International Dairy Journal 131 (2022) 105387

Contents lists available at ScienceDirect

International Dairy Journal

journal homepage: www.elsevier.com/locate/idairyj

Proteomic characterisation and phylogenetic derivation of ovine α_{S1}-CN B and α_{S1}-CN G genetic variants

Sabrina De Pascale [a], Simonetta Caira [a], Giuseppina Garro [b], Rosalba Mauriello [b], Andrea Scaloni [a], Gianfranco Cosenza [b], Lina Chianese [b, *]

[a] Proteomics, Metabolomics & Mass Spectrometry Laboratory, ISPAAM-National Research Council, 80055 Portici, Naples, Italy
[b] Department of Agricultural Sciences, University of Naples Federico II, 80055 Portici, Naples, Italy

ARTICLE INFO

Article history:
Received 16 October 2021
Received in revised form
29 March 2022
Accepted 3 April 2022
Available online 9 April 2022

ABSTRACT

A combination of chromatographic, immunoelectrophoretic and mass spectrometric methods was applied to healthy individual milk of ovine massese breed. Caseome of α_{S1}-casein (α_{S1}-CN) fraction was characterised in homozygous and heterozygous phenotypes for the novel B and G genetic variants. Using ovine α_{S1}-CN C counterpart as reference, amino acid substitutions and phosphorylated components were determined. The replacement of p.Arg90>His and p.Ile194>Thr were found in α_{S1}-CN B variant, while the single p.Ile194>Thr substitution occurred in α_{S1}-CN G variant. The quali-quantitative composition of phosphorylated α_{S1}-CN components in milk of animals bearing each variant was defined, highlighting possible genetic markers for assessment of dairy product quality. The heterogeneity of the electrophoretic profiles of the corresponding caseomes was due to the number of phosphorylated components and their net charge and pI values, and updated the ovine α_{S1}-CN phylogenetic tree, suggesting α_{S1}-CN G as the ancestral variant of α_{S1}-CN B and other variants.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

The genetic polymorphism of milk proteins is the consequence of genic mutations due to impacting external events on the genome of mammals, which gave rise to different coding alleles at the same protein locus. Previous studies on the genetic polymorphism of homologous caprine and ovine α_{S1}-CN showed the occurrence of multiple alleles coding for this protein, differing from each other by: (i) one or more amino acid substitution(s); (ii) the length of the polypeptide chain; (iii) the quantity of expressed protein in the milk. In the latter context, numerous caprine α_{S1}-CN genetic variants, such as A, B1, B2, B3, B4, C, H, L and M, have been distinguished as "strong" variants, since they are expressed in higher amounts in milk (3.2 g L^{-1} per allele) than the others that, for the same reason, have been named "intermediate" (1.2 g L^{-1} per allele), namely α_{S1}-CN E, I, and D_1, and "weak" counterparts (0.6 g L^{-1} per allele), namely α_{S1}-CN D, F and G (Bevilacqua et al., 2002; Chianese, Ferranti, Garro, Mauriello, & Addeo, 1997a; Cunsolo et al., 2005; Garro et al., 2012; Grosclaude, Mahé, Brignon, Di Stasio, & Jeunet, 1987; Martin, 1993). Moreover, individual caprine milk samples lacking α_{S1}-CN for the presence of "null" alleles at this locus, such as α_{S1}-CN 01, 02 and N (Cosenza et al., 2003; Ramunno et al., 2005), have also been ascertained. Molecular mutations causing the above-mentioned different amounts of α_{S1}-CN genetic variants are quite different, from single nucleotide substitutions/deletions to large insertions/deletions (for review see Cosenza et al., 2008). Since the 3D structure of genetic variants are in any case involved, differences in both bio-functionality and technological properties of milk may be expected.

Previous studies in this research field showed a positive relationship between caprine α_{S1}-CN variants with important qualitative and quantitative differences also in the content of fat, urea level, fatty acid profile and milk yield (Cosenza et al., 2015). In general a greater digestibility of caprine milk containing α_{S1}-CN weak or null allele content (Bevilacqua et al., 2001) has been observed.

In comparison with the caprine counterpart, ovine α_{S1}-CN has been characterised later, taking advantage of the combined use of different electrophoretic procedures eventually coupled to specific immunostaining (Chianese et al., 1996) and mass spectrometry (Ferranti et al., 1995) protocols for the primary structure elucidation of protein variants. Until now, 9 genetic variants of ovine α_{S1}-CN have been identified, namely A, B, C, D, E, F, G, H and I, although not all characterised in detail (Calvo et al., 2013) (Table 1). Among α_{S1}-

* Corresponding author.
E-mail address: chianese@unina.it (L. Chianese).

https://doi.org/10.1016/j.idairyj.2022.105387
0958-6946/© 2022 Elsevier Ltd. All rights reserved.

CN variants, A, C, D and E have been characterised by means of a proteomic approach (Chianese et al., 1996; Ferranti et al., 1995; Garro, Caira, Lilla, Mauriello, & Chianese, 2019), B and F by immunoelectrophoresis (Chianese et al., 1996; Pirisi et al., 1999), G by RP-HPLC (Giambra, Chianese, Ferranti, & Erhardt, 2010a; Martin, Bianchi, Cebo, & Miranda, 2013a), and H and I by a combined genomic and proteomic approach (Giambra et al., 2010a; Giambra, Chianese, Ferranti, & Erhardt, 2010b). Genetic variants α_{S1}-CN A, C and D differ for single neutral amino acid substitutions, while E, H and I differ for deleted sequences otherwise encoded by exons 10, 8 and 7, respectively. In general, the occurrence of single amino acid substitutions and/or deleted sequences determined the loss of one or more phosphate groups in each protein variant, with respect to C reference (Table 1), negatively impacting on the corresponding milk mineral carrying activity and renneting (Chianese et al., 1997b; Pirisi et al., 1999). Moreover, each ovine α_{S1}-CN variant showed a very complex compositional heterogeneity due to a variable phosphorylation degree and the occurrence of constitutive non-allelic forms (Ferranti et al., 1998), similarly to caprine counterparts (Ferranti et al., 1997; Ramunno et al., 2005). The latter originated from: (i) the skipping of exon 13 codifying for sequence α_{S1}-CN(f110-117); (ii) the skipping of exon 16 codifying for sequence α_{S1}-CN(f141−148); (iii) a cryptic splice site within exon 11 encoding for Gln^{78} (Ferranti et al., 1998, 2001).

"Weak" α_{S1}-CN alleles expressing a low amount of the encoded protein, or "null" alleles were not fully identified in ovine milk so far. However, a low α_{S1}-CN content was assessed in the milk of ovine homozygotes for α_{S1}-CN D variant as well as a reduced total casein amount was verified in bulk milk associated with this phenotype, when compared with CD and CC counterparts (Pirisi et al., 1999). Similarly, a lower amount of α_{S1}-CN E (Garro et al., 2019) and α_{S1}-CN H (Giambra et al., 2010a) was also determined. Conversely, the amount of α_{S1}-CN I variant was comparable with that of the C counterpart (Giambra et al., 2010b). All these data suggested the need for further studies to clarify the corresponding gene expression machinery.

Frequency studies regarding ovine α_{S1}-CN variants showed the C as the most common in European breeds, while the other ones were found only in some European countries or specific regional districts. In the latter context, α_{S1}-CN D variant, which was firstly detected in Cluny Forest sheep (King, 1967), was rarely found in other European countries, including Italy, except in the Sardinia island, where it was found as the second most common phenotype (Pirisi et al., 1999; Ramunno et al., 1997). To date, E is the most rare α_{S1}-CN variant, being only found in Leccese breed reared in Puglia region (Garro et al., 2019). Similarly, α_{S1}-CN H and I variants were detected only in German East Friesian (Giambra et al., 2010a) and Gray Horned Heath (Giambra et al., 2010b) breed**s**, respectively.

Since the main use of ovine milk is the production of typical cheeses, which are safeguarded by protected designation of origin (PDO) label rules indicating the geographical origin of milking rearing breed, ovine α_{S1}-CN variants can be used as a useful markers to evaluate product quality, as already done in buffalo mozzarella cheese (Caira et al., 2019; Chianese et al., 2009). Having this aim in mind, a genetic screening of ovine α_{S1}-CN variants in individual milks from Italian Massese breed have been carried out by coupling different separation procedures to mass spectrometry (MS)-based protein characterisation techniques. This allowed obtaining original information on the structure and phosphorylation heterogeneity of α_{S1}-CN B and G variants.

In recent decades ruminant phylogenetics has undergone some remarkable revisions using a very large number of data deriving from analysis of genes and proteins. In particular the evolution of the primary structure of milk proteins, caseins and whey proteins, can surely represent the ideal tool to define the evolutionary relationships between the different species/breeds belonging to this mammalian group that have always represented an ideal model in which to study the process of speciation. To this aim the achieved results from B and G α_{S1}-CN characterisation have been used to supplement the phylogenetic tree of the ovine α_{S1}-CN genetic variants proposed for the first time by Ceriotti, Chiatti, Bolla, Martini, and Caroli (2005) and subsequently modified by Giambra (2011).

2. Materials and methods

2.1. Materials

All chemicals were of the highest purity commercially available and were used without further purification. N,N,N′,N′-tetramethylethylenediamine (TEMED), ammonium persulphate, glycine, urea, ammonium bicarbonate (AMBIC), water HPLC-grade, formic acid (FA) acetonitrile (ACN) acetic acid, sodium acetate and phosphate-buffered saline were purchased from Carlo Erba (Milan, Italy). 2-Mercaptoethanol (β-Me), trifluoroacetic acid (TFA), sinapinic acid, α-cyano-4-hydroxycinnamic acid, 2,5-dihydroxybenzoic acid (DHB), iodoacetamide and dithiotreitol (DTT) were obtained from Sigma Aldrich (St. Louis, MO, USA). Modified trypsin, sequencing grade, was purchased from Promega (Madison, WI, USA), while alkaline phosphatase (AP) and C_{18}ZipTip micro-columns were from Roche (Mannheim, Germany) and Millipore (Bedford, MA, USA), respectively. Acrylamide, bis-acrylamide and ampholine buffers (PharmalyteTM buffers) were purchased from (GE Healthcare Amersham Bioscience, Buckinghamshire, UK). Coomassie brillant blue (CBB) R250 and G250, 0.45 μm nitrocellulose membrane and macro-prep ceramic hydroxyapatite (HA) type I were purchased from Bio-Rad (Richmond, CA, USA).

