Analysis of MUC1 gene and its polymorphism in Capra hircus

(Article begins on next page)

08 January 2022

Analysis of the MUC1 Gene and its Polymorphism in *Capra hircus*

P. Sacchi,[1] A. Caroli,[2] E. Cauvin,[1] S. Maione,[1]
S. Sartore,[1] D. Soglia,[1] and R. Rasero[1]
[1]Dipartimento di Produzioni Animali, Epidemiologia ed Ecologia,
University of Torino, Grugliasco 10095, Italy
[2]Dipartimento di Sanità, Patologia,
Farmacotossicologia e Benessere degli Animali,
University of Bari, Valenzano 70010, Italy

ABSTRACT

The objective of our study was to demonstrate the existence of a repetitive region in the goat MUC1 gene and to develop a polymerase chain reaction (PCR) protocol to analyze its polymorphism in different breeds. Using 2 primers derived from the bovine MUC1 sequence, a PCR fragment was obtained and cloned. The sequence analysis shows that the repetitive region of goat MUC1 is an array of 60 bp repeats in accordance with the data reported for other species. The polypeptide sequences encoded by the consensus repeats of goat and bovine were exactly alike. A PCR protocol to improve the detection of goat MUC1 polymorphism was developed, and a total of 178 animals from 6 Italian breeds were analyzed. Fifteen different alleles, ranging in size from 1500 to 3000 bp, were found. The high number of alleles observed shows that the goat MUC1 is a hypervariable gene. These results are the basis for further investigations on the possible role of MUC1 polymorphism in the genetic control of disease susceptibility and production traits in the goat species. (**Key words:** goat, milk, MUC1, polymorphism)

Abbreviation key: AGE = agarose gel electrophoresis, **VNTR** = variable number tandem repeats.

INTRODUCTION

MUC1 is a mucin-type glycoprotein occurring in all mammalian species. It is expressed on the apical surface of simple epithelial tissues that line ducts and glands and exhibits a wide organ distribution, including the mammary gland. During lactation, the apical membrane of the secreting cell envelops cytoplasmic lipid droplets, giving rise to the milk fat globule membrane. MUC1 is an integral component of milk fat globule membrane (Mather, 2000).

MUC1 shows some structural features common to different species: a N-terminal cleavable signal domain, a large exoplasmic domain containing a repetitive region, a hydrophobic transmembrane domain, and a short C-terminal cytoplasmic tail. The repetitive region in humans and cattle shows a 20-amino acid motif, rich in serine and threonine residues, serving as sites for O-glycosylation. This motif is tandemly repeated a variable number of times (**VNTR**) in individual molecules, giving rise to a length polymorphism. The resulting molecular mass differences are detectable by polyacrylamide gel electrophoresis with SDS-PAGE (Patton et al., 1995).

In humans, MUC1 is extensively studied as a potential target for the antigen-specific immune therapy of cancer (Gendler, 2001). In domestic species, important roles in epithelia protection and milk production were suggested based on MUC1 structure and location (Patton, 1999). Furthermore, its VNTR polymorphism was found to be associated with yield, health, and reproductive traits in Holstein dairy cows (Hens et al., 1995).

Electrophoretic analysis has demonstrated that goat MUC1 is structurally related to human and bovine mucin and shows the same pattern of variation (Campana et al., 1992). Six different length variants with molecular mass >200 kDa were detected in 4 Italian goat breeds, suggesting the control of VNTR polymorphism by codominant alleles (Rasero et al., 1996). Nevertheless, analyses at the protein level are sex-limited and time-consuming. Moreover, the mobility of protein bands may be affected by different degrees of glycosylation. Although overall human and cattle MUC1 gene sequences are known (7 exons and 6 introns) and size differences between alleles were described in terms of variable number of 60 bp repeats into the repetitive region (Lancaster et al., 1990, Pallesen et al., 2001; Rasero et al., 2002), no data are available for goat MUC1 at present.

The fist aim of the present study was to demonstrate the existence of a repetitive region in the goat MUC1 gene. The second objective was to develop a PCR protocol suitable to analyze the polymorphism in different

breeds. The polymorphism in 6 Italian goat populations was then investigated.

MATERIALS AND METHODS

Sample Collection and DNA Extraction

Blood samples were collected from 178 individuals belonging to 6 Italian goat breeds (30 Camosciata, 23 Garganica, 32 Ionica, 28 Roccaverano, 32 Saanen, and 33 Vallesana). Genomic DNA was obtained from 200 μL of whole blood using the Invisorb Spin Blood Mini Kit (Invitek GmbH, Berlin, Germany).

