Copresence of Deleted Protein Species Generates Structural Heterogeneity of Ovine α_{s1}-Casein

Pasquale Ferranti,*[,†,‡] Lina Chianese,[†] Antonio Malorni,[‡] Francesco Migliaccio,[†] Vittorio Stingo,[†] and Francesco Addeo[†]

Dipartimento di Scienza degli Alimenti, Università degli Studi di Napoli Federico II, 80055 Portici, Italy, and Servizio di Spettrometria di Massa, CNR, Via Pansini 5, 80131 Napoli, Italy

Multiple forms of mature α_{s1}-casein have been characterized in ovine variants A and D using a combination of mass spectrometry and automated Edman degradation. Mature ovine α_{s1}-casein was found to be a heterogeneous mixture of at least seven molecular species. The main component, representing about 50% total α_{s1}-casein, corresponded to the full-length (199 residues long) protein. The other components were α_{s1}-casein of different lengths: 198 (less Gln78), 191 (less peptide 110−117), 191 residues (less peptide 140−148), 190 (less peptide 110−117 and Gln78), 190 (less peptide 140−148 and Gln78), and 183 (less peptides 110−117 and 140−148) residues long α_{s1}-casein. Each of the α_{s1}-casein multiple forms occurred at three different phosphorylation levels, due to the partial phosphorylation of both Ser115 (at about 50%) and Ser41 (at about 20%). In the case of deleted peptide 110−117, the protein heterogeneity linked to the partially phosphorylated Ser115 was abolished, and only two levels of phosphorylation were observed. These multiple forms differing in molecular weight and degree of phosphorylation may have been developed from an exon skipping during mRNA splicing in ovine α_{s1}-casein, similar to that recently described in the case of its caprine counterpart.

Keywords: *Ovine α_{s1}-casein; nonallelic deletion; polymorphism; mass spectrometry*

INTRODUCTION

Over the last few years, several studies have been devoted to genetic polymorphism of sheep caseins. By electrophoretic, immunochemical, and chromatographic analyses, it has been shown that ovine α_{s1}-casein exists as a number of distinct genetic variants both in the heterozygous and homozygous forms (Addeo et al., 1992). Five variants of ovine α_{s1}-casein have been described so far, associated with quantitative variation (Chianese et al., 1996a). The primary structure of three of them, A, C, and D, has been determined, differing from each other by amino acid substitutions and phosphorylation degree (Ferranti et al., 1995).

Recently, it was reported that ovine α_{s1}-casein consists of a mixture of two protein species differing for the deletion of peptide 141−148, the short form representing about 20% of the total ovine α_{s1}-casein fraction (Ferranti et al., 1995). The copresence of the full-length and the shortened forms of ovine α_{s1}-casein was shown for the five ovine variants reported so far (Chianese et al., 1996a), all of them also occurring at different degrees of phosphorylation. Electrophoretic evidence, however, still indicated for ovine α_{s1}-casein an additional yet unexplained heterogeneity (Chianese et al., 1996b) similar to that observed in the case of the caprine counterpart. In the latter case, the differential splicing of premessenger RNA was found to produce four differently sized forms of caprine α_{s1}-casein (Ferranti et al., 1998). In order to identify the origin of ovine α_{s1}-casein heterogeneity, a detailed study has been carried out on ovine α_{s1}-casein variants A and D, the former associated with a high rate and the latter with a low rate of protein synthesis, by using mass spectrometry techniques and automated sequencing. Mechanisms of primary transcript processing leading to protein formation at different lengths and at different phosphorylation levels are discussed.

MATERIALS AND METHODS

Reagents. Glycerol and thioglycerol were purchased from Sigma Chemical Co. (St. Louis, MO). Sequence-grade bovine pancreatic trypsin and alkaline phosphatase from calf intestine were obtained from Boehringer (Mannheim, Germany). HPLC-grade solvents and reagents were obtained from Carlo Erba (Milan, Italy). A Vydac (Hesperia, CA) C18 (218TP54, 250 × 4.6 mm, 5 μm) column was used for the tryptic peptide separation. A Vydac C4 (214TP52, 250 × 2.1 μm, 5 mm) column was used for LC/MS of caseins. Reagents for automated sequencing were supplied by the sequenator manufacturer (Applied Biosystems, Warrington, U.K.).

HPLC Separation of α_{s1}-Casein. Whole casein was prepared by acid precipitation of skimmed milks from individual animals homozygous for the A and D variants (Ferranti et al., 1995) of ovine α_{s1}-casein. The α_{s1}-casein samples were obtained from whole casein by reversed-phase (RP) HPLC using the procedure developed by Jaubert and Martin (1992) for goat caseins with the following modifications: solvent A was 1 mL/L trifluoroacetic acid (TFA) in water, solvent B was 0.8 mL/L TFA in 800 mL/L acetonitrile; before analysis, casein samples were dissolved in water (5 mg/mL) adjusting the pH to 7.0, reduced for 2 h at 37 °C with 10 mM dithiothreitol, and then diluted with solvent A to reach pH 2.0 and filtered through a 0.45 mm filter (Millipore). Elution was achieved using a linear gradient from 35 to 50% solvent B for 28 min at a flow rate of 1 mL/min. Samples used for analytical runs

* Author to whom correspondence should be addressed (tel +39 81 7463472; fax +39 81 5465484; e-mail ferranti@unina.it).
† Università degli Studi di Napoli Federico II.
‡ Servizio di Spettrometria di Massa.