Rabbit polyclonal antibodies raised against synthetic bovine peptide α_{S1}-CN (187−199), as antigen was purchased from Primm (Milan, Italy). C_{18}ZipTip devices (Millipore).

The selected genetic variants αS1-CN A, B, C and G came from a personal collection in our laboratories. Caseins were extracted using the methodologies described in this paper and used as reference standards.

2.2. Sheep milk sampling and casein fraction preparation

Individual samples of healthy Massese sheep's milk (146 in number), collected from different Tuscan farms (Italy) were skimmed by centrifugation at 2500×*g* for 15 min. Casein isoelectric precipitation (pH 4.6) was performed at 37 °C using 1 N acetic acid, and after 2 min incubation time, to ensure casein precipitation, the same volume of a 1 N sodium acetate buffer. After centrifugation (2500×*g*, 4 °C, 15 min), the acid casein pellet was washed three times before (Aschaffenburg & Drewry, 1959) freeze-dried and stored at −20 °C before further analysis.

2.3. Casein dephosphorylation

Alkaline phosphatase (AP) digestion was carried out on whole casein samples. Briefly, freeze-dried isoelectric casein casein samples (1 mg) were solubilised in 1 mL 0.4% (w/v) AMBIC, pH 8.5, and heated in a boiling water bath for 5 min for plasmin inhibition. After cooling at 4 °C, dephosphorylation was carried out with alkaline phosphatase by use of 2.5 mU enzyme mg^{-1} casein, at 37 °C, overnight. The reaction was stopped by freeze-drying and stored at −20 °C before further analysis.

2.4. RP-HPLC analysis

Native and dephosphorylated whole caseins were analysed using an HP1100 Agilent Technology modular system (Palo Alto, CA, USA) equipped with a Vydac (Hesperia, CA, USA) C4 column (250 × 4.6 mm ID, 5 μm particles) at a detection wavelength of 220 nm. Solvent A was 0.1% (v/v) TFA in HPLC-grade water, while solvent B was 0.1% (v/v) TFA in ACN. Whole freeze dried casein was solubilised (1 mg 1 mL^{-1}) with HPLC-grade water containing 0.1% TFA and 10 mM DTT. An aliquot of 200 μL of native and dephosphorylated samples was loaded onto a C4 column at room temperature. The elution conditions consisted of an isocratic gradient at 30% B for 5 min, followed by a linear gradient from 30 to 50% solvent B in 45 min, and then from 50 to 100% B in 2 min, at a flow rate of 1 mL min^{-1}. The α_{S1}-CN fractions for each selected genotype were manually collected and either used as such for mass spectrometry analysis by flow injection analysis (FIA) method, or freeze-dried for subsequent trypsin digestion.

2.5. Electrophoretic analyses and immunoblotting

Freeze-dried whole casein samples were dissolved in a 9 M urea solution (20 mg mL^{-1}) containing 2-mercaptoethanol (0.1 mg mL^{-1}). Seven microlitres of this solution were used for electrophoresis analysis.

Urea-polyacrylamide gel electrophoresis (urea-PAGE) at pH 8.6 was carried out with a vertical electrophoretic apparatus (Protean II, Bio-Rad) at 200 V and 6 °C, for 7 h. Protein staining was performed with CBB R250, according to the procedure of Chianese et al. (1996).

Ultra-thin layer isoelectric focusing (UTLIEF) on polyacrylamide gels (0.25 mm) was carried out in the pH gradient 2.5–6.5 obtained by mixing 1% (v/v) PharmalyteTM buffers: 2.5–5 (400 μL), 4.5–5.4 (350 μL) and 4–6.5 (250 μL). After electrophoretic run, the gels were stained with CBB G250 according to the procedure of Neuhoff, Arold, Taube, and Ehrardt (1988).

For immunoblotting analysis, the casein fractions resolved by UTLIEF were transferred by capillary diffusion from the gel onto a 0.45 μm nitrocellulose membrane. Immunodetection was carried out according to the procedure described previously (Chianese et al., 1996) using rabbit polyclonal antibodies raised against synthetic bovine peptide α_{S1}-CN (187–199), as antigen.

The relative percentage of each resolved α_{S1}-CN fraction in UTLIEF profiles was determined by scanning immunoblotting patterns with Image J software (http://rsb.info.nih.gov/ij), which was used to perform band aligning/matching between gels and corresponding densitometric analysis. The densitometric measurements for each α_{S1}-CN gel profile were performed in technical triplicate and the calculation of means and standard deviations were performed by Microsoft Excel 2017 to get the coefficient of variation % (CV %) as a statistical measure of the relative dispersion of data points. The values obtained were expressed as the mean with the standard deviation (SD). All statistical analyses were performed at a significance level of 5% ($p \leq 0.05$) using XLSTAT software Version 2020.1.

Theoretical pI and net charge of the different α_{S1}-CN genetic variants, were calculated based on the amino acid sequence of α_{S1}-CN A (UniprotKB P04653 - CASA1_SHEEP) and the different phosphorylation sites as post-translational modifications, using the Prot pi Protein Tool (https://www.protpi.ch).

2.6. Trypsin hydrolysis

HPLC manually collected α_{S1}-CN fractions were submitted to trypsin hydrolysis. The reaction was carried out in 0.4% ammonium bicarbonate (AMBIC), pH 8.0, at 37 °C for 4 h, using a 50:1 substrate-to-enzyme ratio (w/w). The reaction was stopped by freeze-drying and stored at −20 °C before further analysis.

2.7. Mass spectrometry analysis

2.7.1. Electrospray-quadrupole-time-of-flight mass spectrometry (ESI-QTOF–MS) analysis

Seven α_{S1}-CN phenotypes selected from HPLC analysis, i.e., four homozygotes (AA, BB, CC, GG) and three heterozygotes (AC, GC, BC), were submitted to mass spectrometry analysis. Native and dephosphorylated RP-HPLC casein fractions, dissolved at a concentration of 1 mg mL^{-1} in acetonitrile-water 50:50 (v/v) containing 0.1% TFA, were analysed by MS in FIA and positive ion modes using an ESI-Q-TOF™ hybrid mass spectrometer (Micromass Ltd., Manchester, UK) equipped with a Z-spray ion source. The nano flow was achieved using a syringe pump at 0.5 μL min^{-1}; the TOF mass analyser was used to acquire data in both MS and MS/MS modes. The source and desolvation temperature values were 100 and 200 °C, respectively. The TOF analyser operated at an acceleration voltage of 9.1 kV, a cone voltage of 100 V, a cone gas (N_2) of 13 L h^{-1} and a collision energy in MS mode of 10 eV.

2.7.2. NanoLC-ESI-Q-Orbitrap-MS/MS analysis

Tryptic digests of the seven selected α_{S1}-CN RP-HPLC fractions were analysed with a nanoLC-ESI-Q-Orbitrap-MS/MS platform consisting of an UltiMate 3000 HPLC RSLC nano system (Dionex, United States) coupled to a Q Exactive Plus mass spectrometer through a Nanoflex ion source (Thermo-Fisher Scientific). Freeze-dryed tryptic peptides mixture were dissolved at a concentration of 1 mg mL^{-1} in water HPLC-grade containing 0.1% TFA and 2 μL of samples were loaded on an Acclaim PepMap RSLC C18 column (150 mm × 75 μm ID, 2 μm particles, 100 Å pore size, Thermo-Fisher Scientific), and eluted with a gradient of 99.9/0.1 (v/v) water/formic acid (solvent A) and 19.92/80/0.08 (v/v/v) water/acetonitrile/formic acid) (solvent B), at a flow rate of 300 nL min^{-1}.

Peptides were separated applying a 5–60% gradient of B over 125 min at a 300 nL min^{-1} flow rate.

The mass spectrometer operated in data-dependent mode and all MS1 spectra were acquired in the positive ionisation mode with an *m/z* scan range of 375–1500. Up to 10 of the most intense ions in MS1 were selected for fragmentation in MS/MS mode. Precursor spectra were generated at a 70,000 full width at half maximum (FWHM) resolving power, with automatic gain control targets of 1 × 106 and 1 × 105 ions for full MS and MS/MS spectra, respectively, and a maximum ion injection time of 100 ms for both full MS and MS/MS spectra. MS/MS spectra were acquired in the *m/z* range 110–2000, using a normalised collision energy of 32%, an automatic gain control target of 100,000, a maximum ion target of 120 ms, and a resolution of 17,500. A dynamic exclusion value of 30 s was also used.

For protein identification, LC-MS/MS raw files were uploaded in the Proteome Discoverer v. 2.1 software (Thermo Scientific), enabling the database search by Mascot algorithm v.2.6 (Matrix Science, UK) and using the following criteria: (UniProtKB - P04653 - CASA1_SHEEP), including the most common protein contaminants; oxidation of Met, phosphorylation of Ser and Thr, deamidation of Asn and Gln, pyroglutamate formation of N-terminal Gln as variable modifications. Peptide mass tolerance was set to ±10 ppm and fragment mass tolerance to ±0.02 Da. Proteolytic enzyme and maximum number of missed cleavages were set to trypsin and 2, respectively. Peptide spectrum matches (PSMs) were filtered using the target decoy database approach with an e value of 0.01 peptide-level false discovery rate (FDR), corresponding to a 99% confidence score.

