Equine α_{S1} -casein: Characterization of alternative splicing isoforms and determination of phosphorylation levels

A. Matéos,* L. Miclo,* D. Mollé,† A. Dary,* J.-M. Girardet,*¹ and J.-L. Gaillard*

*Unité de Recherche Animal et Fonctionnalités des Produits Animaux (UR AFPA) – Equipe Protéolyse et Biofonctionnalités des Protéines et des Peptides (PB2P), Nancy-Université, Vandœuvre-lès-Nancy, France

†Science et Technologie du Lait et de l'Œuf (STLO), INRA UMR1253, Rennes, France

ABSTRACT

α_{S1} -Casein was isolated from Haflinger mare's milk by hydrophobic interaction chromatography and displayed great micro-heterogeneity by 2-dimensional electrophoresis, probably because of a variable degree of phosphorylation and alternative splicing events. The aim of the present work was to investigate the complexity of the mare's α_{S1} -casein. The different isoforms present in milk were submitted to a double treatment of dephosphorylation, first by using alkaline phosphatase and then acid phosphatase to achieve complete dephosphorylation. The apoforms were then analyzed by electrospray ionization mass spectrometry. The results revealed the existence of a full-length protein and 7 variants resulting from posttranscriptional modifications; that is, exon skipping involving exon 7, exon 14, or both and use of a cryptic splice site encoding a glutamine residue. The determination of the different phosphorylation degrees of the native isoforms of α_{S1} -casein was finally achieved by electrospray ionization mass spectrometry analysis after fractionation of the isoforms by ion-exchange chromatography. Thus, 36 different variants of equine α_{S1} -casein were identified with several phosphate groups ranging from 2 to 6 or 8 depending on whether exon 7 was skipped.

Key words: alternative splicing, cryptic splice site, equine α_{S1} -casein, phosphorylation

INTRODUCTION

Bovine caseins are a family of milk proteins (25 g/kg; Ng-Kwai-Hang, 2003). The α_{S1} -CN, α_{S2} -CN, and β -CN components are multi-phosphorylated, particularly on a conserved cluster sequence of 3 phosphoserine residues (**SerP**) followed by 2 glutamic acid residues: –SerP–SerP–SerP–Glu–Glu–. Unlike bovine milk, human and equine milks are considered to have poor CN content

(4 and 11 g/kg, respectively; Malacarne et al., 2002). The human and equine β -CN have variable degrees of phosphorylation; that is, from zero to 5 phosphate groups (**P**) for the human protein (Greenberg et al., 1984) and 3P to 7P for the equine protein (Girardet et al., 2006), whereas the bovine β -CN contains a mean of 4.9 phosphate groups per protein molecule on its 5 potential phosphorylation sites (Bandura et al., 2002).

Complete amino acid sequences of equine κ -CN (Iametti et al., 2001) and of Haflinger mare's β -CN isoform 1 (Swiss-Prot accession number Q9GKK3; Girardet et al., 2006) and isoform 3 [also called low-molecular-mass (M_r) β -CN; Miclo et al., 2007] have been determined, as well as the cDNA sequences of Hannoverian mare's β -CN isoform 2 (GenBank accession number AAG43954; Lenasi et al., 2003) and κ -CN (GenBank accession number AAK83669; Lenasi et al., 2003). Isoform 1 of the equine β -CN corresponds to the protein produced from a full-length mRNA. Isoform 2 is produced from an mRNA lacking exon 5 that results in the absence of the 27–34 region (β -CN $^{\Delta 5}$). Finally, isoform 3 has a deletion of the 50–181 region encoded by the main part of exon 7. Exon 5 is suppressed by alternative splicing and the removal of 75% of exon 7 used a cryptic splice site (Miranda et al., 2004; Lenasi et al., 2006; Miclo et al., 2007). The presence of α_{S2} -CN in equine milk was suspected by Ochirkhuyag et al. (2000) and Egito et al. (2001, 2002) and has been confirmed by microsequencing of peptides from trypsin-treated CN components extracted from electrophoretic gel (Miranda et al., 2004). Nevertheless, no nucleotide or protein whole sequence is yet available.

The mRNA sequences corresponding to Hannoverian mare's α_{S1} -CN $^{\Delta 7}$ (numbering of α_{S1} -CN exons of all species studied according to Miranda et al., 2004; GenPept AAK83668; GenBank AY040862; Lenasi et al., 2003) and to mare's α_{S1} -CN $^{\Delta 14}$ (GenPept AAL05435; GenBank AY049939; Milenkovic et al., 2002) have been isolated from lactating mammary gland. The presence of the α_{S1} -CN $^{\Delta 7}$ and α_{S1} -CN $^{\Delta 7,14}$ isoforms in Welsh pony's milk is strongly suspected by a mass spectrometry approach, but the mass determination precision of 0.2%

Received February 13, 2009.

Accepted April 17, 2009.

¹Corresponding author: jean-michel.girardet@scbiol.uhp-nancy.fr

was not sufficient to conclude (Miranda et al., 2004). Exon 5 of α_{S1} -CN is constitutively spliced (Lenasi et al., 2003). To our knowledge, complete equine α_{S1} -CN has not been isolated and characterized by proteomic methods. According to the deduced primary structure, the full-length equine α_{S1} -CN displays 8 potential sites of phosphorylation. The real number of phosphate groups per protein molecule is, however, unknown. In the case of the bovine species, α_{S1} -CN contains 8 or 9 phosphate groups per protein molecule, the most phosphorylated isoform being named α_{S0} -CN (Farrell et al., 2004). The equine α_{S1} -CN pattern obtained by urea-PAGE is much more complex than that of the equine β -CN with more bands (at least 7 visible bands for α_{S1} -CN against 5 bands for β -CN; Egito et al., 2002). As for equine β -CN (Girardet et al., 2006; Miclo et al., 2007), the complex pattern of α_{S1} -CN was probably caused by a variable degree of phosphorylation and alternative splicing of exons.