S0021-8561(97)00609-2 CCC: $15.00
Published on Web 01/22/1998

were 100 μg in both systems; for preparative runs, 500 mg were injected per run. The column effluent was monitored at 220 nm.

Enzyme Digestion. Trypsin action was carried out in 4 g/L ammonium bicarbonate, pH 8.5, for 4 h at 37 °C in a substrate:enzyme ratio 100:1 (w/w). Dephosphorylation with calf intestine alkaline phosphatase was performed in the same buffer by using 1 mU enzyme:mg casein at 37 °C for 18 h. Under these conditions, complete sample dephosphorylation was obtained, as previously described (Ferranti et al., 1992, 1998). All the reactions were stopped by freeze-drying.

HPLC Separation of Tryptic Peptides. Tryptic digests of α_{s1}-casein were fractionated by HPLC using a Vydac C18 column at a flow rate of 1 mL/min; solvent A was 1 mL/L TFA in water; solvent B was 0.7 mL/L TFA in acetonitrile. After a 5 min hold at 4% solvent B, elution was performed by a linear gradient from 4 to 25%B for 25 min, followed by a linear gradient from 25 to 39% B for 39 min. The column effluent was monitored at 220 nm. Peptides eluted from the column were manually collected, dried down, and stored at −20 °C.

Electrospray Mass Spectrometry. Electrospray mass spectrometry (ES/MS) analysis of intact α_{s1}-casein was performed with a Platform single-quadrupole mass spectrometer (VG-Biotech, Altrincham, U.K.). Samples from HPLC separation (10 μL, 50 pmol) were injected into the ion source at a flow rate of 2 μL/min; the spectra were scanned from 1400 to 600 at 10 s/scan.

Mass scale calibration was carried out using the multiple charged ions of a separate introduction of myoglobin. Mass values are reported as average masses. The components identified were assigned to the corresponding protein species with the aid of a computer program (F. Barone, unpublished results). Quantitative analysis of components was performed by integration of the multiple charged ions of the single species (Ferranti et al., 1995).

Liquid Chromatography/Mass Spectrometry. Liquid chromatography/electrospray mass spectrometry (LC/ES/MS) was performed using a 2.1 mm i.d. × 250 mm, C_{18}, 5 μm (Vydac) reversed-phase column with a flow rate of 0.2 mL/min on a Kontron modular system. The column effluent was split 1:25 to give a flow rate of about 8 μL/min into the electrospray nebulizer. The bulk of the flow was run through the detector for peaks collection as measured by following the absorbance at 220 nm. Solvent A was 0.03% TFA (v/v) in water; solvent B was 0.02% TFA in acetonitrile. Separation of the peptides was carried out with a gradient 35−50% B in 60 min. The ES-mass spectra were scanned from 1800 to 400 u at a scan cycle of 5 s/scan. The source temperature was 120 °C, and the orifice voltage was 40 V.

Fast Atom Bombardment Mass Spectrometry. FAB/MS spectra were recorded on a ZAB 2SE (VG-Biotech, Altrincham, U.K.) double-focusing mass spectrometer fitted with a cesium gun operating at 25 kV (2 μA). Samples were dissolved in 1 mL/L HCl and loaded onto a glycerol-coated probe tip; thioglycerol was added to the matrix just before introduction of the probe into the ion source. Mass values are reported as monoisotopic masses.

Peptide Recognition. Signals recorded in the FAB spectra were associated with the corresponding tryptic peptides on the basis of the molecular mass calculated using a suitable computer program (F. Barone, unpublished results), taking into account the enzyme specificity and the amino sequence (Ferranti et al., 1995). In order to confirm the assignments, Edman degradation steps were manually performed on single peptides, followed by FAB/MS analysis on the truncated peptides (Pucci et al., 1989).

Sequence Analysis. Automated Edman degradation was performed using an Applied Biosystems Model 477A protein sequencer with on-line phenylthiohydantoinyl amino acid (Pth-Xaa)-HPLC analyzer. Starting with 250−500 pmol of peptide/protein purified by means of RP-HPLC, up to 40 cycles were performed.