2.7.3. Matrix assisted laser desorption-induced ionisation-time-of-flight mass spectrometry (MALDI-TOF-MS) analysis

Tryptic digests of the seven selected α_{S1}-CN RP-HPLC fractions before and after casein phosphopeptide enrichment with hydroxyapatite (Pinto et al., 2010), were dissolved at a concentration of 1 mg mL^{-1} in water HPLC-grade containing 0.1% TFA. After desalting with C_{18}ZipTip devices, samples were mixed 1:1 (v/v) with a solution of 2,5-dihydroxybenzoic acid (DHB) (10 mg mL^{-1}) in 30% (v/v) ACN containing 0.1% (v/v) TFA. 0.5 μL of the resulting mixture was deposited onto a Ground steel target (Bruker Daltonics). Experiments were performed with an UltrafleXtreme instrument (Bruker Daltonics) according to previous studies (Arena, Salzano, & Scaloni, 2016). Spectral acquisition method was developed to maximise the number of signals present in the mass spectra and the corresponding signal-to-noise ratio. Spectra were recorded by automatically collecting 5 independent acquisition of 1000 shots per sample (random walk) within the *m/z* range 500–5000 Da, in positive linear mode (laser frequency, 1000 Hz; ion source 1 voltage, 25.19 kV; ion source 2 voltage, 23.94 kV; lens voltage, 6.50 kV; sample rate, 0.16). Mass spectra were externally calibrated using nearest neighbour positions loaded with Peptide calibration standard I and II (Bruker Daltonics), with quadratic calibration curves. Signals in the mass spectra were assigned to the corresponding tryptic peptides according to the expected molecular mass from the known ovine α_{S1}-CN A casein sequence, taking into account enzyme specificity, using several bioinformatics tools, such as the GPMAW 7.0 software (Lighthouse Data, Odense, Denmark). Semiquantitative analysis of peptide/phosphopeptides having the same sequence was performed assuming the same ionisation tendency for all considered compounds and evaluating the area of the corresponding mass signals. In particular, the relative abundance of each molecular species was calculated by normalising the corresponding area to the sum of the different areas associated with non-modified/modified species having the same sequence.

3. Results and discussion

3.1. RP-HPLC analysis of ovine casein samples

The primary structure of ovine α_{S1}-CN A, C, and D variants differs for "neutral" amino acid substitutions, which are located at key positions within triplet codes affecting protein phosphorylation, thus resulting in different phosphoproteins bearing up to 11, 10 and 6 phosphate groups, respectively (Table 1). Based on their different net charge, isoelectric point and hydrophobicity values, discrimination of above-mentioned genetic variants was obtained with PAGE at pH 8.6 and/or UTLIEF analysis (Chianese et al., 1996) and/or chromatographic investigations (Martin, Cebo, & Miranda, 2013b). Thus, a preliminary genetic screening of caseins from Massese ovine individual milks was carried out by RP-HPLC analysis. This allowed defining two main chromatographic profiles (Fig. 1), in which main ascertained α_{S1}-CN phenotypes were eluted as single (Fig. 1, panel A) (sample 1 to 4) or double (Fig. 2, panel B) (sample 5 to 7) chromatographic peaks respectively thus revealing the homozygous and heterozygous state.

Samples were numbered according to the increasing elution time order and the homozygous/heterozygous appearence exhibited by each phenotype. In particular, the α_{S1}-CN homozygous phenotypes eluted with four different elution time value respectively (Fig. 1, panel A); conversely, α_{S1}-CN phenotypes 5 to 7 eluted as heterozygous forms split in two chromatographic peaks having the same area (Fig. 1, panel B). Parallel RP-HPLC analysis of α_{S1}-CN AA, CC and GG phenotypes used as reference standards, allowed to hypothesise the identity of α_{S1}-CN genetic variants in the 7 samples analysed, considering that the latter variant has been characterised only by chromatography and not by other structurally informative methods (Table 1). Overall, RP-HPLC analysis allowed defining the occurrence of α_{S1}-CN variant A in samples 2 and 6, variant C in samples 4, 5, 6 and 7, variant G in samples 3 and 7, as well as ascertaining the presence of an unknown α_{S1}-CN variant (here named X) in samples 1 and 5. The latter showed the lowest hydrophobicity and occurred as homozygous and heterozygous form in sample 1 (Fig. 1, panel A) and 5 (Fig. 1, panel B), respectively.

3.2. Electrophoretic analysis of ovine casein samples

3.2.1. PAGE analysis at pH 8.6 of ovine casein samples

PAGE analysis (Fig. 2) of whole casein samples (lanes 1–7) selected for the above-mentioned α_{S1}-CN phenotypes was carried out in presence of reference standards of α_{S1}-CN A, B and C variants in the homozygous state (Chianese et al., 1996). A comparison of the corresponding PAGE migration profiles ascertained the occurrence of homozygous α_{S1}-CN B and A variants in lane 1 and 2 respectively, thus suggesting that variant X (detected in RP-HPLC analysis) coincides with variant B. On the other hand, homozygous phenotypes (lane 3 and 4), as well as heterozygous phenotype (lane 7), tentatively associated by RP-HPLC analysis to α_{S1}-CN GG, CC and GC, respectively, showed identical electrophoretic pattern of reference α_{S1}-CN CC phenotype. These findings suggested that a single neutral amino acid substitution characterised the hydrophobicity of α_{S1}-CN GG (lane 3) and CC (lane 4) phenotypes. A comparison of PAGE profiles of α_{S1}-CN CC and α_{S1}-CN AA confirmed the absence of the 11P component in the former (Ferranti et al., 1995). In this regard, even if the anodic mobility of α_{S1}-CN B was higher than that of α_{S1}-CN C, and was very similar to that of α_{S1}-CN A, a neutral amino

Table 1
Amino acid substitutions or sequence deletions characterising ovine α_{S1}-CN variants.[a]

Variant	Number of phosphates	Position											Reference
		12	13	41	43–50	51–58	64	66	68	70–77	90	194	
A*	11	Ser P	Ser										Ferranti et al. (1995)
B‡													Chianese et al. (1996)
C*	10	Ser	Pro	SerP			Ser P	Ser P	Ser P		Arg	Ile	Ferranti et al. (1995)
D*	6			Ser			Ser	Ser	Asn				Ferranti et al. (1995)
E*	6									Del.			Chianese et al. (1997, 2007)
F‡													Pirisi et al. (1999)
G¥													Martin (2013a); Giambra et al. (2010a)
H†	10					Del.							Giambra et al. (2010a)
I†	7				Del.								Giambra et al. (2010b)

[a] C is the reference variant; Del., sequence deletion. The analytical methods applied to characterise variants are as follows: * proteomic;† proteomic and genomic;‡ immunoelectrophoretic technique; ¥ chromatographic technique.

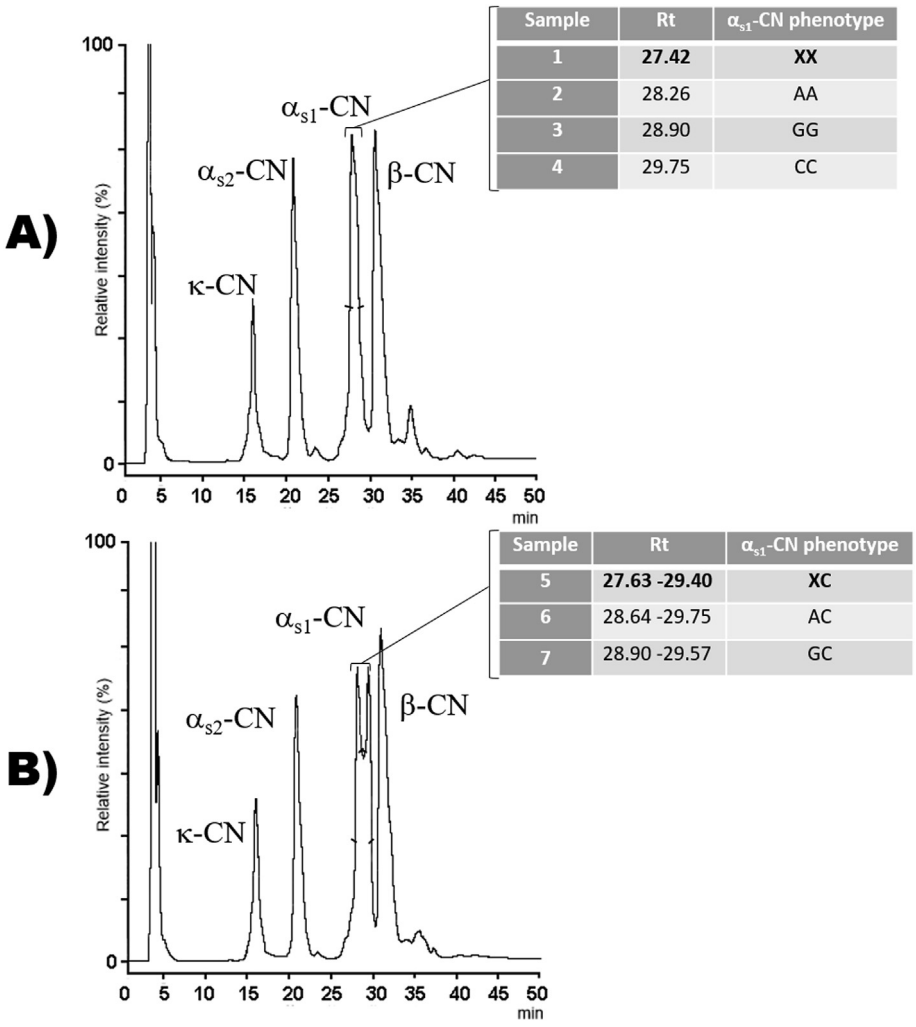

Fig. 1. Representative RP-HPLC profiles of the four homozygous (panel A) and three heterozygous (panel B) α_{S1}-CN phenotypes detected in ovine individual casein samples. Retention time (Rt) values of individual α_{S1}-CN genetic variants are also shown in the corresponding inserts for comparison. The start and end peak collection is marked.