The aim of this study was to elucidate the micro-heterogeneity of α_{S1} -CN of Haflinger mare's milk by analyzing the different isoforms resulting from alternative splicing and displaying different degrees of phosphorylation. For this purpose, α_{S1} -CN was purified from whole sodium caseinate and fractionated by chromatographic methods. The different variants contained in the chromatographic fractions were identified by liquid chromatography electrospray ionization mass spectrometry (LC-ESI/MS) before and after enzymatic dephosphorylation.

MATERIALS AND METHODS

Preparation of Equine Sodium Caseinate

Haflinger mare's milk was obtained from Jum'Vital (Volmunster, France). Milk was stored at -20°C until used. After thawing, 0.05% (vol/vol) toluene was added to the milk to avoid microorganism development. Milk was skimmed by centrifugation at $4,000 \times g$ at 32°C for 30 min and toluene was added again. Sodium caseinate was prepared by isoelectric precipitation at pH 4.2 as previously reported (Egito et al., 2002).

Hydrophobic Interaction Chromatography

Sodium caseinate was fractionated by hydrophobic interaction chromatography (HIC) using a TSK Phenyl-5PW column (21.5×150 mm; $13\text{-}\mu\text{m}$ particle size; Interchim, Montluçon, France) connected to a fast protein liquid chromatography (FPLC) system (ÄKTA-FPLC, Amersham Pharmacia, Uppsala, Sweden) according to the method used by Tauzin et al. (2003) to fractionate bovine caseins. The column was

equilibrated at 25°C in 480 mM sodium phosphate buffer, pH 6.4, containing 2.5 M urea, at a flow rate of 6 mL/min. Sodium caseinate (10 mg/mL) was dissolved in the sodium phosphate buffer and filtered through a $0.45\text{-}\mu\text{m}$ filter; a volume of 10 mL was loaded onto the column. A linear gradient from 480 to 37 mM sodium phosphate was applied for 47 min followed by an isocratic step performed for 12 min in 37 mM sodium phosphate buffer, pH 6.4, containing 2.5 M urea. Absorbance was recorded at 280 nm. The collected fractions were dialyzed in distilled water for 48 h in the presence of thymol and then freeze-dried.

Affinity Chromatography on Immobilized Wheat Germ Agglutinin

Protein sample (15 mg) was dissolved in 10 mL of 20 mM Tris-HCl buffer, pH 8.0, containing 0.02% (wt/vol) sodium azide, filtered through a $0.45\text{-}\mu\text{m}$ filter, and loaded onto a column (10×100 mm) of wheat germ agglutinin (WGA) immobilized on agarose as described by Egito et al. (2001).

Ion-Exchange Chromatography

Protein fractions of interest obtained by HIC were further separated by ion-exchange chromatography (IEC) using a TSK DEAE-5PW column (7.5×75 mm, Amersham Pharmacia) connected to an ÄKTA-FPLC system according to the method used by Aoki et al. (1992) to separate phosphorylation isoforms of human β -CN. The column was equilibrated in 20 mM imidazole buffer, pH 8.0, containing 3.3 M urea and 0.08 M NaCl, at a flow rate of 0.5 mL/min. Protein sample (30 mg) was dissolved in 2 mL of the imidazole buffer and filtered through a $0.45\text{-}\mu\text{m}$ filter; the protein solution was then loaded onto the column. A linear gradient from 0.08 to 0.28 M NaCl was applied for 70 min, and absorbance was detected at 280 nm. The collected fractions were dialyzed and lyophilized.

Enzymatic Dephosphorylation of α_{S1} -CN

A 2-step method was developed to achieve full dephosphorylation of α_{S1} -CN. The first step consisted of dephosphorylation by alkaline phosphatase (EC 3.1.3.1) from bovine intestinal mucosa (1,109 U/mg; 1 unit hydrolyzes 1.0 μmol of *p*-nitrophenyl phosphate per min at pH 9.8 at 37°C ; Sigma Chemical Co., St. Louis, MO) according to Yeung et al. (2001). The enzyme was solubilized (0.5 mg/mL) in a dephosphorylation solution containing 1 mM MgCl_2 and 1 mM $\text{Zn}(\text{CH}_3\text{COO})_2$ leading to the enzyme solution. The protein sample was dissolved (3 mg/mL) in 800 μL of 20 mM Tris-HCl buf-

fer, pH 8.5, containing 1 mM phenylmethanesulfonyl fluoride (PMSF) and 0.02% (wt/vol) sodium azide. A volume of 100 μ L of the dephosphorylation solution was added to the sample and the mixture was incubated at 37°C for 10 min. The enzymatic reaction was started by addition of 10 μ L of enzyme solution and was carried out at 37°C for 24 h. During the alkaline dephosphorylation step, the reaction volume was dialyzed against 20 mM Tris-HCl buffer, pH 8.5, containing 1 mM PMSF and 0.02% (wt/vol) sodium azide to avoid inhibition by phosphate ions released in the medium. At the end of the experiment, the reaction volume was dialyzed against distilled water and freeze-dried.

The protein sample that was partially dephosphorylated by alkaline phosphatase was then treated with acid phosphatase (EC 3.1.3.2) according to Carles and Ribadeau-Dumas (1986) to improve the dephosphorylation of α_{S1} -CN. Acid phosphatase of sweet potato (6.7 U/mg of protein; 1 unit hydrolyzes 1.0 μ mol *p*-nitrophenyl phosphate per min at pH 4.8 at 37°C; Sigma) was solubilized (0.2 mg/mL) in 0.24 M ammonium sulfate buffer, pH 6.0. The sample was dissolved in 950 μ L of 50 mM sodium citrate buffer, pH 5.8, containing 1 mM PMSF and 0.02% (wt/vol) sodium azide. The reaction was started by addition of 8 μ L of enzyme solution and was carried out at 30°C for 24 h. The reaction volume was dialyzed against 50 mM sodium citrate buffer, pH 5.8, containing 1 mM PMSF and 0.02% (wt/vol) sodium azide during the course of the second dephosphorylation step to avoid inhibition by phosphate ions. After incubation, the reaction volume was dialyzed against distilled water and freeze-dried.

