

Besma Saadaoui^{1*}
Leonardo Bianchi^{2*}
Céline Henry³
Guy Miranda²
Patrice Martin^{2**}
Christelle Cebo²

¹Faculté des Sciences de Gabès,
Université de Gabès, Gabès,
Tunisia

²INRA, UMR 1313 Unité
Génétique Animale et Biologie
Intégrative, Equipe Lait,
Génome et Santé (LGS), France

³INRA, UMR 1319 MICALIS,
Plateforme PAPSSO (Plateforme
d'Analyse Protéomique Paris
Sud Ouest), France

Received August 9, 2013

Revised December 6, 2013

Accepted January 24, 2014

Research Article

Combining proteomic tools to characterize the protein fraction of llama (*Lama glama*) milk

Llamas belong to the *Camelidae* family along with camels. While dromedary camel milk has been broadly characterized, data on llama milk proteins are scarce. The objective of this study was thus to investigate the protein composition of llama milk. Skimmed llama milk proteins were first characterized by a 2D separation technique coupling RP-HPLC in the first dimension with SDS-PAGE in the second dimension (RP-HPLC/SDS-PAGE). Llama milk proteins, namely caseins (α_{s1} -, α_{s2} -, β -, and κ -caseins), α -lactalbumin, lactoferrin, and serum albumin, were identified using PMF. Llama milk proteins were also characterized by online LC-ESI-MS analysis. This approach allowed attributing precise molecular masses for most of the previously MS-identified llama milk proteins. Interestingly, α -lactalbumin exhibits distinct chromatographic behaviors between llama and dromedary camel milk. De novo sequencing of the llama α -lactalbumin protein by LC coupled with MS/MS (LC-MS/MS) showed the occurrence of two amino acid substitutions (R62L/I and K89L/I) that partly explained the higher hydrophobicity of llama α -lactalbumin compared with its dromedary counterpart. Taken together, these results provide for the first time a thorough description of the protein fraction of *Lama glama* milk.

Keywords:

Camelids / Liquid chromatography / Mass spectrometry / Milk proteins

DOI 10.1002/elps.201300383



Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Llamas (*Lama glama*) are members of the *Camelidae* family (Artiodactyla order), which includes three genera (*Camelus*, *Lama*, and *Vicugna*) and seven living species (Arabian, Bactrian, and Wild Bactrian camels, llamas, guanacos, alpacas, and vicuñas). Like camels, the llama is known in the domestic state. While dromedary camel milk (referred to as camel milk in the subsequent text) has been extensively characterized with regard to its casein and whey protein fractions [1], almost nothing is known about milk proteins from South American camelids, including llamas.

Early studies on llama milk were limited to determining fat, protein, lactose, or macrominerals content [2, 3]. Authors concluded that llama milk contains considerably more sugar and less fat than milks of domestic ruminants (i.e. cow, goat, or sheep). A recent review also compared fatty acid composition in milk from llamas and other minor dairy animals using cow milk as a reference [4]. A few studies focused on whey proteins in llama milk [5, 6]. However, for these studies, current effective proteomic tools were not available to characterize llama milk proteins, and the protein fraction of *L. glama* milk remains to be elucidated.

A large genome sequencing project has been recently completed to help the annotation of functional elements encoded in the genome. Twenty-nine eutherians were chosen to maximize the branch length of the evolutionary tree while representing the diversity of mammalian species [7]. In this context, a first release of a low-coverage assembly

Correspondence: Dr. Christelle Cebo, Institut National de la Recherche Agronomique (INRA), Unité Génétique Animale et Biologie Intégrative (GABI), Domaine de Vilvert, Bâtiment 440, 78352, Jouy-en-Josas cedex, France
E-mail: christelle.cebo@jouy.inra.fr

Abbreviations: FDR, false discovery rate; LDH, lactadherin; m/z, mass-to-charge ratio; PGRP, peptidoglycan recognition protein

*These authors contributed equally to the work.