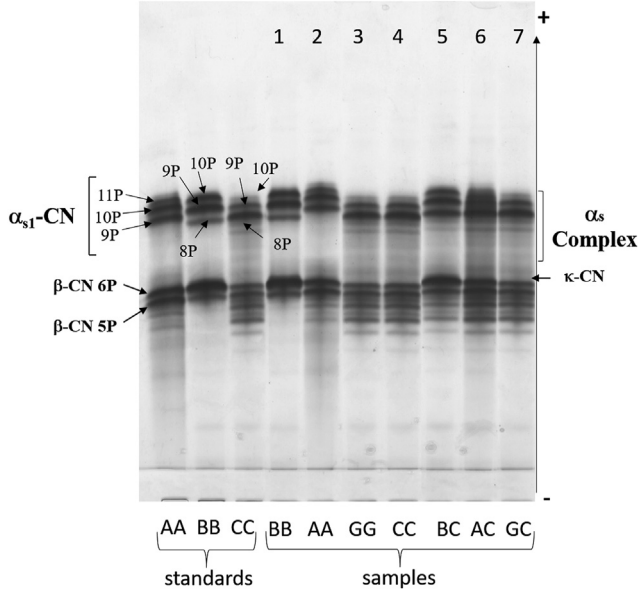

Fig. 2. Urea-PAGE pH 8.6 profiles of whole ovine casein samples containing the α_{S1}-CN phenotypes separated by HPLC analysis in the elution order reported on the top in comparison with the AA, BB, CC α_{S1}-CN references. Staining with Coomassie brilliant blue.

acid substitution differentiating α_{S1}-CN B and α_{S1}-CN A would be excluded based on RP-HPLC results.

3.2.2. UTLIEF analysis of ovine casein samples and anti-α_{S1}-CN specific immunoblotting

Highly resolutive UTLIEF analysis was then performed to further characterise the above-mentioned seven α_{S1}-CN phenotypes (Fig. 3), based on their focalisation in the low pH gel region. Resulting UTLIEF profiles, either after Coomassie brilliant blue (Fig. 3, panel A) or specific anti-α_{S1}-CN polyclonal antibodies (Fig. 3, panel B) staining, were more heterogeneous than the PAGE counterparts. This result was due to the enhanced capability of UTLIEF to highlight, in addition to the higher number of phosphorylated components, also the associated non-allelic forms (asterisk in Fig. 3) (Ferranti et al., 1998). The UTLIEF results were in agreement with the previous findings from chromatographic and PAGE analyses, namely samples 3, 4 and 7 corresponding to α_{S1}-CN G (lane 3), C (lane 4) and GC (lane 7) variants. Moreover, electrophoretic profiles of α_{S1}-CN B (lane 1) and A (lane 2) variants allowed highlighting the occurrence of the same number of phosphorylated components ranging from 7P to 11P. An analogous condition was verified for α_{S1}-CN C and G, both bearing from 5P to 10P. A densitometric analysis on the seven selected chromatographic profiles was then carried out on the corresponding immunostained profiles

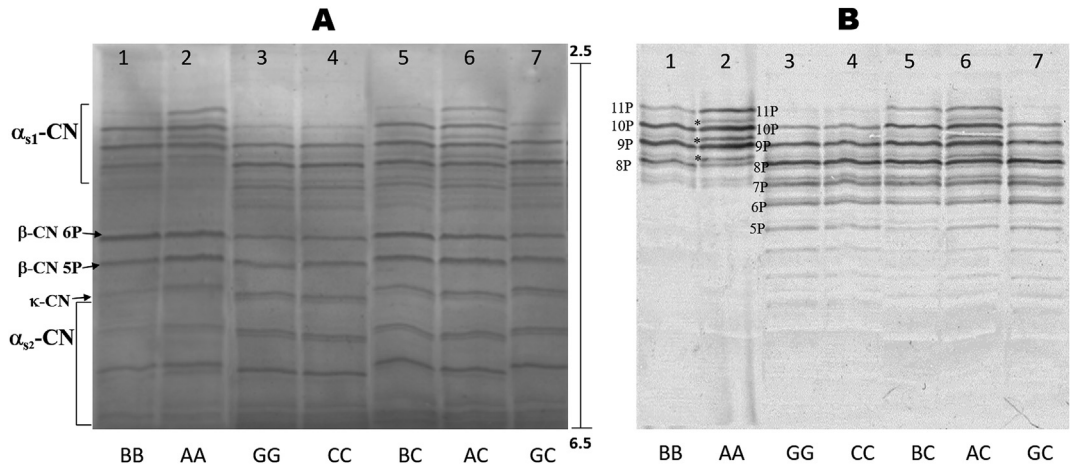

Fig. 3. UTLIEF profiles of the representative ovine α_{S1}-CN selected by HPLC in the elution order reported on the top, after Coomassie brilliant blue staining (panel A) and immunoblotting with polyclonal antibodies against the peptides α_{S1}-CN (f187−199) (panel B).

(Fig. 3, panel B), which allowed evaluating the percentage of the different phosphorylated components among the various α_{S1}-CN phenotypes. Resulting data defined the quali-quantitative composition of the phospho-caseome of each α_{S1}-CN variant, based on both phosphorylation degree and relative content of each component (Fig. 4). A summary of results is reported below: α_{S1}-CN B: 7P<8P<9P>10P>11P; α_{S1}-CN A: 8P<9P>10P>11P; α_{S1}-CN C = G: 5P<6P<7P<8P>9P>10P.

Although the qualitative trend of phosphorylated species present in α_{S1}-CN AA (lane 2) and α_{S1}-CN BB (lane 1) seemed very similar, the densitometric analysis revealed slight higher levels of the protein bearing 11P in α_{S1}-CN A than in B, whereas the component with 9P showed an opposite behaviour. On the other hand, the component with 8P resulted predominant in α_{S1}-CN GG and CC phenotypes (lane 3, 4), which also showed components bearing 7P, 6P and 5P (Fig. 3, panel B). These results would be explained based on the slight different structural characteristics of the various variants, which would have determined differential rates of phosphorylation by kinases.

3.3. Mass spectrometric characterisation of ovine α_{S1}-CN BB and α_{S1}-CN GG variants

3.3.1. ESI-qTOF-MS analysis of ovine α_{S1}-CN variants

RP-HPLC-resolved α_{S1}-CN variants treated or not with alkaline phosphatase were submitted to ESI-qTOF-MS analysis in FIA mode (Table 2) with the aim to determine their exact molecular mass value. Sequence data reported in UniProtKB database (P04653 - CASA1_SHEEP) and ascertained mass values confirmed the presence of the known α_{S1}-CN A and C genetic variants in homozygous form (samples 2 and 4), which showed a $\Delta M = 10$ Da derived from $Ser^{13} \rightarrow Pro^{13}$ substitution (Ferranti et al., 1995). In particular, the α_{S1}-CN A variant was verified as a 199 amino acid-long phosphorylated protein bearing from 8P to 11P. The failed detection of α_{S1}-CN 7P was due to its low relative amount, as deduced from the faintness of the corresponding immunostained band (Fig. 3B, lane 2).

The measured molecular mass of α_{S1}-CN B (sample 1) after alkaline phosphatase treatment resulted as being 31 Da lower than that of the reference α_{S1}-CN C counterpart (sample 4, Table 2). In

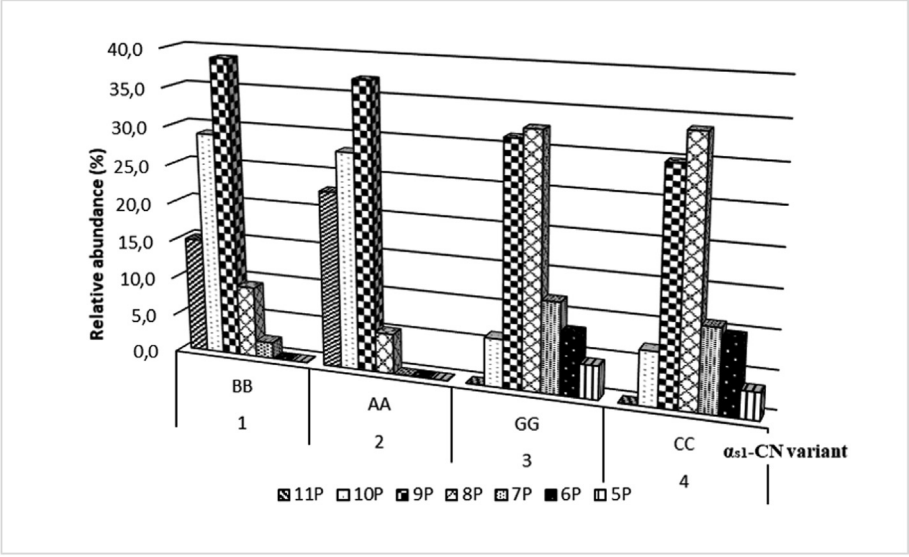

Fig. 4. Content of phosphorylated components in the four α_{S1}-CN variants,B, A, G, C by densitometric analysis of immunostained UTLIEF profiles (Fig. 3).