Phosphorylated peptides were modified following the procedure described by Meyer et al. (1986). The peptide (500 pmol) was dried in a screw cap test tube and dissolved in 50 mL of a mixture containing 10 M ethanethiol/water/dimethyl sulfoxide/ethanol/5 M NaOH in the volume ratio 1/2.5/2.5/1.3/1. After incubation for 1 h at 50 °C under nitrogen, the peptide solution was cooled at room temperature and 10 mL of glacial acetic acid added. Aliquots of modified peptides were directly analyzed in the sequencer, and Pth-S-ethyl-cysteine was identified by HPLC according to the procedure described by Meyer et al. (1986).

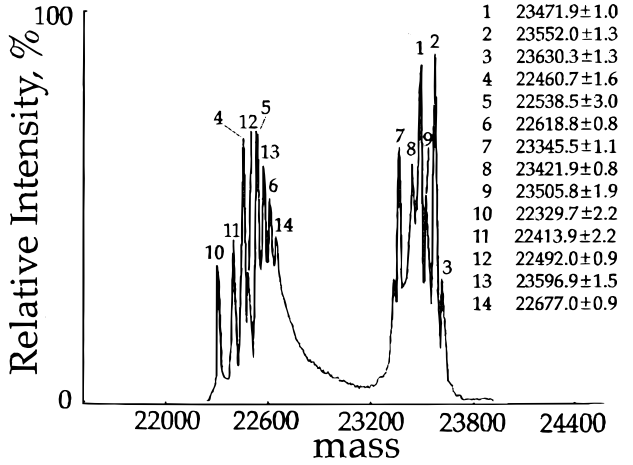

Figure 1. Electrospray mass spectrum of ovine α_{s1}-casein A variant transformed on a real mass scale. The measured molecular mass of each component is reported in Table 1. The multiple series of peaks correspond to differently phosphorylated components. The identification of each component was made by comparison of the measured and the theoretical mass, calculated with a computer program aid.

RESULTS

Characterization of Multiple Species at Different Chain Lengths and Degrees of Phosphorylation in Ovine α_{s1}-Casein. HPLC-purified ovine α_{s1}-casein variant A was analyzed by ES/MS, as shown in Figure 1. Fourteen components were detected, which would justify the band multiplicity previously detected by immunoblotting-two dimensional gel electrophoresis on individual ovine casein samples (Chianese et al., 1996b).

On the basis of the measured molecular mass and of the known protein sequence (Ferranti et al., 1995), these components were tentatively associated to differently sized forms of ovine α_{s1}-casein (Table 1). Each of these components occurred at three different degrees of phosphorylation, namely 9, 10, and 11 phosphate groups/molecule, in the relative ratio of 41/39/20, respectively.

Components 1−3 and 4−6 were soon identified as the two molecular species of ovine α_{s1}-casein, 199 and 191 residues long, respectively, previously characterized by Ferranti et al. (1995), the difference between them consisting in the deletion of peptide 141−148.

Components 7−9 and 10−12 were related to components 1−3 and 4−6 respectively by a −128 mass shift (corresponding to the mass of a Gln or a Lys residue), which suggested the occurrence of a deletion of one of these residues. Components 13−14 having molecular mass about 873 Da lower than the full-length protein might correspond to α_{s1}-casein having peptide 110−117 deleted.

In order to confirm the above assignments, further ES analysis was carried out following complete dephosphorylation of α_{s1}-casein, as shown in Figure 2. The spectrum was found to be simpler than that made on

Table 1. Measured and Predicted Molecular Masses of the Components Occurring in Ovine α_{s1}-Casein Variant A[a]

a_{s1}-casein component	measured mol mass (Da)	predicted mol mass (Da)	protein identification	P#[b]	%
1	23 471.9 ± 1.0	23 471.6	full length	9	19.3
2	23 552.0 ± 1.3	23 551.6	full length	10	19.0
3	23 630.3 ± 1.3	23 631.6	full length	11	8.5
4	22 460.7 ± 1.6	22 459.5	del (141−148)	9	7.1
5	22 538.5 ± 3.0	22 539.5	del (141−148)	10	6.2
6	22 618.8 ± 0.8	22 619.5	del (141−148)	11	4.3
7	23 345.5 ± 1.1	23 343.6	del Gln78	9	7.2
8	23 421.9 ± 0.8	23 423.6	del Gln78	10	7.5
9	23 505.8 ± 1.9	23 503.6	del Gln78	11	3.6
10	22 329.7 ± 2.2	22 331.5	del (141−148) + Gln78	9	2.9
11	22 413.9 ± 2.2	22 411.5	del (141−148) + Gln78	10	2.7
12	22 492.0 ± 0.9	22 491.5	del (141−148) + Gln78	11	1.5
13	22 596.9 ± 1.5	22 598.2	del (110−117)	9	7.4
14	22 677.0 ± 0.9	22 678.2	del (110−117)	10	2.8

[a] The α_{s1}-casein fractions were purified by reversed-phase HPLC and directly analyzed by ES/MS. The measured molecular mass of each component is reported as average mass in Da (mean ± SD). In the spectrum, five series of discretely phosphorylated components were present, differing in polypeptide chain length, which were associated with the corresponding proteins on the basis of the theoretical mass using a computer program. Species 1−3 correspond to the differently phosphorylated forms of the full-length protein 191 residues long; species 4−14 correspond to the deleted forms of α_{s1}-casein. [b] P#: phosphate groups/molecule.