Electrophoresis Methods

Urea-PAGE and 2-dimensional (2D)-PAGE were performed according to Girardet et al. (2006) with minor modifications for the 2D-PAGE. The immobilized pH gradient gel strip was loaded with 100 μ g of protein sample and a 12% (vol/vol) polyacrylamide gel was performed for the second dimension.

MS Analysis

Determination of the molecular mass of the equine native and dephosphorylated α_{S1} -CN isoforms was performed by LC/ESI-MS on a PE-Sciex API III⁺ triple quadrupole mass spectrometer (Sciex, Thornhill, Canada) or on an ESI nanospray source coupled with a hybrid quadrupole time-of-flight (Q/TOF) mass spectrometer Q/Star XL (MDS Sciex, Toronto, Canada) as described previously (Girardet et al., 2006). Mass determination of proteins was done with a precision of 0.01% and 0.001% for API III⁺ triple quadrupole mass

spectrometer and hybrid Q/TOF mass spectrometer Q/Star XL, respectively. The MS calibration was external. The API III⁺ worked online with a conventional HPLC on narrow-bore columns, and collected peaks were infused onto the Q/Star XL with nano-ESI needle (Proxeon Biosystems, Odense, Denmark). The reproducibility, determined in triplicate for measurements of molecular mass of an entire protein, was ± 1 Da.

RESULTS

Preparation of α_{S1} -CN by HIC

Semipreparative HIC, an efficient FPLC method to separate the bovine caseins (Tauzin et al., 2003), was chosen to prepare equine α_{S1} -CN in sufficient quantity for further analyses, whereas reverse phase-HPLC did not allow us to obtain enough amount of material under our experimental conditions.

Equine sodium caseinate was separated by HIC into 4 main fractions (F; F1 to F4) that were analyzed by urea-PAGE (Figure 1). The major bands were numbered from 1 to 14 according to Egito et al. (2002). The fraction F1 mainly contained the different phosphorylation variants 3P to 7P (bands 13 and 14) of the low- M_r β -CN as shown by Miclo et al. (2007), whereas the isoforms 3P to 7P of the mature β -CN (bands 1 to 5) and probably of β -CN Δ^5 were recovered in F2 (Egito et al., 2002; Girardet et al., 2006).

Among the bands 6 to 12 identified as α_{S1} -CN in previous work (Egito et al., 2002), bands 11 and 12 and bands 8 to 10 were mainly recovered in fractions F3 and F4, respectively, corresponding to 2 poorly resolved peaks. Identification of bands 6 and 7 was, however, ambiguous on the present profile. The α_{S1} -CN was more complex than expected, as other minor bands were highlighted on the urea-PAGE profiles of F3 and F4. At least 5 well-focalized minor bands were visible on the profile of F4 after Coomassie Brilliant Blue staining. These bands may correspond to κ -CN isoforms because this casein displays 5 detectable bands with similar slow migration rates as those of the minor bands of F4 (Egito et al., 2001). In contrast to bovine κ -CN, equine κ -CN is always glycosylated, as no trace of κ_1 -CN (the glycan-free form) has been found in Haflinger mare's milk according to Egito et al. (2001). Glycosylated compounds were detected by Schiff's reagent staining and located with precision on the profile of fraction F4. The compounds corresponding to κ -CN were removed by WGA affinity chromatography of F4 to obtain a κ -CN-free fraction. The urea-PAGE profile of this κ -CN-free fraction seemed to be closely similar to that of F4 (electrophoretic control not shown). The Schiff's reagent did not, however, stain the minor bands.

Figure 1. Hydrophobic interaction chromatography of Haflinger mare's sodium caseinate. The sample (100 mg) was loaded onto a TSK Phenyl-5PW column (Interchim, Montluçon, France). A linear gradient from 480 to 37 mM sodium phosphate, pH 6.4, was applied for 47 min at a flow rate of 6 mL/min followed by an isocratic step at 37 mM sodium phosphate for 12 min. Composition of each collected fraction (F1 to F4) was checked by urea-PAGE (inset). For that, quantities of 40 μ g of protein were loaded on the gel. Proteins were stained by Coomassie Brilliant Blue. eCN = equine sodium caseinate; the main bands were numbered from 1 to 14 according to Egito et al. (2002).

Therefore, these minor bands might not correspond to the κ -CN that was bound on immobilized WGA, but rather to minor α_{S1} -CN isoforms not yet characterized.

Investigation of Equine α_{S1} -CN Variability

The α_{S1} -CN-containing fractions (F3 and κ -CN-free F4) were characterized by 2D-PAGE (Figure 2). The profile of F3 displayed numerous spots distributed in at least 4 groups according to their apparent molecular masses (approximately 30, 33, 37, and 39 kDa) with at least 7 isoelectric forms per group distributed in a pH range of 4.85 to 6.02. The main spots on the profile of the κ -CN-free fraction belonged to 2 groups (approximately 27 and 30 kDa) with at least 6 isoelectric forms distributed between pH 5.15 and 6.55. The complexity of the 2D-PAGE pattern of the equine α_{S1} -CN supported the hypothesis of the simultaneous presence of potential alternative splicing variants (leading to a variability in molecular mass) and phosphorylation variants (leading to a variability in isoelectric point, **pI**).

