The cow that was heterozygous for SNP g.600G > T, corresponding to being heterozygous for α -LA protein variants B and D, was either homozygous for the reference alleles (M90645, X06366; g.-1290GG, g.-1000TT, g.-788GG, g.1325GG, g.1675TT, and g.1857GG) or heterozygous (g.-1206CA, g.15GA, g.433TC, and g.1778TC) for each of the 10 SNP. Linkage disequilibrium of SNP g.600G > T could not be established because this SNP was not found in any of the other animals that were screened.

Haplotypes of all 156 sires comprising the 10 polymorphisms in *LAA* were inferred with PHASE v2.1.1 (Stephens et al., 2001), assuming absence of recombination and without using pedigree data. Twelve haplotypes in *LAA* were inferred (Table 2), of which 7 had frequencies >5%. All 7 most frequent haplotypes were found in the Holstein-Friesian bulls. In each of the other breeds, 1 or more of the 7 most frequent haplotypes were not observed. Haplotype 4, comprising

Table 1. Polymorphisms in the coding and promoter regions of the α -LA gene (*LAA*)¹

Polymorphism	Nt. No.	Location	AA change	Reference ²	Allele frequency					
					Ayrshire (n = 11)	Brown Swiss (n = 20)	Simmental (n = 27)	Guernsey (n = 9)	Holstein-Friesian (n = 72)	Jersey (n = 17)
g.-1290G > T	662	Promoter	—	ss410758925	0	0.33	0.02	0	0.17	0
g.-1206C > A	746	Promoter	—	ss410758926	0.32	0.10	0.57	0.11	0.13	0.24
g.-1000T > C	M	Promoter	—	ss410758927	0.50	0.80	0.30	0.06	0.42	0.06
g.-788G > T	M	Promoter	—	ss410758928	0.09	0	0.04	0	0.02	0
g.15G > A	X	5'UTR	—	ss410758929	0.05	0	0	0	0.26	0
g.433T > C	X	Intron 1	—	ss410758930	0.18	0	0.39	0.11	0.08	0.21
g.600G > T	X	Exon 2	65Gln > His	ss469415640	0	0	0	0	0	0
g.1325G > A	X	Intron 3	—	ss410758931	0.05	0	0.02	0.11	0.15	0
g.1675T > C	X	Intron 3	—	ss410758932	0	0	0.02	0	0.15	0
g.1778T > C	X	3'UTR	—	rs43707866	0.05	0	0	0	0.26	0
g.1857G > A	X	3'UTR	—	ss410758934	0.05	0	0.04	0.11	0.16	0

¹Nucleotide changes, nucleotide numbers (Nt. No.), locations relative to *LAA* genomic DNA sequence M90645 (M, promoter region) or X06366 (X, coding region corresponding to α -LA protein variant B), and resulting AA changes. Homozygous differences in all investigated animals as compared to the reference sequence M90645: 486C > A, 1443_1444insT, and X06366: 759C > A, 792A > G, 1207G > A, 1231G > A, 1837-1841AGAGG > TATAA, 2017T > C, 2651T > G. UTR = untranslated region.

²Reference according to the dbSNP database (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp>).

Table 2. Haplotypes comprising 10 polymorphisms in the coding and promoter regions of the α-LA gene (*LAA*)

Haplotype		Haplotype frequency						
No.	Sequence ¹	Ayrshire (n = 11)	Brown Swiss (n = 20)	Simmental (n = 27)	Guernsey (n = 9)	Holstein-Friesian (n = 72)	Jersey (n = 17)	Total (n = 156)
1	GCCGGTGTTG	0.50	0.48	0.26	0.06	0.24	0.06	0.263
2	GCTGGTGTTG	0	0.10	0.09	0.72	0.02	0.71	0.157
3	GATGGCGTTG	0.18	0	0.39	0.11	0.08	0.21	0.144
4	GCTGATGTCG	0.05	0	0	0	0.26	0	0.125
5	TCCGGTGTTG	0	0.33	0.02	0	0.17	0	0.122
6	GATGGTGTTG	0.14	0.10	0.17	0	0.05	0.03	0.077
7	GCTGGTACTA	0	0	0.02	0	0.15	0	0.074
8	GCTTGTGTTG	0.09	0	0	0	0.01	0	0.013
9	GCTGGTATTA	0.05	0	0	0.11	0	0	0.010
10	GCTGGTGTTA	0	0	0.02	0	0.01	0	0.006
11	GCCTGTGTTG	0	0	0.02	0	0.01	0	0.006
12	GATTGTGTTG	0	0	0.02	0	0	0	0.003

¹Haplotypes comprise SNP g.-1290G > T, g.-1206C > A, g.-1000T > C, g.-788G > T, g.15G > A, g.433T > C, g.1325G > A, g.1675T > C, g.1778T > C, and g.1857G > A in the coding and promoter regions of *LAA*.

genotypes g.15A and g.1778C, was the most frequent haplotype in Holstein-Friesian bulls, but it was not detected in Brown Swiss, Simmental, Guernsey, and Jersey bulls. Conversely, haplotype 2 was hardly found in Holstein-Friesian bulls, but it was the most frequent haplotype in Guernsey and Jersey bulls. Haplotype 1 was observed in all 6 breeds.