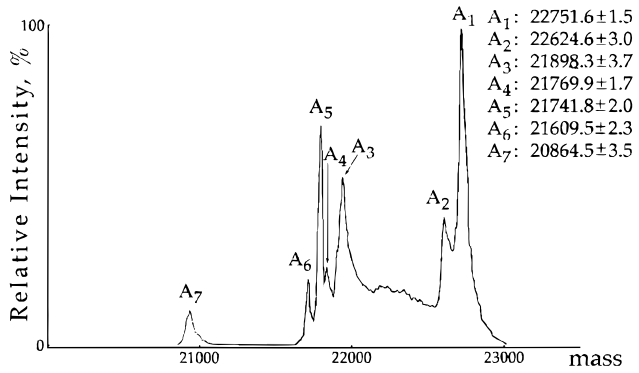

Figure 2. Electrospray mass spectrometry of dephosphorylated ovine α_{s1}-casein variant A. The spectrum contains seven components with their respective average molecular mass in daltons. The molecular masses of the components in the spectrum of native casein (Figure 1; Table 1) were shifted to lower mass values corresponding to the removal of the phosphates ($\Delta_m = -80$/phosphate).

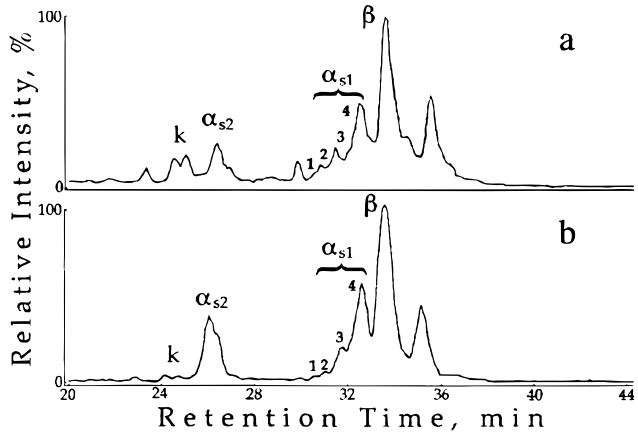

Figure 3. Liquid chromatography/electrospray mass spectrometry chromatogram (LC/ES/MS) of the dephosphorylated individual whole casein containing ovine α_{s1}-casein A (a) and D (b). For details see Materials and Methods.

the native protein, as the heterogeneity depending on the phosphorylation was fully removed by the treatment with alkaline phosphatase. The molecular mass of the dephosphorylated components obviously shifted to lower values as a −80 Da multiple was removed. By determining the difference between the molecular mass figures before and after dephosphorylation, the phosphorylation level of each native component was calculated. In the case of the 199 amino acid long α_{s1}-casein, the formerly phosphorylated species shifted from 23 471.9, 23 552.0, and 23 630.3 to the common mass value of 22 751.6, the molecular weight expected on the basis of the amino acid sequence of variant A. This finding confirmed the occurrence of three species at 9, 10, and 11 phosphate groups/molecule. By using this artifice, it was possible to determine that each of the multiple forms detected in ovine α_{s1}-casein variant A occurred at three differently phosphorylated levels. This heterogeneity was previously shown to correspond to partial phosphorylation at the level of Ser115 (50%) and Ser41 (20%), respectively (Ferranti et al., 1995). The only exception to this was the peptide chain with peptide 110−117 deleted, for which only two species at 10 and 9 phosphate groups/mol were observed (Table 1) because of the loss of Ser115.

Furthermore, two additional α_{s1}-casein components were detected in the ES spectrum, which were lacking in the native casein spectrum, probably because of their low abundance. The seven α_{s1}-casein-related species are indicated in Figure 2 with a capital letter and an increasing number corresponding to the decreasing molecular mass. Components A_1 and A_5 were soon identified by their molecular mass as the full-length protein and the protein deleted of peptide 141−148, respectively. The other five components obviously corresponded to novel deleted species.

On the basis of the relative intensity of the ES/MS multicharged ion, the dephosphorylated α_{s1}-casein A variant contained 45.1% A_1, 17.1% A_2, 10.0% A_3, 3.0% A_4, 16.9% A_5, 4.5% A_6, and 3.6% A_7, values consistent with those resulting from the measurement on the ES/MS of the native protein.