**Additional corresponding author: Dr. Patrice Martin,
E-mail: patrice.martin@jouy.inra.fr

Colour Online: See the article online to view Fig. 3 in colour.

of the genome of alpaca (*Vicugna pacos*), which belongs to the *Camelidae* family along with *Camelus* and *Llama* genera, has been recently released (http://www.ensembl.org/Vicugna_pacos/). Preliminary sequences for α_{s1} -casein (ENSVAG00000004344), β -casein (ENSVAG00000004350), and κ -casein (ENSVAG00000003971) are already available. The gene-building methodology for alpaca low-coverage genome relies on a whole-genome alignment to the well-annotated human genome. But in some cases, this gene-building strategy may lead to annotation errors. This is well exemplified by the β -casein encoding gene in the alpaca genome. Multiple alignment of the amino acid sequence of β -casein from 11 eutherians strongly suggests a species-specific out-splicing of exon 3 in the human β -casein gene [8]. As a consequence, since human genome is used as a reference, a putative sequence encoded by exon 3 of the β -casein encoding gene from alpaca is missing as well. Hence, the exonic sequence for the nonapeptide KEEFK-TAGE encoded by exon 3 in the camel species can be easily retrieved approximately 130 bp upstream of the provided *V. pacos* ENSVAG00000004350 genomic sequence. This observation supports the need to gain more experimental data to help the annotation of *Camelidae* milk proteins.

However, full-annotated sequences for most of camel milk proteins are already available [9]. Thus, identification of major proteins from llama milk by an MS approach can be efficiently performed by comparison with homologous proteins from a closely related species, the dromedary camel (*Camelus dromedarius*).

We combined different proteomic approaches to achieve the identification of llama milk proteins. RP-HPLC coupled with SDS-PAGE (RP-HPLC/SDS-PAGE), RP-HPLC coupled with ESI-MS (LC-ESI-MS) techniques together with LC coupled with MS/MS (LC-MS/MS) were used to characterize thoroughly the protein fraction of llama skimmed milk. In addition, we provide first data for *L. glama* milk fat globule membrane proteins, proteins associated to fat globules in milk.

2 Materials and methods

2.1 Milk samples

Llama milk samples from one 7-year-old female (5 months of lactation) were collected at the “Lamas de la Boule” breeding farm (Domaine de la Balerie, Limoise, France). Three other samples were obtained from the “Lamas des Granades” breeding farm (Restes, Saint-Remy). Two lactating females were 4 years old and 3 years old, respectively. Lactation stages were comparable among milk samples (5 months of lactation). Milk samples were centrifuged at $1000 \times g$ for 20 min at 20°C. Skimmed milk samples were stored at -80°C into small aliquots (20 μL). Dromedary camel (*C. dromedarius*) milk samples were collected from four animals at different lactation stages (50–230 days postpartum) at the experimen-

tal station of the Arid Land Institute of Medenine (Medenine, Tunisia). Camel milk samples were similarly handled.

Fat globules were recovered in the supernatant layer and washed twice with 40 mL of a 0.9% w/v NaCl solution ($1000 \times g$ at 10°C for 15 min) to remove residual caseins and whey proteins eventually adsorbed to fat globules. Milk fat globule membrane (MFGM) proteins were extracted from fat milk with an SDS-containing solution as previously described [10] with minor modifications. Briefly, lysis buffer (63 mM Tris-HCl, pH 7.5, 2%) SDS supplemented with a protease inhibitor cocktail (Complete Mini, EDTA-free, Roche Diagnostics, Meylan, France) was added to washed fat globules at a concentration of 1 mL/g of fat, incubated 1 h at room temperature with periodical vortexing, and centrifuged at $10\,000 \times g$ for 10 min. Floating cream layer was removed, lysates were centrifuged again, and then stored at -80°C for further analysis. Protein concentration (Lowry's method) was assessed with the Bio-Rad RC-DC Protein Assay according to manufacturer's instructions (Bio-Rad, Marnes-la-Coquette, France). A total of 20 μg of MFGM proteins were resolved by 10% SDS-PAGE followed by staining with Bio-Safe (Bio-Rad).

Caseins were precipitated from defatted camel and llama milk at pH 4.6 with acetic acid/sodium acetate [11] and washed twice in acetate buffer, pH 4.6. Dephosphorylation of proteins was performed as follows: samples were heated at 100°C for 5 min to inactivate proteases. Samples were then diluted to the final concentration of 2 mg/mL in 50 mM Tris, pH 7.9, 100 mM NaCl, 10 mM MgCl_2 , and 1 mM DTT. One unit of calf intestinal alkaline phosphatase (Sigma-Aldrich, Saint-Quentin Fallavier, France) was added per 20 μg of protein sample. Samples were incubated for 18 h at 37°C . Dephosphorylated samples were then subjected to online LC-ESI-MS analysis as described below.

