

Heterogeneity of Caprine Beta-Casein Elucidated by RP-HPLC/MS: Genetic Variants and Phosphorylations

Carole Neveu,^{1,3} Daniel Mollé,¹ Javier Moreno,² Patrice Martin,³ and Joëlle Léonil^{1,4}

Received September 11, 2002

Casein variants occurring in milks from goats homozygous at the α_{s1} -Cn locus were separated and identified by an RP-HPLC/ESI-MS method. Preferential haplotypes arose as well as some particularities in posttranslational modifications. A new variant of caprine β -Cn, named C, as well as the phosphorylations pattern of the protein were characterized by the combined use of peptide mass fingerprinting and sequencing by tandem mass spectrometry. The molecular mass of the new variant in its 6P form was measured as 23854 Da and it differs in a mono amino acid substitution, A₁₇₇ $\rightarrow$ V₁₇₇, from the variant A. The phosphorylation pattern of caprine β -Cn is homologous to bovine β -Cn concerning the 5P located on Ser_{15, 17, 18, 19, 35} but it presents a specific additional phosphorylation site on Thr₁₂ that is comparable to human β -Cn phosphorylation located on Thr₃.

KEY WORDS: β -Casein; polymorphism; phosphorylation; mass spectrometry; goat milk.

1. INTRODUCTION

In goat milk, caseins exhibit a great variability associating an important genetic polymorphism to multiple posttranslational modifications. The most extensive and quite unusual polymorphism is recorded at the α_{s1} -casein (α_{s1} -Cn)⁵ locus that associates both structural and quantitative variability at the protein level. To date, at least 18 alleles have been described and shown to be associated with 4 different levels of α_{s1} -Cn in milk designated as "strong," "medium," "low," and "null" (Chianese *et al.*, 1997; Grosclaude and Martin, 1997). Thus, a strong level is observed for alleles As,⁶ B_{1,2,3,4}, C and L, with an estimated amount of 7.2 g of α_{s1} -Cn per liter of milk, and for allele H with 8.4 g/L. These strong alleles only differ in amino acid substitution(s). Medium alleles E and I approximately contribute only 1.6 g of α_{s1} -Cn per liter of

milk but share the same primary protein structure with the strong variants B₄ and A respectively. Low alleles F and G contribute 0.6 g of α_{s1} -Cn per liter of milk and both present an internal deletion of 37 (corresponding to the outsplicing of exons 9, 10 and 11) and 13 (outsplicing of exon 4) amino acid residues, respectively. As far as null alleles are concerned, three different mutations have been identified and are currently under investigation (Leroux and Martin, manuscript in preparation). Polymorphism has also been observed in the three other main fractions of caprine milk caseins, α_{s2} -Cn, κ -Cn and β -Cn. Three variants of α_{s2} -Cn have been identified and characterized: α_{s2} -Cn A, B (Boulanger *et al.*, 1984) and C (Bouniol *et al.*, 1994; Chianese *et al.*, 1992) that only differ in amino acid substitution(s). The existence of a polymorphism of κ -Cn was first suggested by Russo

¹INRA, Laboratoire de Recherche de Technologie Laitière, 35042 Rennes cedex, France.

²Instituto de Fermentaciones Industriales (CSIC), 28006 Madrid, Spain.

³INRA, Laboratoire de Génétique Biochimique et de Cytogénétique, 78352 Jouy-en-Josas cedex, France.

⁴To whom correspondence should be addressed. E-mail: jleonil@rennes.inra.fr

⁵Abbreviations: Cn, Casein; α -lac, α -lactalbumin; β -lg, β -lactoglobulin; ESI-MS, electrospray ionization-mass spectrometry; TPCK, tosyl-phenylalanine-chloromethylketone; FPLC, fast protein liquid chromatography; MS/MS, tandem mass spectrometry; CID, collision induced dissociation; Mr, relative mass

⁶Studies on the polymorphism at the DNA level have revealed that the allele A can be divided into four subtypes on the basis of intronic mutations (Leroux *et al.*, manuscript in preparation).

et al. (1986) and later confirmed by several groups (e.g., Di Luccia *et al.*, 1990; Law and Tziboula, 1993; Recio *et al.*, 1997). Characterization of variants A, B and C was published recently (Yahyaoui *et al.*, 2001) confirming variant B differs from variant A by the Val₁₁₉ → Ile₁₁₉ substitution as previously revealed by Mercier *et al.* (1976) and showing variant C differs from variant A by 4 amino acid substitutions (Val₆₅ → Ile₆₅, Val₁₁₉ → Ile₁₁₉, Ala₁₅₆ → Val₁₅₆, Ser₁₅₉ → Pro₁₅₉).

Caprine β -Cn, that can represent up to 60 % of total casein in goat milk, has long been considered to be monomorphic. To date, four different alleles (A, B, O, O') have been described. The sequence of the protein coded by the allele A was first elucidated by Robert *et al.* (1992) from its cDNA precursor. By electrophoretic techniques, an additional allele, so named B, was described by Mahé and Grosclaude (1993) but DNA and protein sequences still remain undetermined. The existence of null alleles was first presumed by Dall'olio *et al.* (1989) and confirmed by Mahé and Grosclaude (1993) whose segregation analysis has concluded to the existence of two null alleles, O and O', later characterized by Persuy *et al.* (1996) and Rando *et al.* (1996) at the DNA level. Variability also arises from posttranslational processing of the protein. The caprine β -Cn exhibits two principal phosphorylation levels, 5 and 6P (Richardson and Creamer, 1974), generally occurring in similar relative abundance but other phosphorylated forms (3 and 4P) have also been described (Chianese *et al.*, 1993). Homology with bovine β -Cn would locate 4P on the cluster Ser₁₅, 17, 18, 19, and 1P on Ser₃₅ and by electrophoretic separations conducted on plasmin hydrolysates of β -Cn, Trujillo *et al.* (1997) have suggested that the sixth phosphorylation site occurring on caprine β -Cn would correspond to Thr₄₂. Nevertheless, no direct determination being available, ambiguity still remains on the exact location of the six potential phosphorylation sites of caprine β -Cn and complementary work is needed at the protein level.

Moreover, caseins polymorphism, amino acid sequences as well as posttranslational modifications, are known to be deeply involved in the functional and biological properties of milk (e.g., relation between caprine α_{s1} -Cn polymorphism and milk allergenicity; Bevilacqua *et al.*, 2001), and even slight modifications can prove to be of a great importance (e.g., relation between the consumption of bovine β -Cn A₁ and similar variants and the development of some diabetes; Hill and Boland, 2001). Therefore, there is a growing need for characterizing the protein variants (sequences and posttranslational modifications) with more accuracy.