An approach by enzymatic dephosphorylation was undertaken to determine by LC-ESI/MS the masses of each dephosphorylated potential α_{S1} -CN splicing variant and, consequently, the different degrees of phosphorylation of each native isoform. In a preliminary study, either alkaline phosphatase or acid phosphatase was used with the F3 and κ -CN-free F4 fractions. It is indeed reported that these 2 enzymes are efficient to achieve complete dephosphorylation of bovine (Carles and Ribadeau-Dumas, 1986) and equine (Girardet et al., 2006) β -caseins. The 2D-PAGE analysis of the 2 enzymatically treated fractions has shown, however, very limited dephosphorylation of α_{S1} -CN regardless of the type of hydrolysis undergone, as several isoelectric spots remained in each molecular mass group (data not shown). A method consisting of a double dephosphorylation was developed to improve the removal of phosphate. The F3 and κ -CN-free F4 fractions were successively submitted to alkaline phosphatase action and to acid phosphatase action. The 2D-PAGE profiles of the dephosphorylated α_{S1} -CN fractions (termed F3d

Figure 2. Two-dimensional PAGE of Haflinger mare's α_{S1} -casein isoforms contained in the hydrophobic interaction chromatography fractions before (F3 and κ -CN-free F4) and after (F3d and κ -CN-free F4d) double dephosphorylation by alkaline phosphatase and acid phosphatase. Quantities of 100 μ g of proteins were loaded in the immobilized pH gradient gel strips. MM = apparent molecular mass.

and κ -CN-free F4d) show almost complete dephosphorylation (Figure 2). Indeed, each molecular mass group of F3d and κ -CN-free F4d contained one major spot corresponding to one fully dephosphorylated isoform, as the apparent pI of each major spot was the least acid inside each group. Nevertheless, the double dephosphorylation method seemed to be more effective with the κ -CN-free F4 fraction than with fraction F3, as the 2D-PAGE profile of F3d still displayed some minor isoelectric spots. It was noteworthy that the reverse method using acid phosphatase before alkaline phosphatase only resulted in partial dephosphorylation (data not shown).

The F3d and κ -CN-free F4d fractions were submitted to LC/ESI-MS analysis with the API III⁺ or the Q/Star XL mass spectrometer. We observed better separations on narrow-bore columns compared with nano-columns, and a greater amount of material could be loaded onto the narrow-bore columns. The peaks separated with a narrow-bore column were then col-

lected. Thus, all the results presented in this study were obtained with nano ESI needle-infused peaks on the Q/Star XL apparatus (Figure 3). Despite a complex multiplied charged ion spectrum (data not shown), 4 molecular masses (24,613; 24,486; 23,257; and 23,127 Da) have been identified to a full-length apo- α_{S1} -CN (theoretical mass of 24,614 Da) and to 3 dephosphorylated splicing variants, respectively (Table 1). The MS analysis of κ -CN-free F4d revealed 4 other molecular masses (23,751; 23,624; 22,394; and 22,266 Da) related to other dephosphorylated splicing variants of α_{S1} -CN. Theoretical masses of potential splicing variants were calculated and compared with experimental masses found. For example, the 23,257 Da mass found in F3d could be explained by alternative splicing of exon 14 leading to the translation of the α_{S1} -CN Δ^{14} variant (theoretical mass of the apoform of 23,257 Da). In the same manner, the apo- α_{S1} -CN Δ^7 (theoretical mass of 23,751 Da) and apo- α_{S1} -CN $\Delta^{7,14}$ (theoretical mass of 22,393 Da) variants were identified in the κ -CN-free

Figure 3. Reconstructed masses from liquid chromatography electrospray ionization mass spectrometry [Q/Star XL spectrometer, MDS Sciex, Toronto, Canada; molecular mass (M_r) ± 1] of Haflinger mare's α_{S1} -casein isoforms contained in the dephosphorylated fractions F3d and κ -CN-free F4d. α_{S1} -CN = full-length α_{S1} -casein; d = dephosphorylated; ΔX = deletion of the amino acid sequence encoded by exon X; ΔQ = deletion of the N-terminal residue (Gln) of the region encoded by exon 11; 0P = no phosphate group.

F4d. Other forms that presented a mass decrease of 128 ± 1 Da compared with masses found for the full-length protein and each of its splicing variants existed in each case. This corresponded to the deletion of either 1 Gln residue or 1 Lys residue. Lenasi et al. (2003) have isolated a transcript for which they have suspected a deletion of 3 nucleotides corresponding to the first codon of exon 11 encoding Gln-91 (numbered according to the full-length sequence; Figure 4). The presence of protein variants displaying a mass difference of 128 Da in Haflinger mare's milk shows that this kind of transcript is translated. Because of the double dephosphorylation treatment, 7 splicing variants of apo- α_{S1} -CN have been found in Haflinger mare's milk together with the full-length protein: α_{S1} -CN $^{\Delta Q}$ (where Gln-91 was lacking), α_{S1} -CN $^{\Delta 14}$, α_{S1} -CN $^{\Delta 14,Q}$, α_{S1} -CN $^{\Delta 7}$, α_{S1} -CN $^{\Delta 7,Q}$, α_{S1} -CN $^{\Delta 7,14}$, and α_{S1} -CN $^{\Delta 7,14,Q}$.

With knowledge of the structure and molecular masses of the different dephosphorylated splicing variants of the α_{S1} -CN present in equine milk, it was then possible to determine their different degrees of phosphorylation. The F3 and κ -CN-free F4 fractions were directly analyzed by LC/ESI-MS without dephosphorylation treatment. The spectra obtained were, however, very complex and difficult to interpret, probably because of the numerous phosphorylation and splicing isoforms in each fraction (see the 2D-PAGE profiles; Figure 2). To resolve this difficulty, another approach was undertaken. A whole α_{S1} -CN fraction recovered by HIC (corresponding to a pool of the fractions F3 and F4) was fractionated by IEC into 6 subfractions (SF; SF1 to SF6; Figure 5). Each was characterized by urea-PAGE and analyzed by LC/ESI-MS. Every spectrum was correctly interpreted despite the potential presence of glycosylated forms of κ -CN and of α_{S2} -CN not yet characterized (Figure 6). However, it was not possible to assign each experimental mass to a given electrophoretic band. The variants α_{S1} -CN $^{\Delta 7}$ with 2P to 6P and α_{S1} -CN $^{\Delta 7,14}$ with 2P to 6P (apoforms identified in κ -CN-free F4d) were found in the IEC subfractions (Table 2), as well as the variant α_{S1} -CN $^{\Delta 14}$ with 4P to 8P and the full-length protein α_{S1} -CN with 2P and with 4P to 8P (apoforms identified in F3d). It was noteworthy that the phosphorylation degree 3P was never found for variants with the sequence encoded by exon 7. Indeed, the full-length α_{S1} -CN did not present any 3P isoform, as no signal at approximately 24,854 Da was detected on the MS spectra of the subfractions SF3 to SF5 (Figure 6). In the same way, no signal at approximately 23,497 Da was detected on the MS spectra of SF2 and SF3, suggesting the absence of α_{S1} -CN $^{\Delta 14-3P}$. In all cases, no form with 0P and 1P was found.