Primary Structure of the Multiple Forms of α_{s1}-Casein A. In order to fully characterize the seven components detected by analysis of α_{s1}-casein variant A, the dephosphorylated whole casein sample was fractionated by LC/ES/MS (Figure 3a), which allowed the simultaneous identification and purification on a semipreparative scale of the components themselves. In the region of the chromatogram where α_{s1}-casein was

Table 2. Characterization of Deleted Components Observed in the LC/ES/MS Analysis of Dephosphorylated α_{s1}-Casein Variant A, Obtained by FAB/MS and Automated Edman Degradation of the Tryptic Peptides[a]

HPLC peak	component name[b]	measured mol wt	theor mol wt	variated peptides[c]	MH^+ [d]	automated sequence analysis	deleted sequence(s)[e]
1	A_7	20 864.5 ± 3.5	20 866.3	T13/14	1586	YNVPQLEEQLHSMK	(110−117) and (141−148)
				T16*	1316	QPMIAVNQLFR	
2	A_6	21 609.5 ± 2.3	21 611.3	T16*	1316	QPMIAVNQLFR	(141−148) and Gln 78
				T18*	1678	AGSSSSSEEIVPNSAEK	
	A_5	21 741.8 ± 2.0	21 739.4	T16*	1316	QPMIAVNQLFR	(141−148)
3	A_4	21 769.9 ± 1.7	21 769.3	T13/14	1586	YNVPQLEEQLHSMK	(110−117) and Gln 78
				T8*	1678	AGSSSSSEEIVPNSAEK	
	A_3	21 898.3 ± 3.7	21 897.4	T13/14	1586	YNVPQLEEQLHSMK	(110−117)
4	A_2	22 624.6 ± 3.0	22 623.1	T8*	1678	AGSSSSSEEIVPNSAEK	Gln 78
	A_1	22 751.6 ± 1.5	22 751.2	T8	1806	AGSSSSSEEIVPNSAEQK	none
				T16	2327	QPMIAVNQELAYFYPQLFR	
				T13	1299	YNVPQLEIVPK	
				T14	1159	SAEEQLHSMK	

[a] Dephosphorylated ovine α_{s1}-casein variant A was analyzed by LC/ES/MS (Figure 3a) and the molecular weight of the components under each HPLC peak was measured. The purified components were hydrolyzed with trypsin. The digests were fractionated on a Vydac C18 column (Figure 4) and individual peptides analyzed by FAB/MS and automated Edman degradation. Peptide assignment was made taking into account the amino acid sequence of ovine α_{s1}-casein variant A (Ferranti et al., 1995). [b] The name refers to the peaks of the chromatogram shown in Figure 3a. [c] Only peptides containing the amino acid deletions are indicated. [d] MH^+ quasi molecular ion value observed in the FAB mass spectrum. [e] Numbers indicate the amino acid residues at the extremities of each peptide.

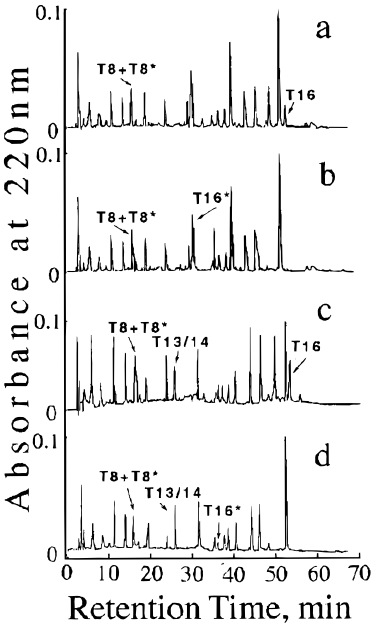

Figure 4. Reversed phase HPLC separation of the tryptic digest of ovine α_{s1}-casein A. The α_{s1}-casein isoforms contained in single peaks from 1 to 4 (Figure 3) were hydrolyzed with trypsin and the digests (0.5 mg) fractionated on a Vydac C_{18} column (250 × 4.6 mm, 5 mm). T, tryptic peptides. See Table 2 for peptide identification. Parts a−d refer to peaks 4, 3, 2, and 1 in Figure 3, respectively.

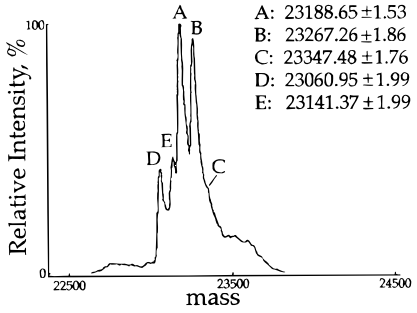

Figure 5. Partial transformed electrospray mass spectrum of ovine α_{s1}-casein D. The two series of differently phosphorylated α_{s1}-casein D were as follows: full-length component, A (5P), B (6P), and C (7P); and Gln78 deleted-short form, D (5P) and E (6P).

eluted, four peaks, which contained the seven components, were collected and labeled 1−4. ES/MS analysis (Table 2) actually showed that peak 1 contained component A_7; peak 2 components A_5 and A_6; peak 3 components A_3 and A_4; peak 4 components A_1 and A_2.