Table 2
Average molecular mass value of α_{S1}-CN samples as determined before and after alkaline phosphatase (AP) treatment.[a]

Sample	Protein	Molecular mass (Da)				Number of phosphate groupS	ΔM (Da)
		Native protein		After AP treatment			
		Theoretical	Measured	Theoretical	Measured		
1	α_{S1}-CN BB	23610.4	23609.7	22730.6	22729.8	11	31
		23530.4	23529.9			10	
		23450.4	23449.7			9	
		23370.4	23369.7			8	
		23290.4	23289.9			7	
2	α_{S1}-CN AA	23631.5	23630.9	22751.7	22750.9	11	10
		23551.5	23551.0			10	
		23471.5	23470.8			9	
		23391.5	23390.9			8	
3	α_{S1}-CN GG	23549.5	23548.9	22749.7	22749.1	10	12
		23469.5	23468.8			9	
		23389.5	23389.0			8	
		23309.5	23308.8			7	
		23229.5	23228.9			6	
4	α_{S1}-CN CC	23561.5	23560.4	22761.7	22761.1	10	
		23481.5	23480.7			9	
		23401.6	23400.7			8	
		23321.6	23321.0			7	
		23241.6	24240.9			6	

[a] α_{S1}-CN variants occurred as different phosphorylated species. ΔM is the difference between the molecular mass of α_{S1}-CN genetic variants with respect to the most common variant C.

agreement with UTLIEF results (Fig. 3, panel B), α_{S1}-CN B occurred as a mixture of phosphorylated species bearing from 7P to 11P. On the other hand, measured molecular mass of the α_{S1}-CN G (sample 3) was 12 Da lower than that of the reference α_{S1}-CN CC (Table 2). This was confirmed from the analysis of homozygous α_{S1}-CN G and C (sample 3 and 4) and heterozygous CG (sample 7) forms. In agreement with UTLIEF results (Fig. 3, panel B), α_{S1}-CN G occurred as a mixture of phosphorylated species whose pattern was very similar to that of α_{S1}-CN C (Table 2).

3.3.2. *MALDI-TOF-MS and nanoLC-ESI-Q-Orbitrap-MS/MS analysis of ovine α_{S1}-CN variants trypsin digest*

To obtain information on amino acid substitutions differentiating the various α_{S1}-CN variants detected in this study, RP-HPLC-resolved α_{S1}-CN species (samples 1–4) were digested with trypsin, and resulting peptide mixtures were analysed by MALDI-TOF-MS. Results highlighted polypeptide regions that were common to all phenotypes, while other portions varied between them, as summarised in Table 3.

In particular, signals at *m/z* 2193.1 Da and 1717.7 were selectively detected in samples 1, 3 and 4 (variant B, G and C), being absent in sample 2 (variant A). Based on their mass value and the known amino acid sequence of α_{S1}-CN C variant (sample 4), sample 1 also showed parallel signals at *m/z* 2274.3 and 1797.7, which were associated with the phosphorylated counterparts (ΔM = 80 Da), thus suggesting phosphorylation at Ser^{12}. Conversely, sample 2 (variant A) showed parallel phosphopeptide signals at *m/z* 2263.0 and 1787.7. All above-mentioned results were confirmed by nanoLC-ESI-Q-Orbitrap-MS/MS analysis, which definitively assigned the occurrence of Pro^{13} in samples 1, 3 and 4 (variant B, G and C) and Ser^{13} in sample 2 (variant A) (Table 3).

Mass signals at *m/z* 1362.4, 830.4 and 1266.6 occurring exclusively in samples 2, 3 and 4 (α_{S1}-CN A, G and C) and absent in sample 1, were assigned respectively to the peptides α_{S1}-CN(f80–90), (f84–90) and (f91–100). On the other hand, novel mass signals at *m/z* 2592.1 and 2059.9 corresponding respectively to the peptides α_{S1}-CN(f80–100) and (f84–100), were detected only in sample 1 (α_{S1}-CN B), indicating that the substitution of the amino acid Arg^{90}→His, no longer represents a specific hydrolysis trypsin site. Fig. 5 shows the tandem mass spectra of the tryptic peptides α_{S1}-CN(f80-100) (panel A) and the unchanged peptides α_{S1}-CN(f84-90) and α_{S1}-CN (f91-100) (panel B) respectively proving the occurrence of Arg^{90} in samples 2, 3 and 4 (α_{S1}-CN A, G and C) and His^{90} in sample 1 (α_{S1}-CN B) (Table 3).

Finally, the signal at *m/z* 759.3 detected in samples 2 and 4 (variant A and C) was absent in samples 1 and 3 (α_{S1}-CN B and G), which, conversely, showed the signal at *m/z* 747.4. Based on the mass value and the amino acid sequence of α_{S1}-CN C variant (sample 4), the two signals were associated with peptides α_{S1}-CN(f194–199) and the $p.Ile^{194}$→Thr amino acid substitution probably occurred. Fig. 6 shows the tandem mass spectra of both tryptic peptides definitively confirming the occurrence of α_{S1}-CN(f194–199)Thr^{194} in samples 2 and 4 (panel A) and α_{S1}-CN(f194–199)Ile^{194} in samples 1 and 3 (Panel B) (Table 3).

Overall, the detection of amino acid substitutions in samples 1 and 3 allowed the characterisation of the primary structure of the α_{S1}- CN variants B and G and the rationalisation of the measured mass difference values for intact proteins with respect to the reference α_{S1}-CN C. In fact, the ΔM = 31 Da value, measured in α_{S1}-CN B with respect to the counterpart α_{S1}-CN C derives from two amino acid substitutions, namely $p.Arg^{90}$→His (ΔM = 19 Da) and $p.Ile^{194}$→Thr (ΔM = 12 Da). In contrast, the mass difference of 12 Da measured in α_{S1}-CN variant G respect to the counterpart α_{S1}-CN C derives from the only amino acid substitution $p.Ile^{194}$→Thr.

The occurrence of an ovine C′ α_{S1}-CN allele with a T→C transition at the 663rd position of the seventh exon, which is responsible for the $p.Ile^{194}$→Thr substitution here described, was already reported at the gene level (Ceriotti et al., 2005, 2004; Martin et al., 2013b), from the sequencing of the 17th exon. The same nucleotide substitution was also found after sequencing of the entire ovine allele α_{S1}-CN G (Calvo et al., 2013) but, in any case, the expressed protein was never decribed so far. The neutral $p.Ile^{194}$→Thr substitution present in α_{S1}-CN C and G variants does not have effect on protein net charge or pI value, making them indistinguishable from other phenotypes in electrophoretic analysis. On the other hand, the $p.Arg^{90}$→His substitution present in α_{S1}-CN B can be responsible for the exhibited higher negative net charge and the lower pI value). Neverthless the slight pI value

Table 3
Theoretical and measured molecular mass of the tryptic peptides identified in α_{S1}-CN variants following MALDI-TOF-MS and nanoLC-ESI-Orbitrap-MS/MS analysis.[a]

Peptide	α_{S1}-CN genetic variant	Sample	Molecular mass (Da)		AA substitutions	Deletion	Phosphate group	Amino acid sequence
			Theoretical	Measured				
4-22	B	1	2274.5	2274.3	Pro^{13}		1xPhospho [S9]	[K].HPIKHQGL**S**PEVLNENLLR.[F]
	B,G,C	1,3,4	2193.2	2193.1	Pro^{13}			[K].HPIKHQGLSPEVLNENLLR.[F]
4-22	A	2	2263.1	2263.0	Ser^{13}		1xPhospho [S9]	[K].HPIKHQGL**S**SEVLNENLLR.[F]
8-22	B	1,	1797.9	1797.7	Pro^{13}		1xPhospho [S5]	[K].HQGL**S**PEVLNENLLR.[F]
	B,G,C	1,3,4	1717.9	1717.7	Pro^{13}			[K].HQGLSPEVLNENLLR.[F]
8-22	A	2	1787.9	1786.7	Ser^{13}		1xPhospho [S5]	[K].HQGL**S**SEVLNENLLR.[F]
	A	2	1707.9	1707.9	Ser^{13}			[K].HQGLSSEVLNENLLR.[F]
23-33	A,B,G,C	All	1306.7	1306.5				[R].FVVAPFPEVFR.[K]
23-34	A,B,G,C	All	1434.8	1434. 6				[R].FVVAPFPEVFRK.[E]
34-42	A,B,G,C	All	1073.6	1073.5				[R].KENINELSK.[D]
34-58	A,B,G,C	All	3032.2	3032.0			3xPhospho [S8; S13; S15]	[R].KENINEL**S**KDIG**S**E**S**IEDQAMEDAK.[Q]
	B,G,C	1,3,4	2952.2	2951.9			2xPhospho [S8; S15]	[R].KENINEL**S**KDIGSE**S**IEDQAMEDAK.[Q]
	B,G,C	1,3,4	2872.3	2872.2			1xPhospho [S8]	[R].KENINEL**S**KDIGSESIEDQAMEDAK.[Q]
35-58	A,B,G,C	All	2904.1	2903.9			3xPhospho [S7; S12; S14]	[K-].ENINEL**S**KDIG**S**E**S**IEDQAMEDAK.[Q-]
	B,G,C	1,3,4	2824.1	2824.0			2xPhospho [S7; S14]	[K-].ENINEL**S**KDIGSE**S**IEDQAMEDAK.[Q-]
	B,G,C	1,3,4	2744.2	2744.1			1xPhospho [S7]	[K-].ENINEL**S**KDIGSESIEDQAMEDAK.[Q-]
43-58	A,B,G,C	All	1896.7	1896.5			2xPhospho [S4; S6]	[K-].DIG**S**E**S**IEDQAMEDAK.[Q-]
	B,G,C	1,3,4	1816.7	1816.5			1xPhospho [S6]	[K-].DIGSE**S**IEDQAMEDAK.[Q-]
62-79	A,B,G,C	All	2285.6	2285.5			6xPhospho [S3; S4; S5; S6; S7; S14]	[K].AG**SSSSS**EEIVPN**S**AEQK.[Y]
	B,C	1,4	2205.7	2205.6			5xPhospho [S4; S5; S6; S7; S14]	[K].AGS**SSSS**EEIVPN**S**AEQK.[Y]
	B,C	1,4	2125.7	2125,4			4xPhospho [S5; S6; S7; S14]	[K].AGSS**SSS**EEIVPN**S**AEQK.[Y]
	C	4	2045.7	2045.6			3xPhospho [S6; S7; S14]	[K].AGSSS**SS**EEIVPN**S**AEQK.[Y]
62-79	A,B,G,C	All	2157.7	2157.5		- Gln^{78}	6xPhospho [S3; S4; S5; S6; S7; S14]	[K].AG**SSSSS**EEIVPN**S**AEK.[Y]
	B,C	1,4	2077.6	2077.5		- Gln^{78}	5xPhospho [S4; S5; S6; S7; S14]	[K].AGS**SSSS**EEIVPN**S**AEK.[Y]
	B,C	1,4	1997.6	1997.6		- Gln^{78}	4xPhospho [S5; S6; S7; S14]	[K].AGSS**SSS**EEIVPN**S**AEK.[Y]
80-90	A,G,C	2,3,4	1362.7	1362.4	Arg^{90}			[K].YIQKEDVPSER.[Y]
80-100	B	1	2592.3	2592.1	His^{90}			[K].YIQKEDVPSEHYLGYLEQLLR.[L]
84-90	A,G,C	2,3,4	830.4	830.4	Arg^{90}			[K].EDVPSER.[Y]
84-100	B	1	2060.0	2059.9	His^{90}			[K].YIQKEDVPSEHYLGYLEQLLR.[L]
91-100	A,G,C	2,3,4	1266.7	1266.6	Arg^{90}			[R].YLGYLEQLLR.[L]
101-124	A,B,G,C	All	1955.1	1954.8		-(110-117)		[R].LKKYNVPQL……..EQLHSMK.[E]
103-114	A,B,G,C	All	1426.8	1426.5				[K].KYNVPQLEIVPK.[S]
104-114	A,B,G,C	All	1298.7	1298.5				[K].YNVPQLEIVPK.[S]
104-124	A,B,G,C	All	1585.8	1585.7		-(110-117)		[K].YNVPQL……..EQLHSMK.[E]
115-124	A,B,G,C	All	1158.5	1158.1				[K].SAEEQLHSMK.[E]
	A,B,G,C	All	1238.5	1238.2			1xPhospho [S1]	[K].**S**AEEQLHSMK.[E]
125-132	A,B,G,C	All	879.4	879.1				[K].EGNPAHQK.[Q]
133-151	A,B,G,C	All	2327.2	2326.9				[K].QPMIAVNQELAYFYPQLFR.[Q]
133-151	A,B,G,C	All	1315.7	1315.4		-(141-148)		[K].QPMIAVNQ……LFR.[Q]
152-193	A,B,G,C	All	4657.2	4656.9				[R].QFYQLDAYPSGAWYYLPLGTQYTDA PSFSDIPNPIGSENSGK.[IT]
194-199	A,C	2,4	759.4	759.3	Ile^{194}			[W].ITMPLW
194-199	B,G	1,3	747.4	747.4	Thr^{194}			[W].TTMPLW