The ability of on line reverse phase-high performance liquid chromatography/electrospray ionization mass spectrometry (RP-HPLC/ESI-MS) to separate and characterize protein variants was exploited here first to investigate the particularities of the caseins polymorphism occurring in milks "homozygous" for known α_{s1} -Cn alleles and in particular in milks α_{s1} -Cn OO for which no mass data have been available so far. The second part of the study was to better characterize caprine β -Cn and its phosphorylations pattern.

2. MATERIALS AND METHODS

2.1. Reagents

All chemicals used were of analytical reagent grade. Tosyl-phenylalanine-chloromethylketone treated trypsin (TPCK-trypsin, EC 3.4.21.4) and Glu-C endoproteinase (EC 3.4.21.19) were obtained from Sigma (Saint Quentin Fallavier, France).

2.2. Preparation of Milks

Raw milks from French Alpine or saanen goats homozygous at the α_{s1} -Cn locus for alleles A_b, E or O₁ were supplied by INRA Station d'Amélioration Génétique des Animaux (Castanet-Tolosan, France). After adding 0.02% sodium azide, samples were immediately cooled down to 4°C. Two types of samples were used: 10 individual milks for each genotype studied and two bulk milks corresponding to genotypes EE and O₁O₁, respectively. All milks were skimmed by centrifugation at 800 × g, 20 min, 25°C.

2.3. Preparation of Whole Casein and β -Cn

2.3.1. Whole Casein

Whole casein from bulk milks were prepared by microfiltration techniques according to the method described by Pierre *et al.* (1992) and were immediately lyophilized until used in experiments.

2.3.2. β -Cn

β -Casein was purified from each freeze-dried whole casein by fast protein liquid chromatography (FPLC) on a cation-exchanger column (S-Sepharose XK 50/30 column, 375 ml gel volume) fitted on a Pharmacia FPLC

system LCC-500 equipped with a GP 250 gradient programmer, a single-path 280 nm UV-1 detector, a two-channel REC-482 recorder and a peristaltic pump (Pharmacia LKB Biotechnology, Uppsala, Sweden). Purification procedure was carried out according to a modified method of Jaubert and Martin (1992). Before the chromatographic separation, 200 ml of whole casein dissolved in 20 mM Na formiate, pH 8 containing 6 M urea and 0.8 mM DTT (1 mg/ml) was reduced for 2 hr at 37°C. After pH 4 adjustment by adding 4 M formic acid and filtration on PVDF membrane (0.45 μ m, Millex-HV Syringe Filter Unit, Millipore, Bedford, MA, USA), the solution was loaded on the column previously equilibrated with 20 mM Na formiate, pH 4 containing 6 M urea and 0.064 mM DTT (buffer A). Separation was achieved at room temperature, at a 5 ml/min flow-rate by a linear gradient from 0% to 30% of buffer B (1M NaCl in buffer A) over 750 min. The collected fractions were adjusted to pH 7 with 2.5% (v/v) ammoniac, diafiltrated against distilled water on a 3 kDa Omega membrane (Minisette membrane cassettes, Filtron, Northborough, MA, USA) and finally lyophilized. The purity of β -Cn determined by urea-polyacrylamide gel electrophoresis (Andrews, 1983) was $\sim$ 98%.

2.4. Tryptic Digestion of β -Cn

Digestion of purified β -Cn by TPCK-trypsin was performed in 20 mM Tris-HCl, pH 7.5 buffer at an enzyme/substrate ratio (mol/mol) of 1:2000. The hydrolysis was carried out at 37°C overnight and the reaction was stopped by pH 2 adjustment by adding 10% (v/v) TFA. The hydrolysates were stored at -20°C until used.

2.5. Glu-C Endoproteinase Digestion of Peptide [1–28]

For subsequent MS/MS analysis, further digestion of a mix of β -Cn tryptic peptides [1–28] and [1–29] by Glu-C endoproteinase was performed in 10 mM ammonium acetate, pH 4 buffer at an enzyme/substrate ratio (mol/mol) of 1:50. The hydrolysis was carried out at 25°C overnight and resulting digest immediately frozen.

2.6. Reverse-Phase HPLC Separations of Milk Proteins

RP-HPLC was carried out on a Spectra Physic SP 8800 system (Spectra Physics, San José, CA) with a

Vydac C4 column 214 TP 5415 (4.5 mm i.d. $\times$ 150 mm, Touzart et Matignon, Vitry sur Seine, France) according to a modified method of Jaubert and Martin (1992). Separation was achieved using 0.106% TFA as eluent A and 80% acetonitrile/20% H_2O /0.1% TFA as eluent B. Separation of individual milks proteins was obtained with a linear gradient from 35% to 62% eluent B over 54 min.

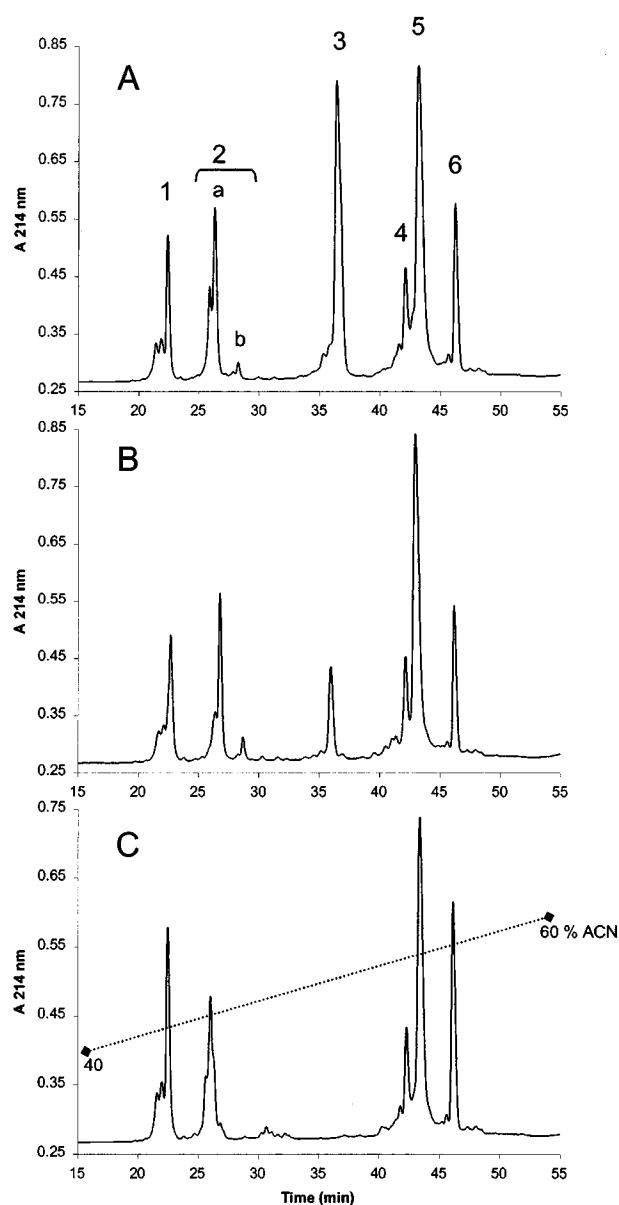


Fig. 1. Typical RP-HPLC fractionation of homozygous milk proteins as analyzed by on-line coupling with ESI-MS. α_{s1} -Cn genotypes: AA (A) – EE (B) – O_1O_1 (C). Protein fractions: κ -Cn (1) – α_{s2} -Cn (2a et 2b) – α_{s1} -Cn (3) – α -lac (4) – β -Cn (5) – β -lg (6).