The cow that was heterozygous for SNP g.600G > T, corresponding to being heterozygous for α-LA protein variants B and D, was deduced to have haplotype 3 and haplotype 4 that each comprised 8 reference alleles (M90645, X06366) and 2 variant alleles. Because LD of SNP g.600G > T could not be established, this SNP was not included in the haplotype analysis.

ACKNOWLEDGMENTS

This study 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 Organisation CRV (Arnhem, the Netherlands), and the Dutch Technology Foundation STW (Utrecht, the Netherlands). The authors thank CRV, GGI Holland (IJsselmuiden, the Netherlands), and Semex (Leeuwarden, the Netherlands) for providing semen of the 156 bulls, and Rama Bangera and Katarzyna Wnuk-Lipińska (both former students at the Animal Breeding and Genomics Centre of Wageningen University) for their help with screening for genetic polymorphisms.

REFERENCES

Anderson, P. J., C. L. Brooks, and L. J. Berliner. 1997. Functional identification of calcium binding residues in bovine α-lactalbumin. *Biochemistry* 36:11648–11654.

Binkley, J., K. Karra, A. Kirby, M. Hosobuchi, E. A. Stone, and A. Sidow. 2010. ProPhyLER: A curated online resource for protein function and structure based on evolutionary constraint analyses. *Genome Res.* 20:142–154.

Bleck, G. T., and R. D. Bremel. 1993. Sequence and single-base polymorphisms of the bovine α-lactalbumin 5'-flanking region. *Gene* 126:213–218.

Bonfatti, V., G. Di Martino, A. Cecchinato, D. Vicario, 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.* 93:3797–3808.

Brew, K., T. C. Vanaman, and R. L. Hill. 1968. The role of α-lactalbumin and the A protein in lactose synthetase: A unique mechanism for the control of a biological system. *Proc. Natl. Acad. Sci. USA* 59:491–497.

Caroli, A. M., S. Chessa, and G. J. Erhardt. 2009. Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *J. Dairy Sci.* 92:5335–5352.

Farrell, H. M., 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.

Fitzgerald, D. K., U. Brodbeck, I. Kiyosawa, R. Mawal, B. Colvin, and K. E. Ebner. 1970. α-Lactalbumin and the lactose synthetase reaction. *J. Biol. Chem.* 245:2103–2108.

Ganai, N. A., H. Bovenhuis, J. A. M. van Arendonk, and M. H. P. W. Visker. 2009. Novel polymorphisms in the bovine β-lactoglobulin gene and their effects on β-lactoglobulin protein concentration in milk. *Anim. Genet.* 40:127–133.

Grobler, J. A., M. Wang, A. C. W. Pike, and K. Brew. 1994. Study by mutagenesis of the roles of two aromatic clusters of α-lactalbumin in aspects of its action in the lactose synthase system. *J. Biol. Chem.* 269:5106–5114.

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. 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.* 92:1192–1202.

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.* 82:1026–1033.

- Kumar, P., S. Henikoff, and P. C. Ng. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4:1073–1081.
- Lundén, A., and M. Lindersson. 1998. α -Lactalbumin polymorphism in relation to milk lactose. Pages 47–50 in Vol. 25 Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia. University of New England, Armidale, Australia.
- Malinovskii, V. A., J. Tian, J. A. Grobler, and K. Brew. 1996. Functional site in α -lactalbumin encompasses a region corresponding to a subsite in lysozyme and parts of two adjacent flexible structures. *Biochemistry* 35:9710–9715.
- Martins, L. F., M. P. Milazzotto, W. B. Feitosa, A. R. S. Coutinho, R. Simões, M. G. Marques, M. E. O. A. Assumpção, and J. A. Visintin. 2008. Sequence variation of the α -lactalbumin gene in Holstein and Nellore cows. *Anim. Biotechnol.* 19:195–199.
- Pettersson, J., A. K. Mossberg, and C. Svanborg. 2006. α -Lactalbumin species variation, HAMLET formation, and tumor cell death. *Biochem. Biophys. Res. Commun.* 345:260–270.
- Pike, A. C. W., K. Brew, and K. R. Acharya. 1996. Crystal structures of guinea-pig, goat and bovine α -lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure* 4:691–703.
- Rychel, A. L., T. W. Reeder, and A. Berta. 2004. Phylogeny of mysticete whales based on mitochondrial and nuclear data. *Mol. Phylogenet. Evol.* 32:892–901.
- 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.* 94:3148–3158.
- Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978–989.
- Stuart, D. I., K. R. Acharya, N. P. C. Walker, S. G. Smith, M. Lewis, and D. C. Phillips. 1986. α -Lactalbumin possesses a novel calcium binding loop. *Nature* 324:84–87.
- Threadgill, D. W., and J. E. Womack. 1990. Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Res.* 18:6935–6942.
- Vilotte, J.-L., S. Soulier, J.-C. Mercier, P. Gaye, D. Hue-Delahaie, and J.-P. Furet. 1987. Complete nucleotide sequence of bovine α -lactalbumin gene: Comparison with its rat counterpart. *Biochimie* 69:609–620.
- Voelker, G. R., G. T. Bleck, and M. B. Wheeler. 1997. Single-base polymorphisms within the 5' flanking region of the bovine α -lactalbumin gene. *J. Dairy Sci.* 80:194–197.