PCR Amplification

Two oligonucleotide primers, P1 and P3, were selected among the primers previously used to amplify the bovine MUC1 repetitive region (Rasero et al., 2002). Amplifications were performed on a GeneAmp PCR System 2400 thermal cycler (Applied Biosystems, Foster City, CA). The best results were obtained by using the Expand Long Template PCR System (Hoffmann-La Roche, Basel, Swizerland) at the following conditions: a denaturation step at 94°C for 2 min, 10 cycles of 94°C for 10 s, 58°C for 30 s, and 68°C for 3 min, followed by 30 cycles of 94°C for 10 s, 50°C for 30 s, and 68°C for 3 min. A final extension step of 68°C for 7 min was carried out for all reactions. The amplified fragments were separated by agarose gel electrophoresis (**AGE**).

The efficiency of the P1-P3 primer set was first tested in some individual DNA samples randomly chosen. Subsequently, 3 (sire-dam-offspring) and 4 parent-offspring pairs were genotyped to analyze the segregation of the different electrophoretic pattern observed.

Sequencing Strategy

The PCR products were purified by the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany) and cloned into the pDrive Cloning Vector using QIAGEN PCR Cloning Kit. A cloned fragment was cycle-sequenced using an ABI PRISM 310 Genetic Analyzer by the dideoxy chain termination method with fluorescence dye terminators (Applied Biosystems). Sequencing on both strands was performed using the 2 vector primers as well as the P1 and P3 and further internal primers subsequently generated on the caprine sequences. A total of 12 primers were used (Roche Diagnostic Italia, Monza, Italy). The sequences obtained were compared and aligned by the ClustalW program (Thompson et al., 1994).
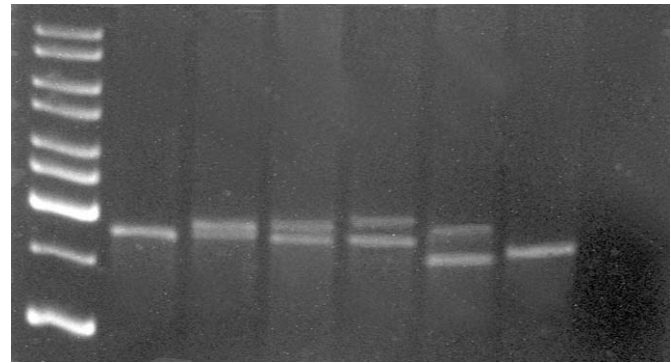
Figure 1. Agarose gel electrophoresis of P1-P3 PCR products of individual goat samples. Gibco 1-kb DNA Ladder, top to bottom: 10, 8, 6, 5, 4, 3.5, 3, 2.5, and 2 kb.

Sequence Analysis

The gene structure and coding sequence were predicted by alignment with the bovine sequence Gen-Bank AF399757. The corresponding peptide was deduced by the GENSCAN program (Burge and Karlin, 1997). Tandem Repeats Finder program (Benson, 1999) was used to identify the repetitive region.

Polymorphism Analysis

A standard PCR protocol was developed using a goat sequence-derived primer set flanking the repetitive region: PG9 (5′-GCAGTACTGTACAAGTTTCCAG-3′) and PG4 (5′-GTGTGCTGAAGTGACCATGGAG-3′). The FastStarTaq DNA Polymerase (Roche Diagnostics GmbH, Penzberg, Germany) was used at the following conditions: a denaturation step 95°C for 10 min and 10 cycles of 95°C for 30 s, 62°C for 30 s, and 72°C for 2.5 min followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 2.5 min. A final extension step of 72°C for 7 min was carried out for all reactions.

All the individual goats belonging to the 6 breeds were genotyped using the PG9-PG4 PCR protocol.