The column was maintained at 40°C, flow rate was 1 ml/min and eluted peaks were detected by UV-absorbance at 214 nm.

Before analysis the 30 individual skim milks were reduced by a 1/3 dilution with a 0.1 M Tris/HCl pH 7 buffer containing 1.3% trisodic citrate (w/v), 8 M urea and 20 mM DTT. After standing at 37°C for 1 hr, samples were further diluted with eluent A and 10% (v/v) TFA to obtain a approximated 1 g/l protein solution at pH 2.2. After filtration on a PVDF membrane (0.45 µm, Millex-HV Syringe Filter Unit, Millipore, Bedford, MA, USA), 100 µl samples were then loaded on the chromatographic column equilibrated with eluent A.

2.7. RP-HPLC/ESI-MS of Milks and Hydrolysates

The MS analysis of milk proteins (for each genotype, one pool of the 10 corresponding individual milks), purified β -Cn and trypsin and Glu-C endoproteinase hydrolysates was performed by an on-line coupling between RP-HPLC and ESI-MS. RP-HPLC was carried out on a Hewlett Packard 1100 system (Agilent Technologies, Massy, France). The column was maintained at 40°C, flow rate was 0.25 ml/min and eluted peaks were detected by UV-absorbance at 214 and 280 nm and by total ion current. The column was directly interfaced

with a Sciex API III Plus mass spectrometer (Perkin-Elmer-Sciex, Thornhill, Ontario, Canada), a post flow splitter permitted to introduce only 1/10 of the HPLC eluate into the mass spectrometer. The ion source voltage and the orifice voltage were set at 4–5 kV and 70/90 V, respectively. Positive ion mode was used and mass scans were acquired over a m/z range of 500–2400 with a step size of 0.3 Da and a dwell time of 1 ms per step. The charge number of the multicharge ions, the deconvoluted mass spectra and the proteins M_r determination were obtained using the BioMultiView Software 1.3.1 (PE-SCIEX).

Measured masses of hydrolysate peptides were matched with predicted fragments by using SWISS-PROT database and the software Biotool box (Sciex).

RP-HPLC separations of milks proteins and purified β -Cn were achieved as described above but with a 2.1 mm i.d. $\times$ 150 mm column and with a different linear gradient from 0 to 80% eluent B over 40 min for purified β -Cn. Separation of hydrolysates was achieved on a Symmetry C18 column (2.1 mm i.d. $\times$ 150 mm, Waters, Saint Quentin en Yvelines, France) with a linear gradient from 3 to 70% eluent B over 40 min for tryptic β -Cn hydrolysates and from 0 to 50% eluent B over 22 min for Glu-C endoproteinase hydrolysate of β -Cn [1–28]/[1–29] mix.

Table 1. Average Experimental Masses of Casein Fractions (See Fig. 1) Obtained by Online RP-HPLC/ESI-MS of Homozygous Milks and Their Assignment by Comparison to Theoretical Data^a

RP-HPLC fraction	α_{s1} -Cn genotype	Main mass measured	Assigned compound	Theoretical mass	Phosphorylation levels detected
1	AA	19 305	κ -Cn B 2P	19 305.4	2 P
	EE	19 306	κ -Cn B 2P	19 305.4	2 P
	O1O1	19 292	κ -Cn A 2P	19 290.4	2 P
2a	AA	25 603	α s2-Cn A 11P	25 599.7	8-9-10-11 P
	EE	25 604	α s2-Cn A 11P	25 599.7	10-11 P
	O1O1	25 523	α s2-Cn A/B 10P	25 519.7 / 25 518.8	6-7-8-9-10 P
2b	AA	25 588	α s2-Cn C 11P	25 584.7	8-9-10-11 P
	EE	25 589	α s2-Cn C 11P	25 584.7	10-11 P
3	AA	23 447	α s1-Cn A 9P	23 443.4	7-8-9 P
	EE	23 375	α s1-Cn E 9P	23370.3	8-9-10 P
4	AA	14 197	α -Lac	14194.1	
	EE	14 197	α -Lac	14194.1	
	O1O1	14 197	α -Lac	14194.1	
5	AA	23 854	β -Cn X 6P?	?	3-4-5-6 P
	EE	23 826	β -Cn A 6P	23 822.0	5-6 P
	O1O1	23 854	β -Cn X 6P?	?	5-6 P
6	AA	18 189	β -Lg	18 191.3	
	EE	18 191	β -Lg	18 191.3	
	O1O1	18 191	β -Lg	18 191.3	

^a ? : unknown variant.

2.8. ESI MS/MS

The amino acid sequence of the peptides of interest detected during LC-ESI/MS was determined by tandem mass spectrometry analysis (MS/MS). Corresponding peaks were collected, dried under vacuum in a Speed Vac concentrator and diluted in 49:49:2 water/acetonitrile/formic acid (v/v/v) before infusion in the mass spectrometer at 5 μ l/min.

The parent ion selected through the first quadrupole (Q1) underwent collision-induced-dissociation (CID) in the second (Q2) by colliding with Argon molecules, the collision energy, chosen as a function of m/z of the parent ion, being in the range of 25 to 50 eV. The product ion mass spectra were acquired by scanning the third quadrupole (Q3) with a m/z ranging from 60 amu to the M_r of the parent ion + 50 amu. The CID mainly acting