For the native full-length protein and for all of its native alternative splicing variants identified (i.e., $\Delta 7$,

Table 1. Molecular masses (M_r) determined by liquid chromatography electrospray ionization mass spectrometry (Q/Star XL spectrometer, MDS Sciex, Toronto, Canada) of the hydrophobic interaction chromatography fractions of α_{S1} -casein after double dephosphorylation by alkaline phosphatase and acid phosphatase

Observed M_r^1	Identification ²	Theoretical M_r	Fraction ³
24,613	α_{S1} -CN-0P	24,614	F3d
24,486	α_{S1} -CN ΔQ -0P	24,486	F3d
23,257	α_{S1} -CN $\Delta 14$ -0P	23,257	F3d
23,127	α_{S1} -CN $\Delta 14, Q$ -0P	23,128	F3d
23,751	α_{S1} -CN $\Delta 7$ -0P	23,751	κ -CN-free F4d
23,624	α_{S1} -CN $\Delta 7, Q$ -0P	23,623	κ -CN-free F4d
22,394	α_{S1} -CN $\Delta 7, 14$ -0P	22,393	κ -CN-free F4d
22,266	α_{S1} -CN $\Delta 7, 14, Q$ -0P	22,265	κ -CN-free F4d

¹ $M_r \pm 1$.
² ΔX = deletion of the region encoded by exon X; ΔQ = deletion of the N-terminal residue (Gln) of the region encoded by exon 11; P = phosphate group.
³F3d and F4d = fractions F3 and F4 dephosphorylated.

$\Delta 14$, and $\Delta 7, 14$), isoforms lacking Gln-91 have been characterized, but not all of the possible phosphorylation degrees were observed (Table 2). The lack of some phosphorylation degrees might be explained by the low amounts of the variants ΔQ in milk (the spectral signal of an isoform ΔQ was systematically lower than that of the corresponding isoform having the Gln residue; Figure 6) leading to difficult detection by MS. In any case, several phosphorylation isoforms were identified for all of the variants ΔQ : α_{S1} -CN ΔQ -2P, α_{S1} -CN ΔQ -5P, α_{S1} -CN ΔQ -7P and α_{S1} -CN ΔQ -8P, α_{S1} -CN $\Delta 7, Q$ -2P to 6P, α_{S1} -CN $\Delta 14, Q$ -7P and α_{S1} -CN $\Delta 14, Q$ -8P, α_{S1} -CN $\Delta 7, 14, Q$ -2P,

α_{S1} -CN $\Delta 7, 14, Q$ -3P, α_{S1} -CN $\Delta 7, 14, Q$ -5P, and α_{S1} -CN $\Delta 7, 14, Q$ -6P.
In our milk sample, the α_{S1} -CN $\Delta 7, 14$ and α_{S1} -CN $\Delta 7$ isoforms were the principal variants of equine α_{S1} -CN, as they were recovered mainly in SF4; that is, the major IEC subfraction of α_{S1} -CN (Figure 5). The full-length protein and the variant $\Delta 14$ were mainly recovered in the minor subfractions SF5 and SF6.
The elution order observed by IEC was roughly correlated with the urea-PAGE migration by taking into account the negative charge content at basic pH (Figure 5). The α_{S1} -CN variants for which the exon 7 has been

Figure 4. Exon modular structure of the Haflinger mare's full-length α_{S1} -casein deduced from mRNA sequence (GenBank accession number AY049939; Milenkovic et al., 2002). The exon nomenclature is that used by Miranda et al. (2004). The potential sites of phosphorylation are in bold characters. The exons that can be alternatively spliced are indicated in italics. The Gln residue that can be deleted is underlined.

Figure 5. Fractionation by ion-exchange chromatography of Haflinger mare's α_{S1} -casein previously prepared by hydrophobic interaction chromatography. The sample (30 mg) was loaded onto a TSK DEAE-5PW column (Amersham Pharmacia, Uppsala, Sweden). A linear gradient from 0.08 to 0.8 M NaCl was applied for 70 min at a flow rate of 0.5 mL/min. Composition of each subfraction (SF1 to SF6) was checked by urea-PAGE (inset). For that, quantities of 40 μ g of proteins were loaded in the gel. SF = subfraction; the main bands were numbered from 8 to 12 according to Egito et al. (2002).

spliced were eluted before the full-length and $\Delta 14$ proteins and had slower rates of electrophoretic migration (Figures 5 and 6). This could be explained by the fact that the $\Delta 7$ and $\Delta 7,14$ variants displayed a less acidic pI (theoretical pI of the respective apoforms of 5.86 and 5.85) than the full-length protein and the variant $\Delta 14$ (theoretical pI of the respective apoforms of 5.47 and 5.45). Unlike reversed-phase HPLC, IEC fractionation appeared to be an efficient method to elute the different α_{S1} -CN isoforms according to their phosphorylation degree (i.e., from the least to the most phosphorylated isoforms).