[a] Reported monoisotopic measured molecular mass values derived from MALDI-TOF-MS analysis, while sequence information was obtained from nanoLC-ESI-Orbitrap-MS/MS. **S** indicates phosphorylated residue.

difference between α_{S1}-CN B and α_{S1}-CN A genetic variants was too low (0.001 pH unit) (Table 4), to properly discriminate their phosphorylated components in the corresponding UTLIEF analysis (Fig. 3, panel B).

Based on above-reported considerations, others might be the structural characteristics responsible for the electrophoretic heterogeneity exhibited by different ovine α_{S1}-CN variants. In this context, worth mentioning are the p.Pro^{13}→Ser substitution present uniquely in variant A and the significant protein phosphorylation heterogeneity observed following ESI-qTOF-MS analysis of these molecules (Table 2).

3.3.3. *MALDI-TOF-MS semiquantitative analysis of phosphorylated tryptic peptides*

To tentatively determine the quali-quantitative phosphorylation heterogeneity of ovine α_{S1}-CN variants, we evaluated the relative intensity ratios of the mass signals of the different non-phosphorylated and phosphorylated forms of the four tryptic peptides α_{S1}-CN (f8−22), (f35−58), (f62−79) and (f115−124) present in ovine α_{S1}-CN variants. Results are summarised in Table 5.

In particular, no phosphorylation was observed at Ser^{12} in α_{S1}-CN(f8−22) in both α_{S1}-CN C and α_{S1}-CN G variants, while a slight and a significant modification was detected in B and A counterparts, respectively (Table 3). Thus, the presence of Pro^{13} or other substitutions in other polypeptide regions may have exerted a certain structural hindrance against the action of kinases phosphorylating Ser^{12}.

On the other hand, phosphopeptide α_{S1}-CN(f35−58) was associated with mass signals at *m/z* 2905.1, 2825.1, and 2745.2 Da, which correspond to components having 3, 2 and 1 phosphate groups, respectively. The latter two species were absent in the α_{S1}-CN A digest, which otherwise showed only α_{S1}-CN(f35−58)3P. Conversely, the relative intensity of signals in the B, G and C always highlighted the same trend, namely α_{S1}-CN(f35−58)3P > 2P > 1P.

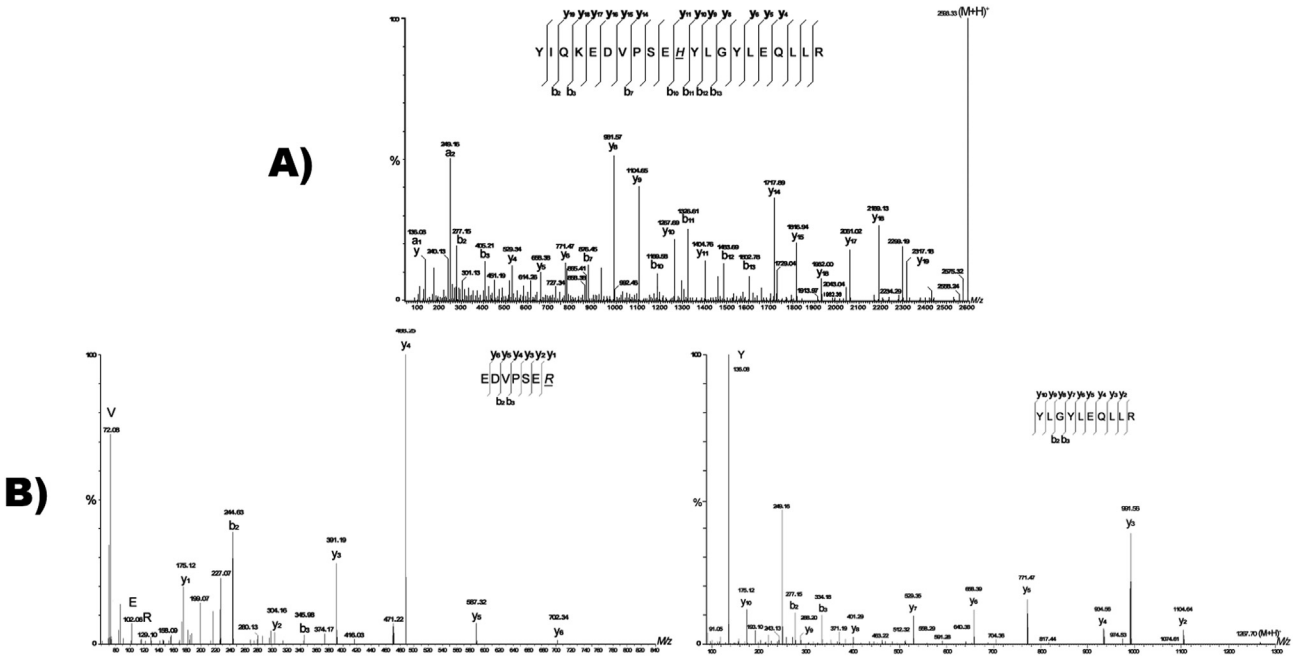

Fig. 5. NanoLC-ESI-Q-Orbitrap-MS/MS analysis of peptide at *m/z* 2593.3 Da (panel A) selectively detected in sample 1 (B variant of α_{S1}-CN) and assigned to peptide α_{S1}-CN(f80–100) bearing His^{90}. The MSMS spectrum of the corresponding reference tryptic peptides (α_{S1}-CN A variant) at *m/z* 831. 4 and 1267.70 were shown (panel B). The His^{90} and Arg^{90} amino acid substitutions are underlined.

Table 4
Theoretical values of pI and of net charge at pH 8.6 shown by AA, BB, CC and GG α_{S1}-CN variants at different degrees of phosphorylation.[a]

Variant	Number of phosphate groups	pI	Net charge (pH 8.6)
α_{S1}-CN BB	11	4.328	−32.2
α_{S1}-CN AA	11	4.329	−31.2
α_{S1}-CN BB	10	4.392	−30.2
α_{S1}-CN AA	10	4.393	−29.2
α_{S1}-CN CC/GG	10	4.393	−29.2
α_{S1}-CN BB	9	4.456	−28.2
α_{S1}-CN AA	9	4.457	−27.2
α_{S1}-CN CC/GG	9	4.457	−27.2
α_{S1}-CN BB	8	4.521	−26.2
α_{S1}-CN AA	8	4.523	−25.2
α_{S1}-CN CC/GG	8	4.523	−25.2
α_{S1}-CN BB	7	4.586	−25.2
α_{S1}-CN AA	7	4.588	−24.2
α_{S1}-CN CC/GG	7	4.588	−24.2

[a] pI and net charge derived using https://www.protpi.ch/Calculator/ProteinTool.

However, while the former two homozygous forms showed identical phosphopeptide ratios (about 75:18:7), the latter one evidenced another distribution (57:33:10), suggesting a general reduction of peptide modification in the C variant. In all cases, tandem mass spectrometric analysis demonstrated that the lack of the phosphate group in α_{S1}-CN(f35–58)2P and α_{S1}-CN(f35–58)P generally affected Ser^{46} and Ser^{48}, being SerP^{41} always detected (Table 3).