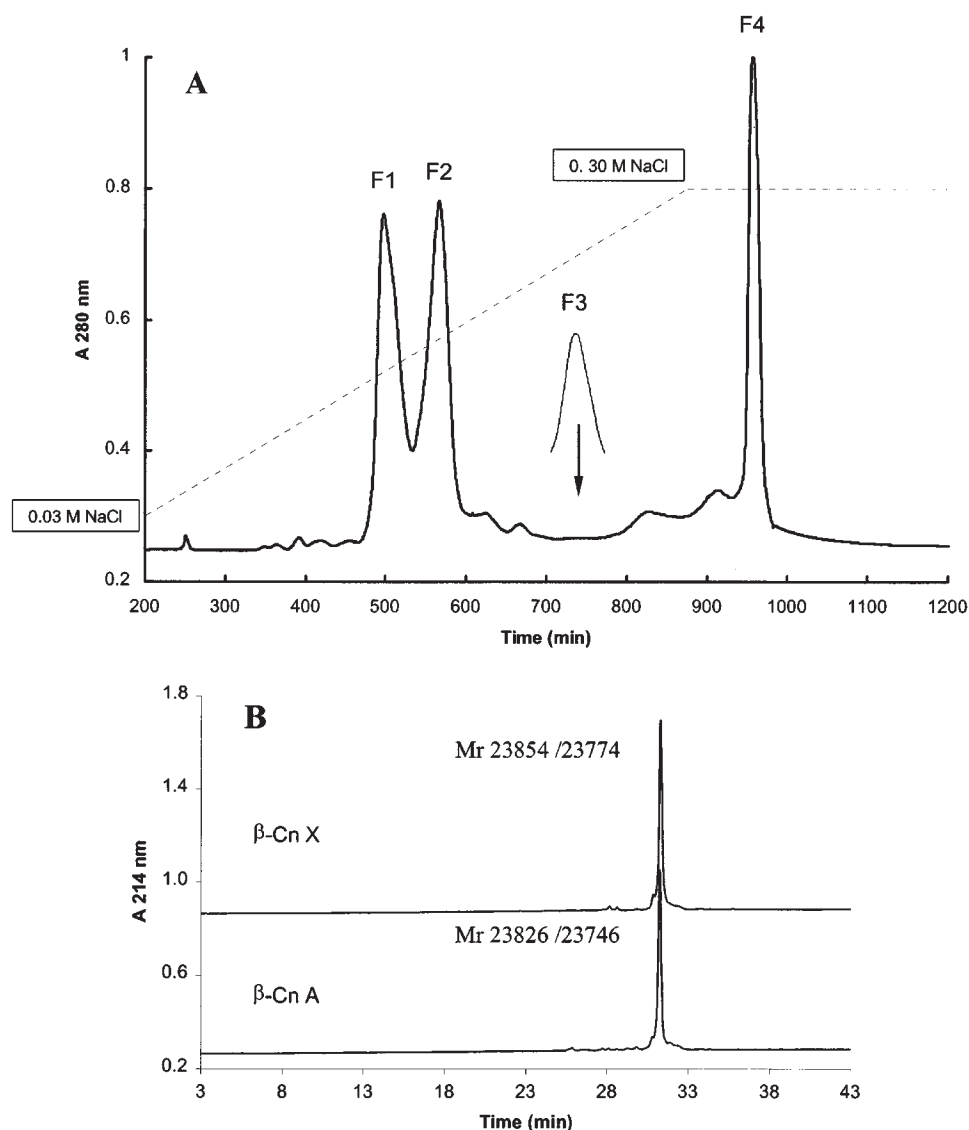


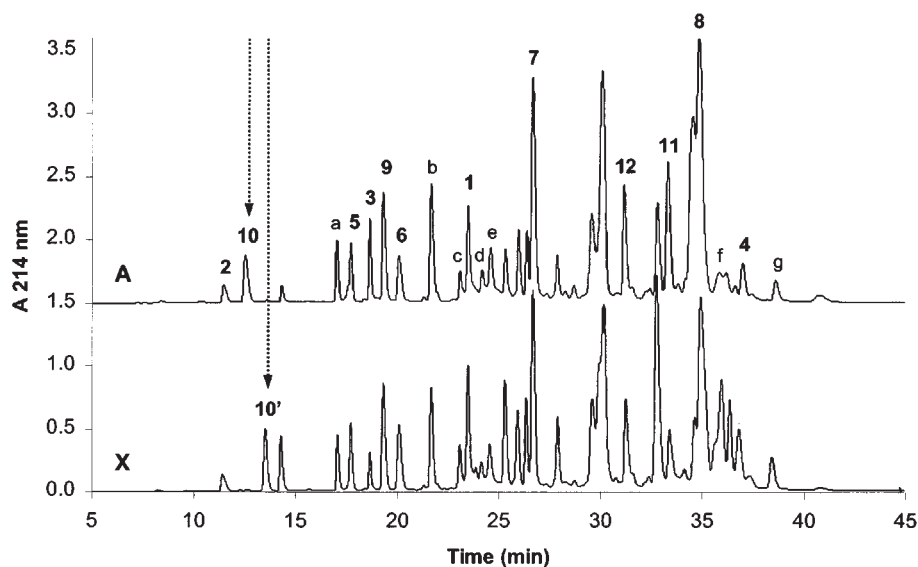
Fig. 2. A. Separation of O_1O_1 bulk milk whole caseins by chromatography on a S-Sepharose XK 50/30 column in 20 mM Na formate, pH 4 containing 6 M urea and 0.064 mM DTT. Elution was performed by a linear NaCl gradient (---). Flow rate was 5 ml/min. Location of supplementary fraction F3 (α_{s1} -Cn) occurring in EE bulk milk whole casein is indicated. B. RP-HPLC profiles of purified β -Cn A and X and corresponding measured M_r as analyzed by on-line coupling with ESI-MS.

on the peptide bonds, the fragmentation produces a series of charged ions corresponding to the C-terminal fragment ions, labeled y_n , and the complementary series of the corresponding N-terminal fragment ions, labeled b_n . The M_r increments between successive y_n and y_{n+2} or b_n and b_{n+1} ions indicate the M_r of the amino acid residue. Observed ions in the CID spectrum are matched with the theoretical M_r of y_n and b_n ions to reconstitute the amino acid sequence.

3. RESULTS AND DISCUSSION

3.1. RP-HPLC/ESI-MS Analysis of Milks from Goats Homozygous at the α_{s1} -Cn Locus

The fractions observed in the RP-HPLC profiles of the different milks (Fig. 1) could be identified as κ -Cn (fraction 1), α_{s2} -Cn (fraction 2), α_{s1} -Cn (fraction 3), α -lactalbumin (α -lac, fraction 4), β -Cn (fraction 5) and



Peak label	Tryptic fragment	M_r		
		Measured for β -Cn A	Measured for β -Cn X	Theoretical (β -Cn A)
1	1-28 5P	3544.4	3544.7	3544.27
2	29-32	516.4	516.4	516.33
3	33-48 1P	2062.3	2062.3	2061.48
4	49-97	5334.2	5334.6	5333.33
5	98-105	931.0	930.7	930.52
6	106-113	1012.9	1012.9	1012.52
7	114-132	2183.6	2183.8	2183.40
8	133-169	4157.9	4157.4	4156.97
9	170-176	779.8	779.8	779.49
10	177-181	569.5	597.4	569.33
11	182-200	2186.4	2186.6	2186.60
12	201-207	742.0	741.6	741.44
a	100-105	703.6	703.6	703.36
b	108-113	747.4	747.4	747.36
c	1-29 5P	3673.0	3673.0	3672.45
d	1-29 4P	3592.7	3592.7	3592.45
e	1-28 4P	3464.8	3464.8	3464.27
f	33-97 1P	7378.5	7378.6	7377.30
g	182-207	2910.5	2910.8	2910.52

Fig. 3. RP-HPLC profiles of tryptic digests of purified β -Cn A and X as analyzed by on-line coupling with ESI-MS. The tryptic peptides are numbered according to the splitting site of the theoretical sequence. Additional peptides are labeled by letters. The table reports the identification of tryptic peptides by comparison of their measured M_r with theoretical molecular mass and supposed posttranslational modifications of the variant A. Arrows indicate the differential peptide.