DISCUSSION

Exon-Skipping Variants and Variants Involving a Cryptic Splice Site

Equine α_{S1} -CN is very complex in term of mass and charge. The presence in Haflinger mare's milk of the full-length α_{S1} -CN and of various splicing variants resulting from 2 splicing processes (i.e., exon-skipping and usage of a cryptic splice site) was clearly shown after double dephosphorylation. The molecular masses determined with an accuracy of ± 1 Da for the fully de-

phosphorylated α_{S1} -CN isoforms revealed the existence of variants α_{S1} -CN $\Delta 7$, α_{S1} -CN $\Delta 14$, and α_{S1} -CN $\Delta 7,14$ that resulted from skipping exon 7, exon 14, or both. Milenkovic et al. (2002) and Lenasi et al. (2003) have isolated the mRNAs encoded by α_{S1} -CN $\Delta 7$ and α_{S1} -CN $\Delta 14$. In Welsh pony's milk, Miranda et al. (2004) have strongly suggested the existence of the exon-skipping variant α_{S1} -CN $\Delta 7,14$ as well as the existence of α_{S1} -CN $\Delta 7$. These authors have not, however, observed any protein corresponding to α_{S1} -CN $\Delta 14$ or to the full-length protein. Therefore, our results showed that the 3 possible kinds of mRNAs generated by exon-skipping were translated into milk proteins and that the full-length α_{S1} -CN also existed in equine milk. Additionally, the major isoforms of α_{S1} -CN were the $\Delta 7$ and $\Delta 7,14$ splicing variants. This may suggest that the mechanism of alternative splicing mainly involved exon 7. Nevertheless, this hypothesis should be verified with individual milk samples because the Haflinger mare's milk sample studied in the present work corresponded to a mixture of milks from a few individuals. Indeed, we could not exclude the possibility that the isoforms detected could be attributed, in part, to the mixture of α_{S1} -caseins belonging to individuals with different polymorphic variants of α_{S1} -CN gene. However, it must be kept in mind that the experimental

Table 2. Molecular masses (M_r) determined by liquid chromatography electrospray ionization mass spectrometry (Q/Star XL spectrometer, MDS Sciex, Toronto, Canada) of the different ion-exchange chromatography subfractions of α_{S1} -casein

Observed M_r ¹	Identification ²	Theoretical M_r	Fraction ³
24,775	α_{S1} -CN-2P	24,774	SF3
24,933	α_{S1} -CN-4P	24,934	SF5
25,013	α_{S1} -CN-5P	25,014	SF5
25,094	α_{S1} -CN-6P	25,094	SF6
25,174	α_{S1} -CN-7P	25,174	SF6
25,255	α_{S1} -CN-8P	25,254	SF6
24,646	α_{S1} -CN $^{\Delta Q}$ -2P	24,646	SF3
24,886	α_{S1} -CN $^{\Delta Q}$ -5P	24,886	SF5
25,047	α_{S1} -CN $^{\Delta Q}$ -7P	25,046	SF6
25,127	α_{S1} -CN $^{\Delta Q}$ -8P	25,126	SF6
23,577	α_{S1} -CN $^{\Delta 14}$ -4P	23,577	SF5
23,656	α_{S1} -CN $^{\Delta 14}$ -5P	23,657	SF5
23,737	α_{S1} -CN $^{\Delta 14}$ -6P	23,737	SF6
23,817	α_{S1} -CN $^{\Delta 14}$ -7P	23,817	SF6
23,897	α_{S1} -CN $^{\Delta 14}$ -8P	23,897	SF6
23,689	α_{S1} -CN $^{\Delta 14,Q}$ -7P	23,688	SF6
23,769	α_{S1} -CN $^{\Delta 14,Q}$ -8P	23,768	SF6
23,911	α_{S1} -CN $^{\Delta 7}$ -2P	23,911	SF1, SF2
23,991	α_{S1} -CN $^{\Delta 7}$ -3P	23,991	SF2, SF3
24,070	α_{S1} -CN $^{\Delta 7}$ -4P	24,071	SF3
24,150	α_{S1} -CN $^{\Delta 7}$ -5P	24,150	SF4, SF5
24,230	α_{S1} -CN $^{\Delta 7}$ -6P	24,231	SF4, SF5
23,783	α_{S1} -CN $^{\Delta 7,Q}$ -2P	23,783	SF1
23,863	α_{S1} -CN $^{\Delta 7,Q}$ -3P	23,863	SF2
23,941	α_{S1} -CN $^{\Delta 7,Q}$ -4P	23,943	SF3
24,022	α_{S1} -CN $^{\Delta 7,Q}$ -5P	24,023	SF4
24,102	α_{S1} -CN $^{\Delta 7,Q}$ -6P	24,103	SF4
22,553	α_{S1} -CN $^{\Delta 7,14}$ -2P	22,553	SF1, SF2, SF3
22,633	α_{S1} -CN $^{\Delta 7,14}$ -3P	22,633	SF2, SF3
22,713	α_{S1} -CN $^{\Delta 7,14}$ -4P	22,713	SF3, SF4
22,793	α_{S1} -CN $^{\Delta 7,14}$ -5P	22,793	SF4
22,873	α_{S1} -CN $^{\Delta 7,14}$ -6P	22,873	SF4
22,425	α_{S1} -CN $^{\Delta 7,14,Q}$ -2P	22,425	SF2
22,505	α_{S1} -CN $^{\Delta 7,14,Q}$ -3P	22,505	SF2, SF3
22,665	α_{S1} -CN $^{\Delta 7,14,Q}$ -5P	22,665	SF4
22,744	α_{S1} -CN $^{\Delta 7,14,Q}$ -6P	22,745	SF4

¹ $M_r \pm 1$.² ΔX = deletion of the region encoded by exon X; ΔQ = deletion of the N-terminal residue (Gln) of the region encoded by exon 11; P = phosphate group.³SF = subfraction of α_{S1} -CN.

masses found corresponded accurately to the masses of splicing variants.

Exon skipping is frequently observed in α_{S1} -CN of other mammalian species studied and is not restricted to ruminants (Martin et al., 2002). This mechanism has been already described for α_{S1} -CN of caprine milk (Ferranti et al., 1997), ovine milk (Passey et al., 1996), and human milk (Johnsen et al., 1995). In sow's milk, 2 forms of cDNA are thought to result from exon-skipping events in mRNA (Alexander et al., 1992). The formation of several splicing variants could be explained by the α_{S1} -CN gene structure that is divided into many short exons (Martin et al., 2002). Among the species studied, human α_{S1} -CN (Johnsen et al., 1995; Martin et al., 1996) and equine α_{S1} -CN are the only α_{S1} -caseins for which exon 7 can be alternatively spliced.