A similar condition was observed for the tryptic phosphopeptide α_{S1}-CN (f62–79), which occurred as hexaphosphorylated species (6P) in α_{S1}-CN AA, while in the other variants showed the intensity trend 6P>5P>4P (Table 5). In the latter context, the G variant showed a higher predominant representation (91%) of the peptide α_{S1}-CN(f62–79)6P with respect to B and C counterparts (about 70–74%), which also showed identical relative levels of α_{S1}-CN(f62–79)5P (13–14%). A progressive lack of modification was observed in this peptide at residues 65, 66 and 67 (Table 5), suggesting a lower rate of phosphorylation of this molecular portion probably due to the steric hindrance of the phosphate group located on the Ser at the C terminus of the triplet code.

Finally, the signal intensity distribution of non-phosphorylated and phosphorylated peptide α_{S1}-CN(f115–124) was 96:4, 70:30, −72:28 and 55:45 in α_{S1}-CN A, B,C and G digests, respectively (Table 5).

3.4. Phylogenetic derivation of ovine B and G α_{S1}-CN genetic variants

The construction of α_{S1}-CN phylogenetic tree was based on the observed similarities and differences within the amino acids sequence derived from studies on bovine (for a review see Caroli, Chessa, & Erhardt, 2009), caprine (for a review see Cosenza et al., 2008) and ovine species (Martin et al., 2013a) and the results achieved in the present research work. The new phylogenetic tree (Fig. 7), enriched with the two characterised genetic variants G and B, allowed us to conclude the first as the ancestral ovine α_{S1}-CN variant according to higher amino acid sequence homology with cattle α_{S1}-CN B and goat α_{S1}-CN B1variants. Moreover, by several mutational events responsible for amino acid substitution or deletions starting from α_{S1}-CN G, we propose two different phylogenetic road maps. The first map is generated by the amino acid substitution p.Thr^{194}>Ile which characterises the more common α_{S1}-CN C variant. From the latter the variants A, D, E, H and I originated one by one in successive mutations. The second phylogenetic road map is generated directly from variant G by the amino acid substitution p.Arg^{90}→His responsible of the α_{S1}-CN B variant.

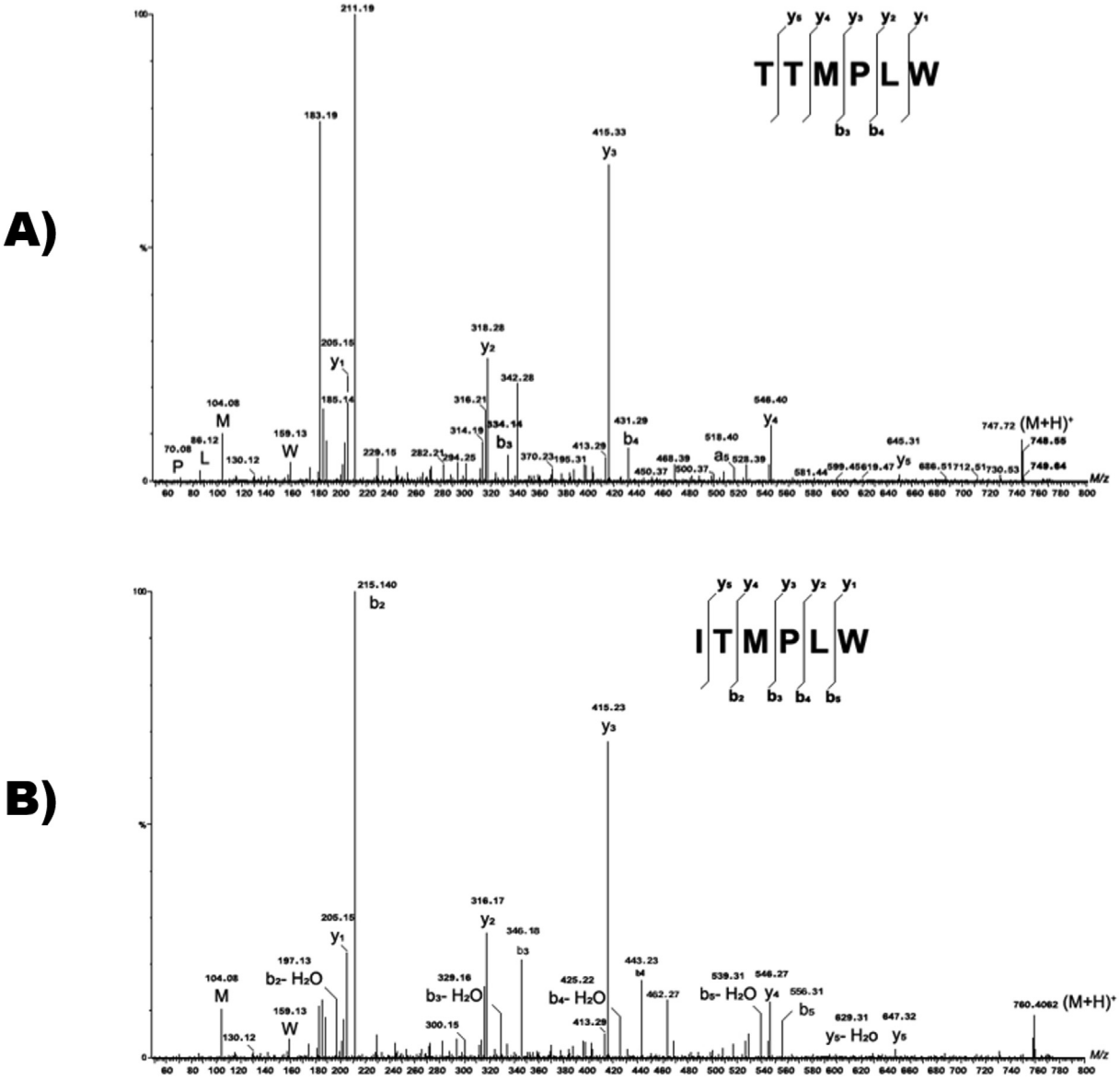

Fig. 6. NanoLC-ESI-Q-Orbitrap-MS/MS analysis of peptide at *m/z* 747.7 (panel A) selectively detected in samples 1 and 3 (α_{S1}-CN B and G variants) and assigned to peptide α_{S1}-CN(f194–199) bearing Thr[194]. The MSMS spectrum of the corresponding reference tryptic peptides with Ile[194] (α_{S1}-CN A variant) at *m/z* 760.4 was shown (panel B). The Thr[194] and Ile[194] amino acid substitutions are underlined.

Table 5
Mass signal intensity obtained by *MALDI-TOF-MS* analysis of the different forms of the tryptic phosphopeptides present in ovine α_{S1}-CN variants and their relative percentage ratios.

Peptide	Phosphate group	Sample/phenotype							
		1		2		3		4	
		BB		AA		GG		CC	
		Int	%	Int	%	Int	%	Int	%
8–22 Pro[13]	0P	94508	96			135979	100	587	100
	1P	3659	4			609	0		
8–22 Ser[13]	0P			12868	19				
	1P			55035	81				
35–58	3P	740	77	39	100	2389	75	111	57
	2P	168	17			551	18	64	33
	1P	61	6			233	7	19	10
62–79	6P	285	70	237	100	1793	91	2033	74
	5P	54	13			110	5		39514
	4P	69	17			72	4		2329
	3P								923
115–24	1P	9299	70	389	96	16586	55	29	72
	0P	3922	30	15	4	13815	45	11	28

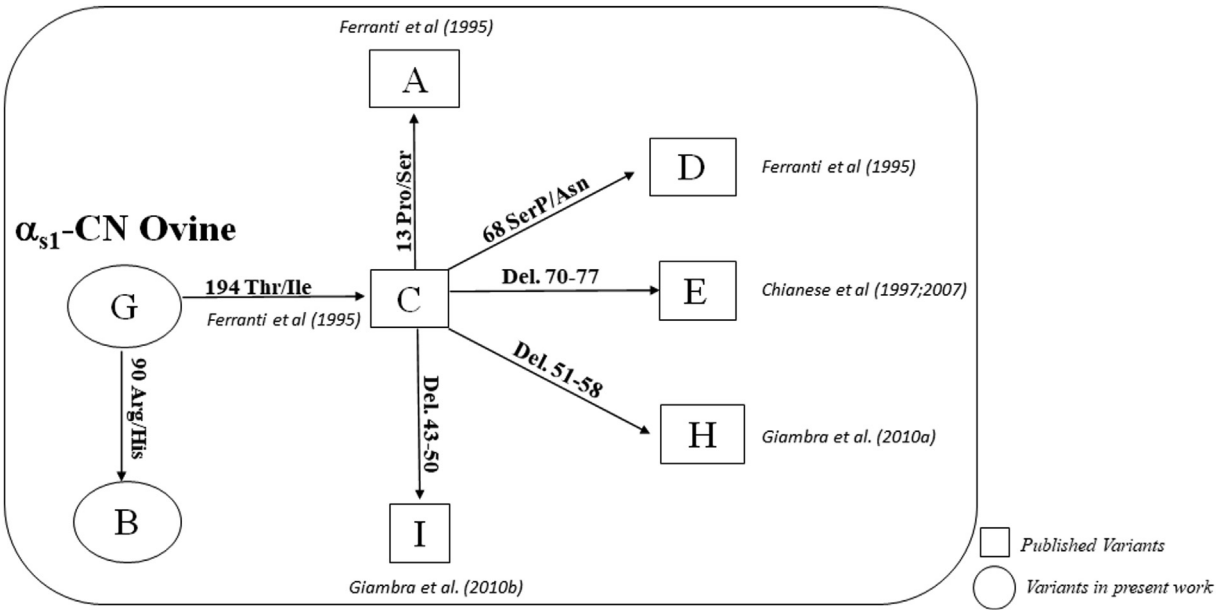

Fig. 7. Phylogenetic tree of ovine α_{S1}-CN genetic variants.