β -lactoglobulin (β -lg, fraction 6) as previously reported by Jaubert and Martin (1992) and Trujillo *et al.* (2000).

Exact determination of the protein variants arose from on line RP-HPLC/ESI-MS analysis of milks and comparison of the measured masses to the theoretical ones obtained from the amino acid sequence of the proteins taking into account the potential posttranslational modifications, particularly, phosphorylations. Experimental masses and ascribed protein variants are given in Table 1.

Contrary to published results of Trujillo *et al.* (2000), not only β -lg but also α -lac could be detected and masses calculated by on line RP-HPLC/ESI-MS; both were found monomorphic in the three types of milk.

Numerous masses were found for caseins revealing a large protein variability. The majority of the masses found are in agreement with those reported by Chianese *et al.* (1993) and Pierre *et al.* (2001) but here we put forward the polymorphism specificities in relation with the α_{s1} -Cn genotype especially the null one whose milk composition has not been well established yet and offers in particular a still uncharacterized β -Cn variant.

Milks AA and EE contain κ -Cn B (Ile₁₁₉) whereas milk O₁O₁ contains κ -Cn A (Val₁₁₉). In both cases, the major form of the protein was the 2P form.

α_{s2} -Cn phenotype determination is more complex as variants A and B only differ by an amino acid substitution Glu₆₄ $\rightarrow$ Lys₆₄ that induces a 1 Da mass difference and thus cannot be discriminated by the technique used here. A slight shift in the retention time of the 2 variants was noted by Jaubert and Martin (1992) but did not appear to be significant and reproducible in our experiments. Nevertheless, variant B is characterized by a single mutation affecting the determinant of the phosphorylation site Ser₆₂ and the protein so exhibits a maximum of 10 phosphorylations compared with the 11P major form of α_{s2} -Cn A. One can so expect that the 10P form occurring in milk O₁O₁ is the variant B whereas the 11P form occurring in milks AA and EE is the A one. In addition, phosphorylated forms detected differ between milks: 5 levels of phosphorylation were observed in milk O₁O₁ ranging from 5 to 10 P, at least 4 levels occur in milk AA with 8 to 11 P and only two forms were detected in milk EE with 10 and 11P respectively. Finally α_{s2} -Cn C was never detected in milk O₁O₁ whereas it regularly occurred in milks AA and EE and here again, phosphorylation levels differ between milks AA and EE and the differences observed are similar to those noted for the variant A.

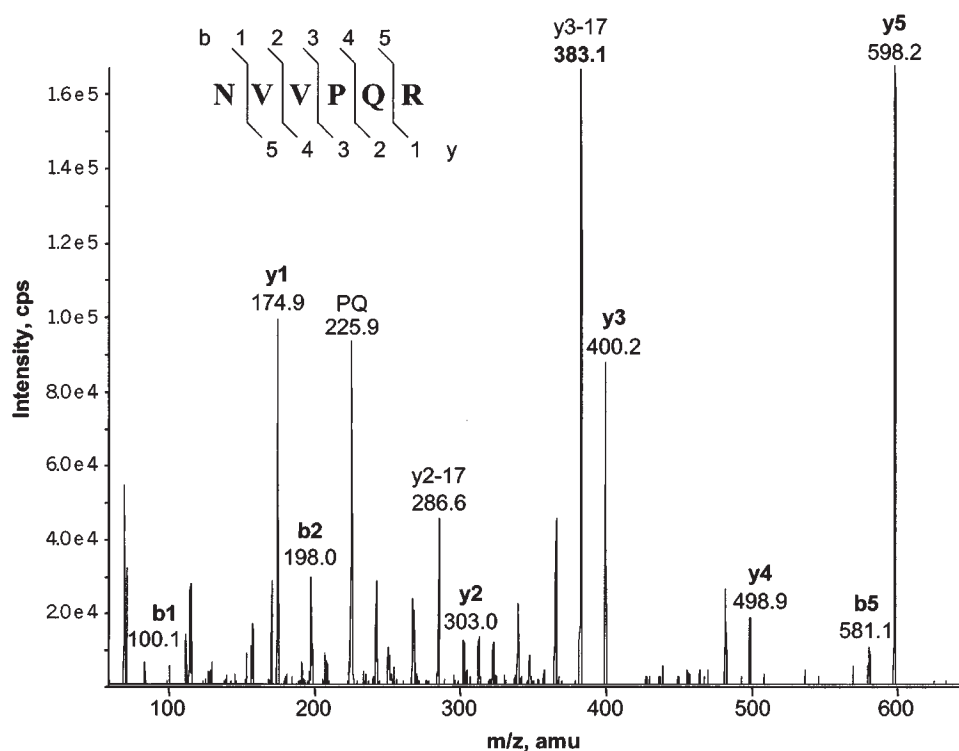


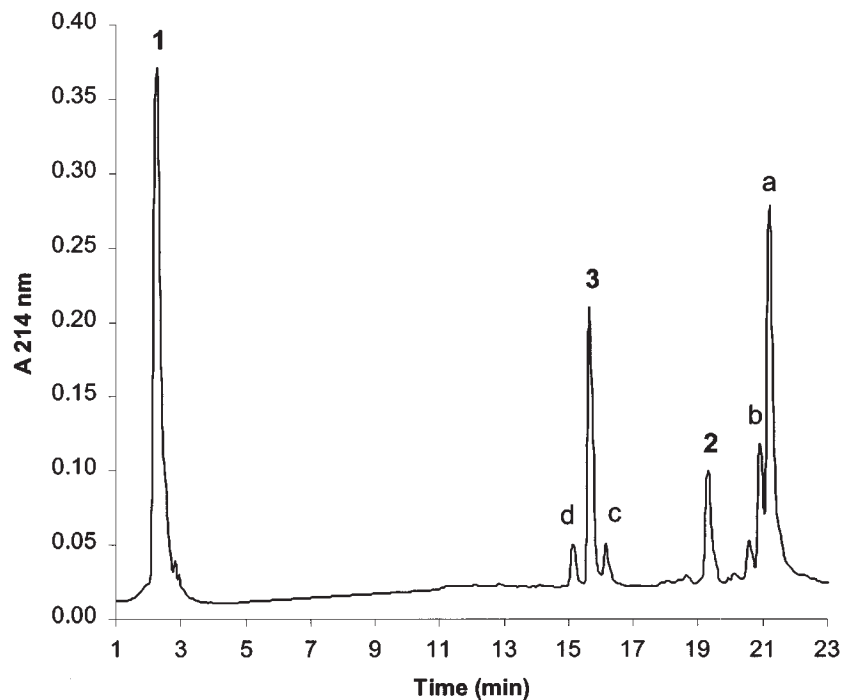
Fig. 4. Tandem mass spectra of the fragmentation process of the doubly charged ion arising from peptide 10' (M_r = 597.4 Da) obtained by ESI-MS. The corresponding amino acid sequence is indicated (insert).