The protein components appearing under each of the four HPLC peaks were recovered and digested with trypsin. The resulting peptide mixtures were fractionated by HPLC (Figure 4) and the peptides either analyzed by FAB/MS, which allowed immediate identification of the peptides containing the amino acid deletions (labeled "T"), or sequenced by automated Edman degradation. The results are summarized in Table 2. It was found that components A_1 and A_2, both contained under the same HPLC peak, actually corresponded to the full-length protein (199 residues long) and to the protein with Gln78 (198 residues long) deleted, respectively. The other species corresponded to α_{s1}-casein having peptide 110−117 (A_3), peptide 110−117 plus Gln78 (A_4), peptide 141−148 (A_5), peptide 141−148 plus Gln78 (A_6), and peptide 110−117 plus peptide 141−148 (A_7) deleted.

Multiple Forms of α_{s1}-Casein D. The partial ES spectrum of native α_{s1}-casein D variant is shown in Figure 5. The presence of the full-length protein (expected molecular mass 23 188.3, 23 268.2, and 23 348.2 for the protein with five, six, and seven phosphate groups, respectively) together with the protein species with Gln78 deleted (expected molecular mass 23 060.2 and 23 144.1 for the protein with five and six phosphate groups, respectively) is evident by comparison of the measured molecular masses with those calculated on the basis of the sequence reported by Ferranti et al. (1995).

The ES spectrum of the dephosphorylated α_{s1}-casein D variant (Figure 6) was similar to that of the A variant, except that the mass of the components (D_1−D_7) shifted to values accounting for the amino acid substitutions already reported: Ser13(A) → Pro(D), Ser*(P)*12(A) → Ser(C), Ser68(A) → Asn(D), Ser*(P)*64(A) → Ser(D) and Ser*(P)*66(A) → Ser(D) (Ferranti et al., 1995).

Analysis by LC/ES/MS of the dephosphorylated casein sample (Figure 3b) allowed us to separate the protein species and collect each at the semipreparative scale, allowing determination of the primary structure. Analy-

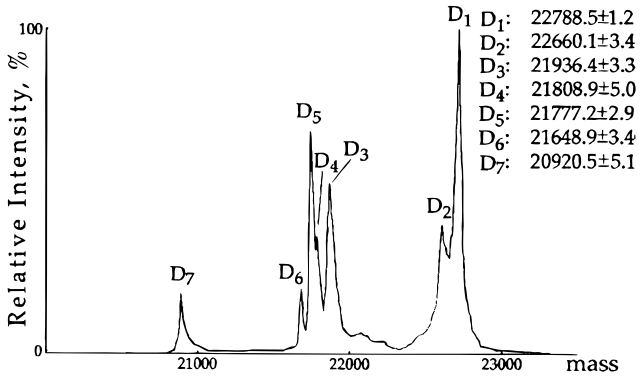

Figure 6. ES/MS analysis of the purified and dephosphorylated α_{s1}-casein. Seven components (C1−C7) were present, which were associated with the corresponding proteins on the basis of the theoretical mass, with the aid of a computer program. Other information as in Figure 2.

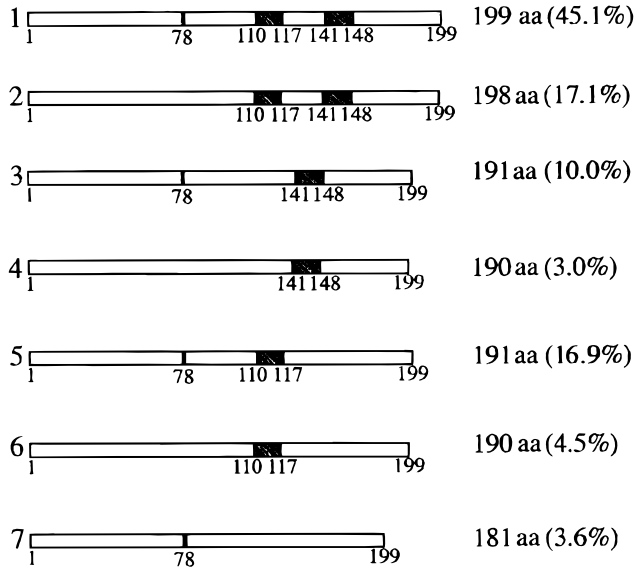

Figure 7. Diagram showing the multiple forms identified in ovine α_{s1}-casein A and D. The main component, corresponding to the full-length α_{s1}-casein, contains Gln_{78} (codified by the first triplet of exon 11), peptide 141−148 (codified by exon 16), and peptide 110−117 (codified by exon 13), indicated with dark boxes. The short protein forms differ from the complete protein for the deletion of one or more of the above elements. The species numbering corresponds to the forms identified in the spectra of Figure 2 for variant A and Figure 6 for variant D. The relative abundance of the single components (in parentheses) was determined by integrating the intensity of the multiple charged ions relative to each species.

sis of the tryptic peptides of the D variant produced results similar to those obtained for the A variant.