For each isoform of α_{S1} -CN identified (including the full-length isoform), a shorter isoform lacking a Gln residue was characterized. This supported the occurrence of an alternative splice mechanism using a cryptic splice site located at the beginning of exon 11 and corresponding to Gln-91 (Lenasi et al., 2003). The deletion of the first codon of exon 11 is a rather frequent phenomenon occurring in other species (Martin et al., 2002) including ewe (Ferranti et al., 1995), goat (Ferranti et al., 1997), cow, and water buffalo (Ferranti et al., 1999). In the human, the first codon of exon 6' (a supplementary exon in the gene of human α_{S1} -CN between exon 6 and exon 7) encoding a Gln residue is spliced (Johnsen et al., 1995; Martin et al., 1996). According to Martin et al. (2002), the loss of a Gln residue could be explained by a splicing error by the

spliceosome. The codon CAG (encoding a Gln residue) at the beginning of an exon could be confused with the AG splice acceptor site of the adjacent intron and could be used alternatively (Smith et al., 1993).

Phosphorylation Degrees of the α_{S1} -CN Variants

Variable phosphorylation from 3P to 7P has already been shown for equine β -CN (Girardet et al., 2006). In the same way, variable phosphorylation of equine α_{S1} -CN was also suspected to explain the isoelectric heterogeneity of the protein. The MS analysis of the native α_{S1} -CN isoforms revealed that α_{S1} -CN could carry a maximum of 6P or 8P depending on the loss or conservation of exon 7, respectively. This observation is in agreement with the fact that exon 7 codes for an amino acid sequence with 2 potential sites of phosphorylation. The maximum number of phosphate groups of α_{S1} -CN agreed well with the presence of 8 potential phosphorylation sites involving Ser residues (Ser-18, Ser-58, Ser-61, Ser-75, Ser-77, Ser-79, Ser-80, and Ser-81; numbered according to the full-length sequence) located in Ser-Xxx-Glu/SerP motifs, which are recognition sequences for bovine mammary gland casein kinase (Mercier, 1981). The MS analysis did not detect any supplementary phosphate group, suggesting that Thr-62, Thr-78, and Thr-104 in Thr-Xxx-Glu/SerP motifs would not be phosphorylated. It is reported that proteins with the sequence Thr-Xxx-Glu are poor substrates for mammary gland casein kinase (Bingham and Groves, 1979; Sørensen and Petersen, 1994). On the other hand, the MS results supported the idea that the presence of exon 7 increased by 2 the minimum and maximum numbers of possible phosphorylations on the amino acid sequence encoded. Thus, the minimum number of phosphorylations went from 2P to 4P and the maximum from 6P to 8P. This suggested that the 56–63 sequence encoded by exon 7 would be systematically phosphorylated on its 2 potential sites (Ser-58 and Ser-61). Indeed, variants lacking the sequence encoded by exon 7 possessed from 2P to 6P, whereas the variant $\Delta 14$ containing the sequence encoded by exon 7 carried from 4P to 8P. Nevertheless, the full-length α_{S1} -CN displayed a minimum number of phosphorylation of 2P instead of 4P. Therefore, further analyses should be undertaken to locate precisely the sites of phosphorylation of the different isoforms.

In comparison, the bovine α_{S1} -CN is fully phosphorylated on its 8 potential sites in Ser-Xxx-Glu/SerP motifs and partly phosphorylated on a Ser residue belonging to Ser-Xxx-Asp motif (Farrell et al., 2004). In other ruminants studied, α_{S1} -caseins have high variable degrees of phosphorylation. Caprine and ovine α_{S1} -caseins carry from 7P to 10P (Ferranti et al., 1997) and 5P to 11P

(Chianese et al., 1996), respectively (the highest degree of phosphorylation depends on the allelic variant). In contrast, the human α_{S1} -CN contains only 4 identified phosphorylation sites (Ser-18, Ser-26, Ser-73, and Ser-75) despite the presence of 9 potential phosphorylation sites (Kjeldsen et al., 2007). The α_{S1} -CN of mare (monogastric species), which was phosphorylated mainly at a high level (6P or 8P) as in the case of ruminants, displayed different structural characteristics compared with the human protein.

CONCLUSIONS

In the present work, 36 different isoforms of α_{S1} -CN were identified in Haflinger mare's milk showing the translation of mRNA sequences isolated in other works. The isoforms resulted from posttranscriptional modifications (alternative splicing) and posttranslational modifications (number of phosphate groups) leading to the very complex polymorphism observed by 2D-PAGE as well as by urea-PAGE. Nevertheless, we could not exclude the possibility of the existence of different allelic variants at low frequency in the Haflinger breed. Characterization of individual milk samples is needed to investigate this possibility. Moreover, further experiments will be undertaken to locate the phosphate groups in the primary structure by tandem MS after in-gel digestion of 2D-PAGE spots.

REFERENCES

- Alexander, L. J., N. A. Das Gupta, and C. W. Beattie. 1992. The sequence of porcine α_{S2} -casein cDNA. *Anim. Genet.* 23:365–367.
- Aoki, T., T. Umeda, and Y. Kako. 1992. The least number of phosphate groups for crosslinking of casein by colloidal calcium phosphate. *J. Dairy Sci.* 75:971–975.
- Bandura, D. R., V. I. Baranov, and S. D. Tanner. 2002. Detection of ultratrace phosphorus and sulfur by quadrupole ICPMS with dynamic reaction cell. *Anal. Chem.* 74:1497–1502.
- Bingham, E. W., and M. L. Groves. 1979. Properties of casein kinase from lactating bovine mammary gland. *J. Biol. Chem.* 254:4510–4515.
- Carles, C., and B. Ribadeau-Dumas. 1986. Influence of the phosphorylation on the behaviour of bovine β -casein tryptic peptides obtained by reverse phase high performance liquid chromatography. *Lait* 66:247–256.
- Chianese, L., G. Garro, R. Mauriello, P. Laezza, P. Ferranti, and F. Addeo. 1996. Occurrence of five α_{S1} -casein variants in ovine milk. *J. Dairy Res.* 63:49–59.
- Egito, A. S., J.-M. Girardet, L. Miclo, D. Mollé, G. Humbert, and J.-L. Gaillard. 2001. Susceptibility of equine κ - and β -caseins to hydrolysis by chymosin. *Int. Dairy J.* 11:885–893.
- Egito, A. S., L. Miclo, C. Lopez, A. Adam, J.-M. Girardet, and J.-L. Gaillard. 2002. Separation and characterization of mares' milk α_{S1} -, β -, κ -caseins, γ -casein-like, and proteose peptone component 5-like peptides. *J. Dairy Sci.* 85:697–706.
- Farrell, H. M. Jr., R. Jimenez-Flores, G. T. Bleck, E. M. Brown, J. E. Butler, L. K. Creamer, C. L. Hicks, C. M. Hollar, K. F. Ng-Kwai-Hang, and H. E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk—Sixth revision. *J. Dairy Sci.* 87:1641–1674.
- Ferranti, P., F. Addeo, A. Malorni, L. Chianese, C. Leroux, and P. Martin. 1997. Differential splicing of pre-messenger RNA produces