The parallel use of genotype allowed us to discriminate between α_{s1} -Cn variants as direct identification is not always possible by using mass spectrometry. For example, variants A and B1 only differ by one mass unit, which is below the discriminating power of the method. Masses calculated for α_{s1} -Cn corresponded to the expected genetic variants, major forms being the 7, 8, and 9P α_{s1} -Cn A and the 8, 9 and 10P α_{s1} -Cn E.

Experimental masses found for β -Cn in milks EE could be ascribed to the 5 and 6P forms of the variant A occurring in similar proportions. The variant occurring in

milks AA and O₁O₁ remains unidentified (labeled X). A similar mass has previously been observed by Chianese *et al.* (1993) and Pierre *et al.* (2001) but the characterization of this variant has never been conducted. Considering the same major phosphorylation levels of 5 and 6P the mass difference between the 2 variants would be ~28 Da.

The preferential associations suggested above between casein variants result from the linkage disequilibrium existing in the alleles of the Cn loci as the four corresponding genes are located at chromosome 6 within a region of ~250 kb only. This stresses the importance to



Peak label	Glu-C endoproteinase fragment	Mr	
		Measured	Theoretical
1	1-5	689.4	689.30
2	6-21 5P	2079.0	2078.67
3	22-28	811.5	811.46
a	6-28 5P	2872.8	2872.13
b	6-29 5P	3001.0	3000.76
c	21-28	940.5	940.50
d	22-29	939.6	939.55

Fig. 5. RP-HPLC profiles of Glu-C endoproteinase digest of a mix of peptide [1–28] and [1–29] as analyzed by on-line coupling with ESI-MS. The peptides are numbered according to the splitting site of the theoretical sequence. Additional peptides are labeled by letters. The table reports the identification of Glu-C endoproteinase peptides by comparison of their measured M_r with theoretical molecular mass and supposed posttranslational modifications of the variant A.

consider caseins all together when studying the effects of Cn polymorphism on milk properties and should incite to implement such study.

The second part of our work was to took advantage of the milks EE and O_1O_1 to purify β -Cn A and X respectively to further characterize the variant X and to determine the phosphorylation pattern of the two variants by a direct MS investigation.

3.2. Characterization of β -Casein Variants

3.2.1. Purified β -Cn from Whole Casein EE and O_1O_1

β -Cn A and X were purified by ion exchange chromatography of whole caseins prepared from milk EE and milk O_1O_1 , respectively. Four major distinct fractions, F1, F2, F3, F4, were separated (Fig. 2) and considering their elution order (Jaubert and Martin, 1992), corre-

spond to the β , κ , α_{s1} and α_{s2} -Cn respectively. Identification was confirmed by the RP-HPLC/ESI-MS analysis of the different peaks (data not shown). The chromatographic profiles and mass data obtained for purified β -Cn are presented in Fig. 2.

3.2.2. Characterization of β -Cn Variant X

Characterization of β -Cn X was carried out by comparing β -Cn A and X peptide mass fingerprinting after complete trypsin digestion of both proteins. On-line coupling RP-HPLC/ESI-MS analysis of tryptic digests of β -Cn A and X allowed us to identify 19 tryptic peptides covering both whole protein (Fig. 3). The only difference observed between the two hydrolysates lied in peptide 10 with a longer elution time and higher measured mass (597.7 Da vs. 569.5 Da) in the β -Cn X tryptic hydrolysate. Moreover, the mass difference of 28 Da ex-