These results demonstrate that at least seven protein forms of α_{s1}-casein occurred in ovine α_{s1}-casein, each 199, 198, 191 (two forms), 190 (two forms), and 183 residues long, respectively. The inventory of the seven forms detected in the α_{s1}-casein variants A and D together with the amount of each form are shown in Figure 7. It is noteworthy that the proportion of full-length α_{s1}-casein to relative short forms is about half and half. This result warrants further study on the function of the short forms in the micelle assembly.

DISCUSSION

The present study has definitively ascertained that the electrophoretic heterogeneity of the ovine α_{s1}-casein variants A and D observed by Chianese et al. (1996a) actually originated from the copresence of a long main protein species and six shorter forms, each having at least two or three levels of phosphorylation. Thus, the number of short forms was higher than the lone form having peptide 141−148 deleted as previously detected by Ferranti et al. (1995).

The short forms differed from the complete protein for the deletion of the peptide segments 141−148 and 110−117 of Gln78 or for a combination of such deletions. While the Gln78 deletion has no effect on the electrophoretic mobility, those of peptides 110−117, containing the charged residues Ser*(P)*115, Glu110, Glu117 and Lys114, and of peptide 141−148, containing Glu141, produce species with a different net charge. Only the protein lacking peptide 141−148 has already been localized along the gel isoelectric patterns of ovine α_{s1}-casein (Chianese et al., 1996a), whereas the other short forms still need to be localized. Taking into account that four out of seven α_{s1}-casein protein forms are each distinguishable from the other according to the negative charge and that two out of four were at two degrees of phosphorylation and two other at three, one can confidently associate the electrophoretic heterogeneity previously shown by Chianese et al. (1996b) with the high complexity of the ovine α_{s1}-casein.

The presence of the deleted α_{s1}-casein species has been shown either in a variant associated with a high level of protein synthesis (variant A), as well as in variant D associated with a low level of synthesis. This finding suggests that it constitutes a common occurrence in ovine species, not related to allelic variability. More interestingly, the deleted species, especially those lacking Gln78 and peptide 141−148, represent a considerable percentage, nearly 50%, of the α_{s1}-casein fraction. This opens the discussion on both their origin and possible function in the overall casein composition.

Internal deletions of 11 and 37 residues have been described for caprine α_{s1}-caseins D and F (Brignon et al., 1990) variants, respectively. These shortened proteins were not the result of genomic deletions but rather of mutations inducing an incorrect splicing with an exon loss during pre-mRNA processing, which in turn produces multiple forms of α_{s1}-casein mRNA (Leroux et al., 1992).

More recently, the occurrence of three short counterparts as a consequence of abnormal mRNA splicing, together with the full length species has been shown also for the spread goat α_{s1}-$casein_{s1}$ A, B, and C variants at high level of synthesis (Ferranti et al., 1998). Analysis of caprine α_{s1}-casein mRNA from milk somatic cells demonstrated that these forms originated from skipping events at the level of exon 13 (codifying for peptide 110−117) and 16 (codifying for peptide 141−148) and from the presence of a cryptic splice site within exon 11 (whose first CAG triplet encodes Gln78) during primary transcript processing. A fifth short transcript was also found, whose protein counterpart was not detected. It is interesting to note that these short goat α_{s1}-casein species correspond perfectly to some of the proteins here identified in sheep.

Therefore, the discovery in homozygous ovine samples of seven α_{s1}-casein forms differing only in the polypeptide chain length suggests that they also could have originated from multiple mRNA splicing. Furthermore, it seems very unlikely that seven multiple forms can arise in homozygous individuals from an alternative

source, such as genetic polymorphism. This hypothesis is also supported by a recent study showing that the absence of exon 16, encoding peptide 141−148, from ovine α_{s1}-casein mRNA is not due to genomic deletion but to exon skipping, and the simultaneous occurrence of mRNA species containing the exon 16 sequence as well as species in which it had been spliced out (Passey et al., 1996).

More importantly, this event does not seem to be related to allelic variants to which a low level of protein synthesis is associated such as caprine α_{s1}-casein F, but it is possibly constitutive of both ovine and caprine α_{s1}-casein species. These results suggest that there could be a stabilized mechanism of simultaneous synthesis of α_{s1}-casein chains of different length common to sheep and goat. A genomic analysis on ovine samples is in progress to verify this hypothesis.

Furthermore, multiple forms are not restricted to ovine and caprine α_{s1}-casein, since ovine α_{s2}-casein was found to occur in two nonallelic forms produced by different mRNAs (Boisnard et al., 1991). Therefore, the exon skipping is a common event in the case of α_{s1}-casein and α_{s2}-casein genes, as the coding frame of these mosaic genes is divided into many short exons. Maturation of long cognate primary transcripts, such as the one encoding ovine and caprine α_{s1}-casein scattered by a score of introns (Leroux et al., 1992; Mercier and Vilotte, 1993), appears to be an intricate process obviously requiring many successive steps. Known deletions of α_{s1}-casein and α_{s2}-casein simply reflect the lack of accuracy of the processing mechanism, whenever mutations induce conformational modifications of pre-mRNA, preventing the normal progress of events.