RESULTS AND DISCUSSION

PCR Amplification

Fragments of different size (2500 to 3000 bp) were obtained using the P1-P3 primer set with individual samples showing either 1 or 2 bands on AGE (Figure 1). The family data analysis showed that the observed differences segregated from parent to offspring according to the codominant Mendelian pattern of inheritance, as in humans and cattle (Swallow et al., 1987; Rasero et al., 2002).
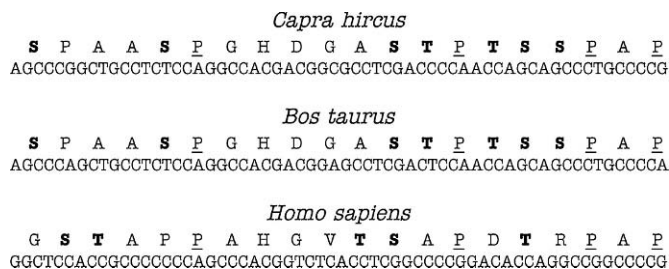
Capra hircus

S P A A **S** P G H D G A **S T** P **T S S** S P A P
AGCCCGGCTGCCTCTCCAGGCCACGACGGCGCCTCGACCCCAACCAGCAGCCCTGCCCCG

Bos taurus

S P A A **S** P G H D G A **S T** P **T S S** S P A P
AGCCCAGCTGCCTCTCCAGGCCACGACGGAGCCTCGACTCCAACCAGCAGCCCTGCCCCA

Homo sapiens

G **S T** A P P A H G V **T S** A P D **T** R P A P
GGCTCCACCGCCCCCCCAGCCCACGGTCTCACCTCGGCCCCGGACACCAGGCCGGCCCCG

Figure 2. Alignment of goat, bovine, and human MUC1 repeat consensus sequences at the DNA as well as at the protein levels. (The potential O-glycosylation sites are in boldface; the fully conserved proline residues are underlined).

Cloning and Sequence Analysis

An individual sample, showing a single band, was selected, and the corresponding PCR fragment (2800 bp) was cloned and sequenced. We were able to assemble 2 nonoverlapping sequences of 1053 bp (GenBank AY388993) and 540 bp (GenBank AY388994) corresponding to the 5′ and 3′ ends of the fragment, respectively. It was not possible to obtain the full-length sequence of the fragment because of the presence of repetitive structures from bp 706 to 1053 and from bp 1 to 163 in the 2 sequences, respectively. Such structures did not allow us to design further internal primers.

The nonrepetitive portions of the 2 goat sequences showed 92% homology both with the portion identified upstream (exon 1, intron 1, 5′ end of exon 2) and with that identified downstream (3′ end of exon 2) from the bovine repetitive region.

An array of 60 bp repeats was found in the goat-repetitive structure sequenced, in accordance with other species (Spicer et al., 1995), showing a 95% homology on average with the derived consensus repeat. Although the repeats are not all perfectly conserved, the degree of sequence homology is sufficiently high to preserve the polymorphism. In fact, the generation of new length alleles is based on misalignment of tandem repeat sequences followed by unequal exchange between sister chromatids. Because a minimum degree of homology is necessary for the repeats to recognize each other, the most polymorphic minisatellite loci are those with the most precise repeats (Olsen, 1999). In the mouse MUC1, which exhibits no polymorphism, the homology between repeats declines to 75% on average (Spicer et al., 1991).

The goat consensus repeat showed 93% homology with the bovine MUC1 consensus repeat (Rasero et al., 2002). The 2 sequences shared the high G and C content typical of mucin genes. On the other hand, the polypeptide sequences encoded by goat and bovine consensus repeats are perfectly identical (Figure 2).
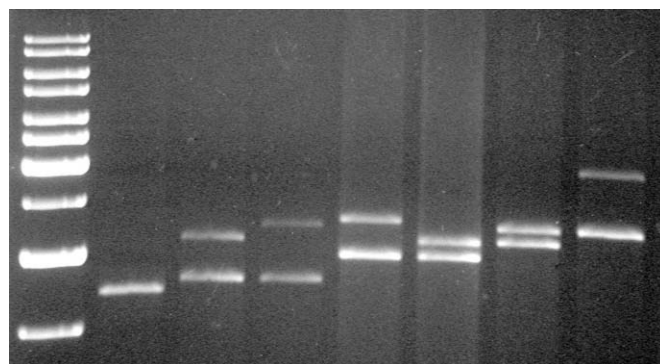

Figure 3. Agarose gel electrophoresis of PG9-PG4 PCR products of individual goat samples. Gibco 1-kb DNA Ladder, top to bottom: 10, 8, 6, 5, 4, 3.5, 3, 2.5, 2, and 1.5 kb.