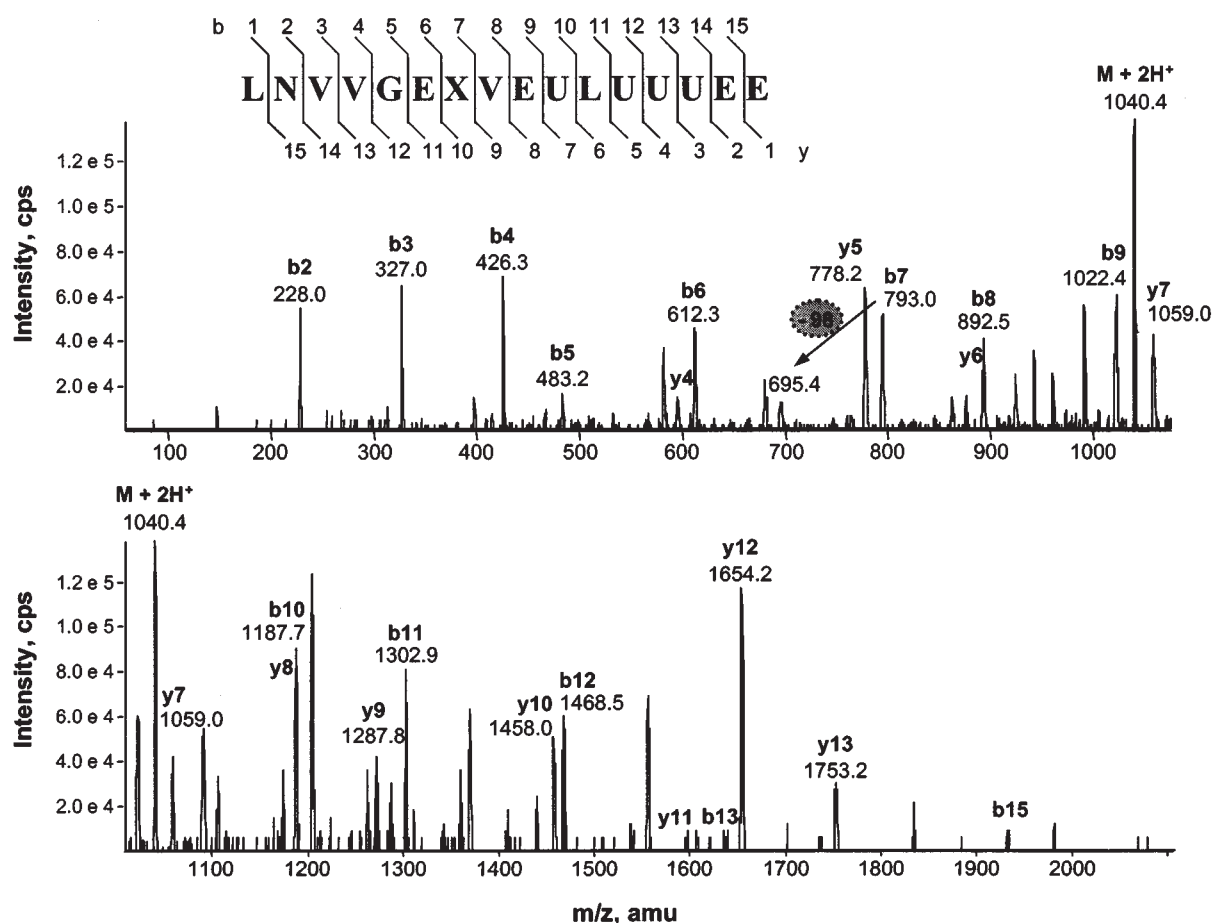


Fig. 6. Tandem mass spectra of the fragmentation process of the doubly charged ion ($M + 2H^+$, m/z 1040.4) arising from peptide [6–21] 5P (M_r = 2079.0 Da) obtained by ESI-MS. The corresponding amino acid sequence is indicated (insert). X stand for phosphorylated Thr and U for phosphorylated Ser.

actly fitted the mass difference found between whole proteins which backed us up in locating the amino acid substitution(s) in peptide 10.

Amino acid sequence of peptide 10 was obtained by ESI-MS/MS (Fig. 4). The following sequence was deduced from the y and b series of fragment ions so obtained: VVPQR. In variant A, corresponding sequence being AVPQR, the mutation characterizing the variant X is so $A_{177} \rightarrow V_{177}$. This substitution totally explains the mass difference reported between the 2 proteins. Recent work carried out at the DNA level (Wang *et al.*, 2001 direct submission) confirms this mutation and reveals it arises from a single base mutation in the coding triplet: A_{177} (GCA) (Robert *et al.*, 1992) to V_{177} (GTA). No relation can be established yet with the variant B of β -Cn described by Mahé and Grosclaude (1993) but still uncharacterized. The new variant of β -Cn characterized here will be named C.

3.2.3. Location of the Six Phosphorylation Sites of Caprine β -Cn

Direct location of the sixth phosphorylation sites occurring in caprine β -Cn was performed by mass spectrometry sequencing of specific β -Cn peptides.

RP-HPLC/ESI-MS analysis of tryptic hydrolysates of β -Cn A and C (Fig. 3) located 5P in peptide [1–28] and 1P in peptide [33–48]. Considering the phosphorylation codes stated by Mercier (1981), potential phosphorylation sites would be Ser_{15, 17, 18, 19, 35} and Thr_{12, 41}.

To locate the 5P occurring in peptide [1–28], complete hydrolysis by Glu-C endoproteinase was performed on a mix of tryptic peptides [1–28] and [1–29] and resulting digest further analyzed by on-line coupling RP-HPLC/ESI-MS. Figure 5 shows that 7 peptides covering the whole tryptic peptide were identified and corresponding masses indicates that 5P occur in peptide [6–21].

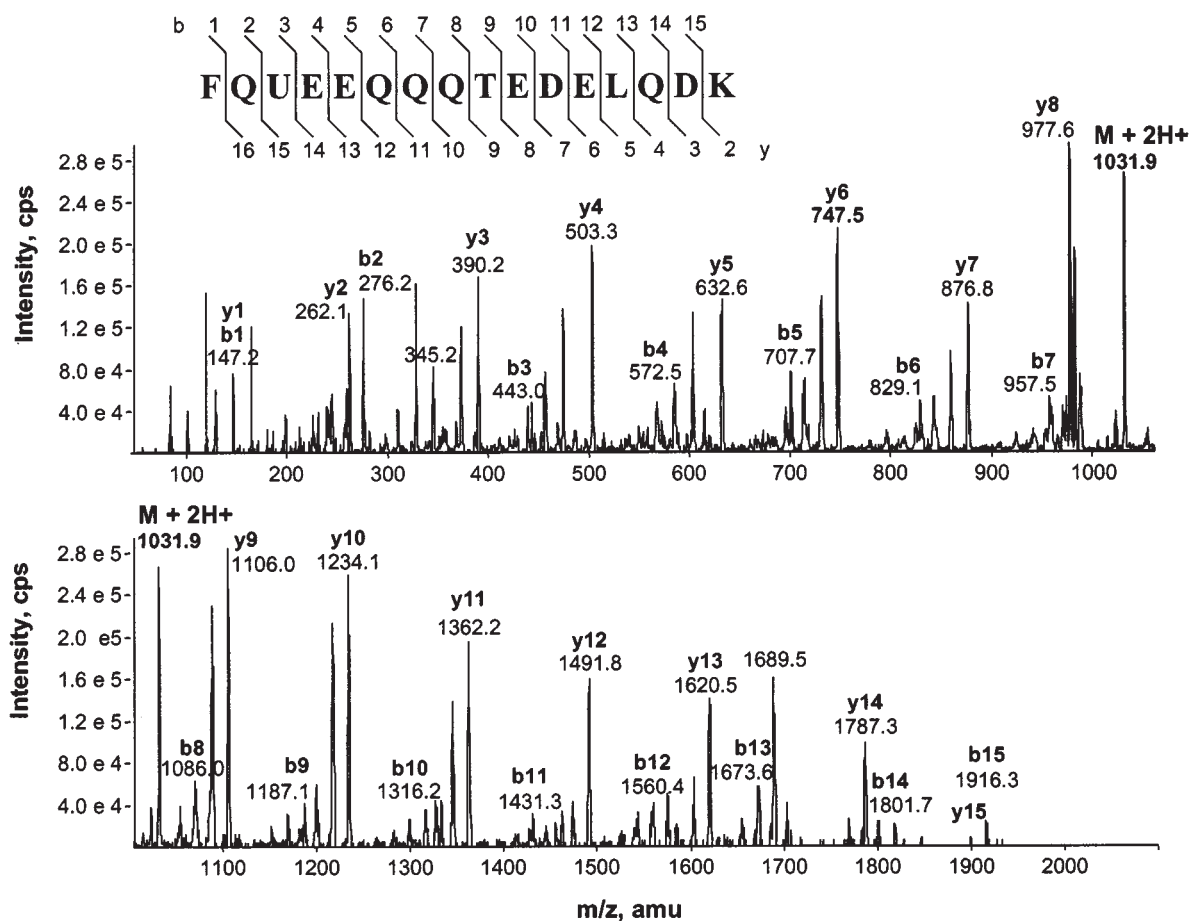


Fig. 7. Tandem mass spectra of the fragmentation process of the doubly charged ion ($M + 2H^+$, m/z 1031.9) arising from tryptic peptide [33–48] 1P ($M_r = 2062.3$ Da) obtained by ESI-MS. The corresponding amino acid sequence is indicated (insert). U stands for phosphorylated Ser.