We are currently investigating the occurrence of similar forms for the other casein components, whose presence could possibly affect the properties of differently-sized milk micelles. A deeper knowledge of both the structure and function of milk protein genes, which has already allowed the use of powerful techniques for the rapid identification of alleles, would offer the potential for a better understanding of the complexity of the milk protein composition.

ABBREVIATIONS USED

ES/MS; electrospray mass spectrometry; FAB/MS; fast atom bombardment mass spectrometry; liquid chromatography/electrospray mass spectrometry, LC/ES/MS; Pth-Xaa, phenylthiohydantoinyl amino acid.

LITERATURE CITED

Addeo, F.; Mauriello, R.; Moio, L.; Laezza, P.; Di Luccia, A. Ovine casein variant identification using electrophoretic, immunochemical and chromatographic techniques. *Milchwissenshaft* **1992**, *47*, 283−287

Barone, F. Istituto Sperimentale Lattiero Caseario at Lodi, Italy, unpublished results.

Boisnard, M.; Hue, D.; Bouniol, C.; Mercier, J-C.; Gaye, P. Multiple mRNA species code for two non-allelic forms of ovine α_{s2}-casein. *Eur. J. Biochem.* **1991**, *201*, 633−641.

Brignon, B.; Mahé, M.-F.; Ribadeau-Dumas, B.; Mercier, J.-C.; Grosclaude, F. Two of the three genetic variants of goat α_{s1}-casein which are synthesized at a reduced level have an internal deletion possibly due to altered RNA splicing. *Eur. J. Biochem.* **1990**, *193*, 237−241.

Chianese, L.; Garro, G.; Mauriello, R.; Laezza, P.; Ferranti, P.; Addeo, F. Occurrence of five α_{s1}-casein variants in ovine milk. *J. Dairy Res.* **1996a**, *63*, 49−59.

Chianese, L.; D'auria, R.; Ferranti, P.; Garro, G.; Mauriello, R.: Rubino, R.; Addeo, F. Occurrence of novel α_{s1}-casein variants in Italian goat breeds. In *Production and Utilization of Ewe and Goat Milk*; International Dairy Federation: Brussels, Belgium, 1996b; pp 141−147.

Ferranti, P.; Malorni, A.; Marino, G.; Pucci, P.: Goodwin, G. H.; Manfioletti, G.; Giancotti, V. Mass spectrometric analysis of the HMGY protein from Lewis lung carcinoma: identification of phosphorylation sites. *J. Biol. Chem.* **1992**, *267*, 22486−22489.

Ferranti, P.; Malorni, A.; Nitti, G.; Laezza, P.; Pizzano, R.; Chianese, L.; Addeo, F. Primary structure of ovine α_{s1}-caseins: localization of phosphorylation sites and characterization of genetic variants A, C, and D. *J. Dairy Res.* **1995**, *62*, 281−296.

Ferranti, P.; Addeo, F.; Malorni, A.; Chianese, L.; Leroux, C.; Martin, P. Differential splicing of pre-messenger RNA produces multiple forms of mature caprine α_{s1}-casein. *Eur. J. Biochem.* **1998**, *243*, 1−7.

Jaubert, A.; Martin, P. Reverse-phase HPLC analysis of goat caseins. Identification of α_{s1} and α_{s2} genetic variants. *Lait* **1992**, *72*, 235−247.

Leroux, C.; Mazure, N.; Martin, P. Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat α_{s1}-casein transcripts. *J. Biol. Chem.* **1992**, *267*, 6147−6157.

Mercier, J.-C.; Vilotte, J.-L. Structure and function of milk protein genes. *J. Dairy Sci.* **1993**, *76*, 3079−3098.

Meyer, H. E.; Hoffmann-Posorske, E.; Korte, H.; Heilmeyer j.r., L. M. G. Sequence analysis of phosphoserine-containing peptides. Modification for picomolar sensitivity. *FEBS Lett.* **1986**, *204*, 61−66.

Passey, G.; Glenn, W.; Mackinley, A. Exon skipping in the ovine α_{s1}-casein gene. *Comp. Biochem. Physiol. B* **1996**, *114*(4), 389−394.

Pucci, P.; Ferranti, P.; Marino, G.; Malorni, A. Characterization of abnormal human hemoglobins by fast atom bombardment mass spectrometry. *Biomed. Environ. Mass Spectrom.* **1989**, *18*, 20−26.

Received for review July 16, 1997. Revised manuscript received October 24, 1997. Accepted October 24, 1997.[⊗]

JF970609S

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1997.