- multiple forms of mature caprine α_{s1} -casein. *Eur. J. Biochem.* 249:1–7.
- Ferranti, P., S. Lilla, L. Chianese, and F. Addeo. 1999. Alternative nonallelic deletion is constitutive of ruminant α_{s1} -casein. *J. Protein Chem.* 18:595–602.
- Ferranti, P., A. Malorni, G. Nitti, P. Laezza, R. Pizzano, L. Chianese, and F. Addeo. 1995. Primary structure of ovine α_{s1} -caseins: Localization of phosphorylation sites and characterization of genetic variants A, C and D. *J. Dairy Res.* 62:281–296.
- Girardet, J.-M., L. Miclo, S. Florent, D. Mollé, and J.-L. Gaillard. 2006. Determination of the phosphorylation level and deamidation susceptibility of equine β -casein. *Proteomics* 6:3707–3717.
- Greenberg, R., M. L. Groves, and H. J. Dower. 1984. Human β -casein. Amino acid sequence and identification of phosphorylation sites. *J. Biol. Chem.* 259:5132–5138.
- Iametti, B. S., G. Tedeschi, E. Oungre, and F. Bonomi. 2001. Primary structure of κ -casein isolated from mares' milk. *J. Dairy Res.* 68:53–61.
- Johnsen, L. B., L. K. Rasmussen, T. E. Petersen, and L. Berglung. 1995. Characterization of three types of human α_{s1} -casein mRNA transcripts. *Biochem. J.* 309:237–242.
- Kjeldsen, F., M. M. Savitski, M. L. Nielsen, L. Shi, and R. A. Zubarev. 2007. On studying protein phosphorylation patterns using bottom-up LC-MS/MS: The case of human α -casein. *Analyst (Lond.)* 132:768–776.
- Lenasi, T., B. M. Peterlin, and P. Dovc. 2006. Distal regulation of alternative splicing by splicing enhancer in equine β -casein intron 1. *RNA* 12:498–507.
- Lenasi, T., I. Rogelj, and P. Dovc. 2003. Characterization of equine cDNA sequences for α_{s1} -, β - and κ -casein. *J. Dairy Res.* 70:29–36.
- Malacarne, M., F. Martuzzi, A. Summer, and P. Mariani. 2002. Protein and fat composition of mare's milk: Some nutritional remarks with reference to human and cow's milk. *Int. Dairy J.* 12:869–877.
- Martin, P., G. Brignon, J.-P. Furet, and C. Leroux. 1996. The gene encoding α_{s1} -casein is expressed in human mammary epithelial cells during lactation. *Lait* 76:523–535.
- Martin, P., M. Szymanowska, L. Zwierzchowski, and C. Leroux. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reprod. Nutr. Dev.* 42:433–459.
- Mercier, J.-C. 1981. Phosphorylation of caseins, present evidence for an amino acid triplet code posttranslationally recognized by specific kinases. *Biochimie* 63:1–17.
- Miclo, L., J.-M. Girardet, A. S. Egitto, D. Mollé, P. Martin, and J.-L. Gaillard. 2007. The primary structure of a low- M_r multiphosphorylated variant of β -casein in equine milk. *Proteomics* 7:1327–1335.
- Milenkovic, D., P. Martin, G. Guerin, and C. Leroux. 2002. A specific pattern of splicing for the horse α_{s1} -casein mRNA and partial genomic characterization of the relevant locus. *Genet. Sel. Evol.* 34:509–519.
- Miranda, G., M.-F. Mahé, C. Leroux, and P. Martin. 2004. Proteomic tools to characterize the protein fraction of Equidae milk. *Proteomics* 4:2496–2509.
- Ng-Kwai-Hang, K. F. 2003. Milk Proteins. Heterogeneity, Fractionation and Isolation. Pages 1881–1894 in *Encyclopedia of Dairy Sciences*. H. Roginski, J. W. Fuquay, and P. F. Fox, ed. Academic Press, London, UK.
- Ochirkhuyag, B., J.-M. Chobert, M. Dalgalarondo, and T. Haertlé. 2000. Characterization of mare caseins. Identification of α_{s1} - and α_{s2} -caseins. *Lait* 80:223–235.
- Passey, R., W. Glenn, and A. Mackinlay. 1996. Exon skipping in the ovine α_{s1} -casein gene. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 114:389–394.
- Smith, C. W., T. T. Chu, and B. Nadal-Ginard. 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell. Biol.* 13:4939–4952.
- Sørensen, E. S., and T. E. Petersen. 1994. Identification of two phosphorylation motifs in bovine osteopontin. *Biochem. Biophys. Res. Commun.* 198:200–205.
- Tauzin, J., L. Miclo, S. Roth, D. Mollé, and J.-L. Gaillard. 2003. Tryptic hydrolysis of bovine α_{s2} -casein: Identification and release kinetics of peptides. *Int. Dairy J.* 13:15–27.
- Yeung, A. C., R. P. Glahn, and D. D. Miller. 2001. Dephosphorylation of sodium caseinate, enzymatically hydrolysed casein and casein phosphopeptides by intestinal alkaline phosphatase: Implications for iron bioavailability. *J. Nutr. Biochem.* 12:292–299.