Amino acid sequence of peptide [6–21] was obtained by ESI-MS/MS (Fig. 6). The sequence deduced from the y and b series of fragment ions locates the 5 phosphorylations of peptide [1–28] on Ser₁₅, 17, 18, 19 and Thr₁₂.

Homology with bovine β -Cn is so confirmed on the cluster of Ser-P and the specific additional phosphorylation of Thr₁₂ is comparable to the phosphorylation of Thr₃⁷ described for human β -casein (Greenberg *et al.*, 1984).

The last phosphorylation, detected in peptide [33–48], must correspond either to the primary site Ser₃₅ as described for bovine β -Cn or to Thr₄₁ as suggested by Trujillo *et al.* (1997). It is noteworthy that in this previous study, phosphorylation was observed on Ser₃₅ alone or on both Ser₃₅ and Thr₄₁, Thr₁₂ thus not appearing to be phosphorylated. Amino acid sequence of peptide [33–48] was obtained by ESI-MS/MS (Fig. 7) and the sequence deduced from the y and b series of fragment ions locates the sixth phosphorylation on Ser₃₅ as described for bovine β -Cn.

4. CONCLUSION

In this work, a new variant of β -Cn, named C, was described and is characterized by a single amino acid substitution on residue 177: A₁₇₇→V₁₇₇. Analyzing the phosphorylations pattern of caprine β -Cn, we confirmed the supposed homology with the bovine protein for Ser₁₅, 17, 18, 19, 35 and we located the additional phosphorylation on Thr₁₂.

Study of the casein polymorphism occurring in milks from goats homozygous for known α _{s1}-Cn alleles has put forward some preferential associations between protein variants but the exact determination of caprine casein haplotypes to improve the efficiency of breeding selection programs would require a complete genetic approach. Nevertheless, mass spectrometry offers an efficient and rapid complementary method giving access to the additional protein polymorphism triggered by posttranslational modifications that remain unreachable through genetics. These are known to deeply influence the functional and biological properties of milk and should be considered when studying casein polymorphism effects.

REFERENCES

- Andrews, A. T. (1983). *J. Dairy Res.* **50**: 45–55.
- Bevilacqua, C., Martin, P., Candalh, C., Fauquant, J., Piot, M., Roucayrol, A. M., *et al.* (2001). *J. Dairy Res.* **68**: 217–227.
- Boulanger, A., Grosclaude, F., and Mahé, M. F. (1984). *Génét. Sé. Evol.* **16**: 157–176.
- Bouniol, C., Brignon, G., Mahé, M. F., and Printz, C. (1994). *Anim. Genet.* **25**: 173–177.
- Chianese, L., Ferranti, P., Garro, G., Mauriello, R., and Addeo, F. (1997). In Proceedings IDF Seminar on Milk Protein Polymorphism, Palmerston North, New Zealand, pp. 259–267.
- Chianese, L., Garro, G., Nicolai, M. A., Mauriello, R., Ferranti, P., Pizzano, R., *et al.* (1993). *Lait* **73**: 533–547.
- Chianese, L., Mauriello, R., Intorcia, N., Moio, L., and Addeo, F. (1992). *J. Dairy Res.* **59**: 299–305.
- Dall'Olio, S., Davoli, R., and Russo, V. (1989). *Sci. Tec. Latt. Casearia* **40**: 24–28.
- Di Luccia, A., Mauriello, R., Chianese, L., Moio, L., and Addeo, F. (1990). *Sci. Tec. Latt. Casearia* **41**: 305–314.
- Greenberg, R., Groves, M. L., and Dower, H. J. (1984). *J. Biol. Chem.* **259**: 5132–5138.
- Grosclaude, F., and Martin, P. (1997). In Proceedings IDF Seminar on Milk Protein Polymorphism, Palmerston North, New Zealand, pp. 241–253.
- Hill, J., and Boland, M. (2001). *Austr. J. Dairy Technol.* **56**: 117–122.
- Jaubert, A., and Martin, P. (1992). *Lait* **72**, 235–247.
- Law, A. J. R., and Tziboula, A. (1993). *Milwissenschaft* **48**: 68–71.
- Leroux, C., and Martin, P. (1992). *Biochem. Biophys. Res. Commun.* **183**: 750–757.
- Mahé, M. F., and Grosclaude, F. (1993). *Génét. Sé. Evol.* **25**: 403–408.
- Mercier, J. C. (1981). *Biochimie* **63**: 1–17.
- Mercier, J. C., Chobert, J. M., and Addeo, F. (1976). *FEBS Lett.* **72**: 208–214.
- Persuy, M. A., Printz, C., Medrano, J. F., and Mercier, J. C. (1996). *Anim. Genet.* **27**: 87–100.
- Pierre, A., Fauquant, J., Le Graët, Y., Piot, M., and Maubois, J. L. (1992). *Lait* **72**: 461–474.
- Pierre, A., Molle, D., and Zahoute, L. (2001). *Lait* **81**: 667–678.
- Recio, I., Perez-Rodriguez, M., Amigo, L., and Ramos, M. (1997). *J. Dairy Res.* **64**: 515–523.
- Rando, A., Pappalardo, M., Capuano, M., Di Gregorio, P., and Ramunno, L. (1996). In Proceedings International Conference on Animal Genetics, Tours, France p50.
- Richardson, B. C., and Creamer, L. K. (1974). *Biochim. Biophys. Acta* **365**: 133–137.
- Robert, B., Ditullio, P., Vitale, J., Hehir, K., and Gerdon, K. (1992). *Gene* **121**: 255–262.
- Trujillo, A. J., Casals, I., and Guamis, B. (2000). *J. Dairy Sci.* **83**: 11–19.
- Trujillo, A. J., Guamis, B., and Carretero, C. (1997). *J. Dairy Sci.* **80**: 2258–2263.
- Wang, Q., Huang, Z., Chen, M. J., Huang, S. Z., and Zeng, Y. T. (2001). Available: <http://www.ncbi.nlm.nih.gov/>
- Yahyaoui, M. H., Coll, A., Sanchez, A., and Folch, J. P. (2001). *J. Dairy Res.* **68**: 209–216.

⁷Residue corresponding to the caprine Thr₁₂ because of the constitutive skipping of exon 3 occurring during the maturation of human β -Cn mRNA (Leroux and Martin, 1992).