The alignment of goat, bovine, and human consensus repeats at the protein level suggested that the presence of proline, serine, and threonine are selected for (Figure 2). Serine and threonine are the relevant amino acids for O-glycosylation. Therefore, goat core protein may be a highly O-glycosylated molecule, as already shown for human MUC1 (Patton et al., 1995).

On the basis of these results, it is possible to conclude that the fragment amplified using P1-P3 primer set corresponds to a portion of goat MUC1 gene containing the repetitive region. A size of about 1700 bp, corresponding to 28 repeats, was estimated for the repetitive region of the cloned fragment. Moreover, we suppose that the individual differences observed in goat are the expression of a different number of re-
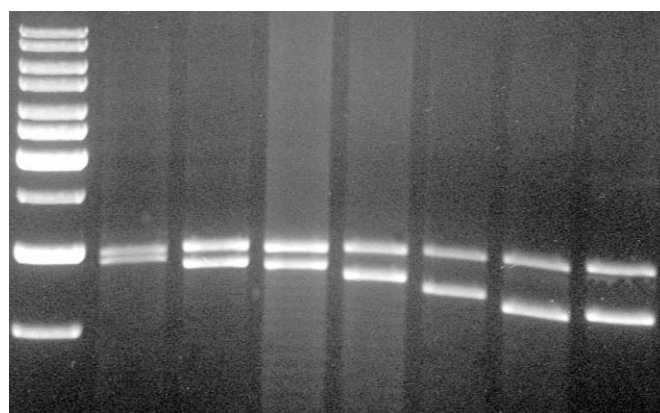

Figure 4. Agarose gel electrophoresis of PG9-PG4 products of individual goat samples. Lane 1, Gibco 1-kb DNA Ladder, top to bottom: 10, 8, 6, 5, 4, 3.5, 3, 2.5, 2, and 1.5 kb; lane 2, goat MUC1 heterozygote 5/6; lane 3, reference sample from a bovine heterozygote showing alleles differing for 2 repeats; lane 4, goat MUC1 4/6; lane 5, goat MUC1 3/6; lane 6, goat MUC1 2/6; lane 7, goat MUC1 1/6; and lane 8, reference sample from a bovine heterozygote showing alleles differing for 5 repeats.

Table 1. Number of alleles (n), observed heterozygosity (H), and allele frequency.

Breed	n	H	Allele frequency														
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Camosciata	9	0.77	0.00	0.15	0.24	0.10	0.07	0.18	0.07	0.05	0.00	0.00	0.00	0.12	0.00	0.02	0.00
Garganica	9	0.61	0.02	0.00	0.09	0.00	0.07	0.26	0.00	0.28	0.00	0.02	0.00	0.07	0.04	0.00	0.15
Ionica	9	0.78	0.00	0.05	0.14	0.00	0.06	0.35	0.06	0.17	0.11	0.03	0.00	0.00	0.03	0.00	0.00
Roccaverano	12	0.96	0.05	0.05	0.05	0.02	0.00	0.29	0.02	0.16	0.00	0.18	0.05	0.05	0.04	0.04	0.00
Saanen	10	0.84	0.06	0.19	0.11	0.05	0.11	0.19	0.06	0.02	0.01	0.00	0.00	0.20	0.00	0.00	0.00
Vallesana	7	0.82	0.05	0.15	0.30	0.09	0.09	0.08	0.00	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00

peats, according to the VNTR model assessed in other species, such as humans and cattle (Patton et al., 1995; Rasero et al., 2002).

Polymorphism Analysis

Some individual samples previously analyzed by P1-P3 primer set were also typed by the 2 most internal goat sequence-derived primers, PG9-PG4. A full correspondence was found between the individual number of bands and their relative mobility obtained by the 2 primer sets. The PG9-PG4 PCR protocol improved the AGE resolution, as the fragment size was reduced by 854 bp (Figure 3).

This approach allowed us to identify 15 different alleles ranging from 1500 to 3000 bp, which were numbered from 1 to 15 in ascending size order.

The differences between alleles could not be computed as number of repeats. In fact, the repetitive region had not been completely sequenced because of its size. So, an estimate of allelic differences was attempted using 2 different bovine heterozygous reference samples. Alleles differed for 2 (120 bp) to 5 (300 bp) repeats. Goat alleles 5 and 6 differed for a single repeat unit, whereas alleles 4 and 6 differed for 2 repeat units and so on (Figure 4). Therefore, we consider PG9-PG4 primers as particularly suitable for the analysis of goat MUC1 polymorphism, allowing the AGE detection of 60 bp differences between alleles.