4. Conclusions

A combined use of sensitive and accurate chromagraphic, immunoelectrophoretic and mass spectrometric procedures allowed characterising the ovine α_{S1}-CN caseome of B and G variants, which evidenced charge-affecting and neutral residue replacements with respect to the most common α_{S1}-CN C counterpart, respectively. Above-mentioned variants and α_{S1}-CN A variant were also investigated for the number and distribution of corresponding phosphorylated components, which can be used in a near future as a genetic markers, when a breed origin certification is requested for trade reasons. Altogether, these data allowed us to define an updated phylogenetic tree for ovine α_{S1}-CNs, where the G variant was identified as the ancestral one preceding evolutionary development of the others reported in this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Arena, S., Salzano, A. M., & Scaloni, A. (2016). Identification of protein markers for the occurrence of defrosted material in milk through a MALDI-TOF-MS profiling approach. *Journal of Proteomics, 16*, 56–65.

Aschaffenburg, R., & Drewry, J. (1959). New procedure for the routine determination of the various non-casein proteins of milk. In *XVth International Dairy Congress* (vol. 3, pp. 1631–1637). London, UK.

Bevilacqua, C., Ferranti, P., Garro, G., Veltri, C., Lagonigro, R., Leroux, C., et al. (2002). Interallelic recombination is probably responsible for the occurrence of a new α_{S1}-casein variant found in the goat species. *European Journal of Biochemistry, 269*, 1293–1303.

Bevilacqua, C., Martin, P., Candalh, C., Fauquant, J., Piot, M., Roucayrol, A. M., et al. (2001). Goats' milk of defective α_{S1}-casein genotype decreases intestinal and systemic sensitization to β-lactoglobulin in Guinea pigs. *Journal of Dairy Research, 68*, 217–227.

Caira, S., Pinto, G., Nicolai, M. A., Novi, G., Addeo, F., & Scaloni, A. (2019). A non-canonical phosphorylation site in β-casein A from non-Mediterranean water buffalo makes quantifiable the adulteration of Italian milk with foreign material by combined isoelectrofocusing-immunoblotting procedures. *Food Chemistry, 277*, 195–204.

Calvo, J. H., Dervishi, E., Sarto, P., Gonzalalez-Calvo, L., Berzal-Herranz, B., Molino, F., et al. (2013). Structural and functional characterisation of the α_{S1}-casein (CSN1S1) gene and association studies with milk traits in Assaf sheep breed. *Livestock Production Science, 157*, 1–8.

Caroli, A. M., Chessa, S., & Erhardt, G. J. (2009). Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *Journal of Dairy Science, 92*, 5335–5352.

Ceriotti, G., Chessa, S., Bolla, P., Budelli, E., Bianchi, L., Duranti, E., et al. (2004). Single nucleotide polymorphisms in the ovine casein genes detected by polymerase chain reaction-single strand conformation polymorphism. *Journal of Dairy Science, 87*, 2606–2613.

Ceriotti, G., Chiatti, F., Bolla, P., Martini, M., & Caroli, A. (2005). Genetic variability of the ovine α_{S1}-casein. *Italian Journal of Animal Science, 4*, 64–66.

Chianese, L., Ferranti, P., Garro, G., Mauriello, R., & Addeo, F. (1997a). Occurrence of three novel α_{S1}-casein in goat milk. In *Milk protein polymorphism* (pp. 259–267). Brussels, Belgium: International Dairy Federation.

Chianese, L., Garro, G., Mauriello, R., Laezza, P., Ferranti, P., & Addeo, F. (1996). Occurrence of five α_{S1}-casein variants in ovine milk. *Journal of Dairy Research, 63*, 49–59. https://doi.org/10.1017/S0022029900031538

Chianese, L., Mauriello, R., Ferranti, P., Tripaldi, C., Taibi, L., & Dell'Aquila, S. (1997b). Relationship between α_{S1}-casein variants and clotting capability of ovine milk. In *Milk protein polymorphism* (pp. 316–323). Brussels, Belgium: International Dairy Federation.

Chianese, L., Quarto, M., Pizzolongo, F., Calabrese, M. G., Caira, S., Mauriello, R., et al. (2009). Occurrence of genetic polymorphism at the α_{S1}-casein locus in Mediterranean water buffalo milk. *International Dairy Journal, 19*, 181–189.

Cosenza, G., Illario, R., Rando, A., Di Gregorio, P., Masina, P., & Ramunno, L. (2003). Molecular characterization of the goat *CSN1S1*01 allele. *Journal of Dairy Research, 70*, 237–240.

Cosenza, G., Pauciullo, A., Gallo, D., Colimoro, L., D'avino, A., Mancusi, A., et al. (2008). Genotyping at the *CSN1S1 locus* by PCR-RFLP and AS-PCR in a Neapolitan goat population. *Small Ruminant Research, 74*, 84–90.

Cosenza, G., Pauciullo, A., Macciotta, N. P. P., Apicella, E., Steri, R., La Battaglia, A., et al. (2015). Mediterranean river buffalo *CSN1S1* gene: Search for polymorphisms and association studies. *Animal Production Science, 55*, 654–660.

Cunsolo, V., Galliano, F., Muccilli, V., Saletti, R., Marletta, D., Bordonaro, S., et al. (2005). Detection and characterization by high performance liquid chromatography and mass spectrometry of a goat β-casein associated with a CSN2 null allele. *Rapid Communications in Mass Spectrometry, 19*, 2943–2949.

Ferranti, P., Addeo, F., Malorni, A., Chianese, L., Leroux, C., & Martin, P. (1997). Differential splicing of pre-messenger RNA produces multiple forms of mature caprine α_{S1}-casein. *European Journal of Biochemistry, 249*, 1–7.

Ferranti, P., Chianese, L., Malorni, A., Migliaccio, F., Stingo, V., & Addeo, F. (1998). Copresence of deleted protein species generates structural heterogeneity of ovine α_{S1}-casein. *Journal of Agricultural and Food Chemistry, 46*, 411–416.

Ferranti, P., Malorni, A., Nitti, G., Laezza, P., Pizzano, R., Chianese, L., et al. (1995). Primary structure of ovine α_{S1}-caseins: Localization of phosphorylation sites and characterization of genetic variants A, C and D. *Journal of Dairy Research, 62*, 281–296.

Ferranti, P., Pizzano, R., Garro, G., Caira, S., Chianese, L., & Addeo, F. (2001). Mass spectrometry-based procedure for the identification of ovine casein heterogeneity. *Journal of Dairy Research, 68*, 35–51.

Garro, G., Caira, S., Lilla, S., Mauriello, R., & Chianese, L. (2019). Characterisation of the heterogeneity of ovine deleted variant α_{S1}-casein E by a proteomic approach. *International Dairy Journal, 89*, 53–59.

Garro, G., Ferranti, P., De Pascale, S., Nicolai, M. A., Mauriello, R., Quarto, M., et al. (2012). The occurrence of genetic polymorphism and related non-allelic proteins increases the composition complexity of goat α_{S1}-CN. *Electrophoresis, 33*, 2337–2344.

Giambra, I. J. (2011). *Ovine milk proteins: DNA, mRNA, and protein analyses and their associations to milk performance traits* (PhD thesis). Germany: Justus-Liebig-University Gießen, Faculty of Agricultural Sciences, Nutritional Sciences and Environmental Management.

Giambra, I. J., Chianese, L., Ferranti, P., & Erhardt, G. (2010a). Molecular genetic characterization of ovine α_{S1}-casein allele H caused by alternative splicing. *Journal of Dairy Science, 93*, 792–795.

Giambra, I. J., Chianese, L., Ferranti, P., & Erhardt, G. (2010b). Genomics and proteomics of deleted ovine CSN1S1*I. *International Dairy Journal, 20*, 195–202.

Grosclaude, F., Mahé, M. F., Brignon, G., Di Stasio, L., & Jeunet, R. (1987). A Mendelian polymorphism underlying quantitative variations of goat α_{S1}-casein. *Genetique Selection Evolution, 19*, 399–412.

King, J. B. W. (1967). The caseins of sheep's milk. In *European Conference on Animal Blood Groups and Biochemical Polymorphisms, no. 10, 1966. Polymorphisme biochimique des animaux* (pp. 427–431). Paris, France: INRA.

Martin, P. (1993). Polymorphisme génétique des lactoprotéines caprines. *Le Lait, 73*, 511–532.

Martin, P., Bianchi, L., Cebo, C., & Miranda, G. (2013a). Genetic polymorphism of milk proteins. In P. L. H. McSweeney, & P. F. Fox (Eds.), *Advanced dairy chemistry* (pp. 463–514). Boston, MA, USA: Springer.

Martin, P., Cebo, C., & Miranda, G. (2013b). Interspecies comparison of milk proteins: Quantitative variability and molecular diversity. In P. L. H. McSweeney, & P. F. Fox (Eds.), *Advanced dairy chemistry* (pp. 387–429). Boston, MA, USA: Springer.

Neuhoff, V., Arold, N., Taube, D., & Ehrhardt, W. (1988). Improved staining of polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis, 9*, 255–262.

Pinto, G., Caira, S., Cuollo, M., Lilla, S., Fierro, O., & Addeo, F. (2010). Hydroxyapatite as a concentrating probe for phosphoproteomic analyses. *Journal of Chromatography B, 878*, 2669–2678.

Pirisi, A., Piredda, G., Papoff, C. M., Di Salvo, R., Pintus, S., Garro, G., et al. (1999). Effects of sheep α_{S1}-casein CC, CD and DD genotypes on milk composition and cheesemaking properties. *Journal of Dairy Research, 66*, 409–419.

Ramunno, L., Cosenza, G., Rando, A., Macciotta, N. P. P., Pappalardo, M., & Masina, P. (1997). Identification of carriers of the Welsh CASA1 variant using an allele-specific polymerase chain reaction method. *Animal Genetics, 28*, 154–155.

Ramunno, L., Cosenza, G., Rando, A., Pauciullo, A., Illario, R., Gallo, D., et al. (2005). Comparative analysis of gene sequence of goat *CSN1S1* F and N alleles and characterization of *CSN1S1* transcript variants in mammary gland. *Gene, 345*, 289–299.