Number of alleles, observed heterozygosity, and allele frequencies for each breed are reported in Table 1.

All goat populations showed a high degree of polymorphism, with a mean heterozygosity value of 0.8. Moderate size alleles were the most common as observed in 5 cattle breeds (Sacchi et al., 1995). The genetic structure of Vallesana was quite different from the other breeds, showing a low frequency of allele 6 and missing the longest alleles.

CONCLUSIONS

The goat MUC1 contains a repetitive region of 60 bp units similar to the homologous genes in other species.

The repetitive region of the goat MUC1 mucin is a modular domain that is rich in serine and threonine residues for addition of large quantities of glycans. The PG9-PG4 PCR protocol is suitable to detect differences of a 60 bp repeat unit between alleles. The high number of alleles observed in different breeds, despite the small number of individual samples tested, shows that the goat MUC1 is a hypervariable gene, and this can be the basis for further investigations about the possible role of MUC1 in the genetic control of disease susceptibility and production.

REFERENCES

Benson, G. 1999. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acid Res. 27:573–580.

Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268:78–94.

Campana, W. M., R. V. Josephson, and S. Patton. 1992. Presence and genetic polymorphism of an epithelial mucin in milk of the goat (*Capra hircus*). Comp. Biochem. Physiol. 103B:261–266.

Gendler, S. 2001. MUC1, the renaissance molecule. J. Mamm. Gland Biol. Neoplasia 6:339–353.

Hens, J. R., J. W. Rogers, M. L. Huott, and S. Patton. 1995. Association of the epithelial mucin, PAS-1, with yield, health and reproductive traits in Holstein dairy cows. J. Dairy Sci. 78:2473–2480.

Lancaster, C. A., N. Peat, T. Duhig, D. Willson, J. Tylor-Papadimitriou, and S. J. Gendler. 1990. Structure and expression of the human polymorphic epithelial mucin gene: An expressed VNTR unit. Biochem. Biophys. Res. Commun. 173:1019–1029.

Mather, I. H. 2000. A review and proposed nomenclature for major proteins of milk-fat globule membrane. J. Dairy Sci. 83:203–247.

Olsen, K. M. 1999. Minisatellite variation in a single-copy nuclear gene: Phylogenetic assessment of repeat length homoplasy and mutational mechanism. Mol. Biol. Evol. 16:1406–1409.

Pallesen, L. T., M. H. Andersen, R. L. Nielsen, L. Berglund, T. E. Petersen, L. K. Rasmussen, and J. T. Rasmussen. 2001. Purification of MUC1 from bovine milk-fat globules and characterization of a corresponding full-length cDNA clone. J. Dairy Sci. 84:2591–2598.

Patton, S., S. J. Gendler, and A. P. Spicer. 1995. The epithelial mucin, MUC1, of milk, mammary gland and other tissues. Biochim. Biophys. Acta 1241:407–424.

Patton, S. 1999. Some practical implications of the milk mucins. J. Dairy Sci. 82:1115–1117.

Rasero, R., E. Macchi, P. Giaccone, M. Panzera, and P. Sacchi. 1996. Electrophoretic analysis of the polymorphic epithelial mucin PAS-1 in four Sicilian goat populations. Agric. Med. 126:270–274.

Rasero, R., P. Sacchi, S. Rosati, E. Cauvin, and S. Maione. 2002. Molecular analysis of the length polymorphic MUC1 gene in cattle. J. Anim. Breed. Genet. 119:342–349.

Sacchi, P., E. Macchi, R. Rasero, and P. Fiandra. 1995. Two new variants of the bovine PAS-1 glycoprotein. Anim. Genet. 26:197–198.

Spicer, A. P., T. Duhig, B. S. Chilton, and S. J. Gendler. 1995. Analysis of mammalian MUC1 genes reveals potential functionally important domains. Mammal. Genome 6:885–888.

Spicer, A. P., G. Parry, S. Patton, and S. J. Gendler. 1991. Molecular cloning and analysis of the mouse homologue of the tumour-associated mucin, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism. J. Biol. Chem. 266:15099–15109.

Swallow, D. M., S. Gendler, B. Griffiths, G. Corney, J. Taylor-Papadimitriou, and M. E. Bramwell. 1987. The human tumour associated epithelial mucins are coded by an expressed highly variable gene locus PUM. Nature 328:82–84.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.