2.2 RP-HPLC/SDS-PAGE

A total of 20 μL (llama milk analysis) or 40 μL (camel milk analysis) of skimmed samples were first clarified in a final volume of 500 μL by the addition of 0.1 M bis-Tris buffer, pH 8.0, containing 8 M urea, 1.3% tri-sodium citrate w/v, and 0.3% DTT w/v as previously described [12]. Clarified milk samples were injected on a Jupiter C5 Reverse Phase chromatographic column (300 Å, 5 μm , 250×4.6 mm; Phenomenex, Paris, France). Protein separation was achieved on a GOLD apparatus (Beckman Coulter) by applying a multistep gradient of solvent A (0.1% TFA in ultrapure water v/v) and solvent B (0.08% TFA in ACN v/v) as follows: 0–33% B in 5 min, 33–39% B in 5 min, isocratic elution at 39% B for 15 min, and 39–50% B in 15 min. The column temperature was maintained at 47°C , the flow rate was set at 1 mL/min, and the absorbance was detected at 214 nm. Eluted fractions corresponding to the major chromatographic peaks were manually collected along the run. Chromatographic fractions corresponding to each peak were pooled and then dried in a Speed Vac concentrator set at medium temperature. Samples were solubilized in Laemmli buffer containing 5% v/v β -mercaptoethanol, and denatured

5 min at 95°C. Samples were subsequently resolved by 15% SDS-PAGE. Gels were stained with Biosafe (Bio-Rad) to reveal protein bands. Prestained molecular weight markers (15–175 kDa range) were from Dutscher (Brumath, France).

2.3 LC-ESI-MS

Determination of molecular masses of milk proteins was achieved by RP-HPLC coupled with ESI-MS. Skimmed milk samples (10 and 20 µL for llama and camel milk, respectively) were first clarified by adding 0.1 M Bis-Tris buffer pH 8.0, containing 8 M urea, 1.3% trisodium citrate w/v, and 0.3% DTT w/v. Final volume was 500 µL.

A total of 20 µL clarified milk samples were injected into a Discovery® BIOWide Pore (Supelco) C5 column (150 × 2.10 mm, 300 Å). RP-HPLC analysis was carried out on an Ultimate 3000 system (Dionex), and conditions for analysis were as follows. The autosampler was kept at 10°C, and the column was maintained at 52°C. The mobile phase of the column consisted of a gradient mixture of solvent A (0.025% TFA in ultra pure water, v/v) and solvent B (0.02% TFA in ACN, v/v) at a flow rate of 0.2 mL/min. The elution condition was as follows: 28% for 2 min, 33% for 15 min, and 41% for 15 min. Protein elutes were detected by UV-absorbance at 214 nm. The column was directly interfaced with a micrOTOF™ II focus ESI-TOF mass spectrometer (Bruker Daltonics), and liquid effluent was introduced to a mass spectrometer equipped with an ESI source via silica capillary. Positive ion mode was used and mass scans were acquired over a mass-to-charge ratio (m/z) range of 20–3000. End plate offset voltage was set at –500 V and capillary voltage to 4500 V. Nebulizer gas (N_2) pressure was maintained at 17.4 psi and drying gas (N_2) flow was set at 8.0 L/min. The LC-MS system was controlled by the Hystar software version 2.3 (Bruker Daltonics). The charge number of the multicharge ions, the deconvoluted mass spectra, and the determination of M_r were obtained using Data Analysis version 3.4 software (Bruker Daltonics). Hystar postprocessing software was used to analyze UV data (Bruker Daltonics).

2.4 Identification of proteins from 1D-PAGE by MALDI-TOF-MS

Bands of interest were excised from SDS-PAGE and gel pieces were equilibrated for 5 min with 100 µL of a 50 mM ammonium bicarbonate (NH_4HCO_3) buffer. Gel pieces were then spun down and the liquid was discarded (step 1). A total of 100 µL ACN was added and samples were incubated for 5 min (step 2). Steps 1 and 2 were repeated at least twice or until the gel pieces shrunk. Samples were spun down and the supernatant was discarded. Gel pieces were then swelled in 100 µL reduction buffer containing 10 mM DTT and incubated for 1 h at 37°C. Gel pieces were shrunk with 100 µL ACN again. Samples were spun down and 100 µL alkylation buffer was added (50 mM iodoacetamide in 50 mM NH_4HCO_3). Samples were incubated for 20 min in the dark and the alky-

lation buffer was discarded. Gel slices were incubated with 100 µL ACN until they shrink. Gel particles were then spun down and the liquid was discarded (twice). Gel pieces were then rehydrated in the digestion buffer (50 mM NH_4HCO_3) containing 12.5 ng/µL trypsin (sequencing grade, Promega, Charbonnières, France) for 16 h at 37°C.

The identity of peptides was obtained using an Applied Biosystems Voyager DE super STR MALDI-TOF mass spectrometer (AB-sciex) equipped with a nitrogen laser emitting at 337 nm. One microliter of tryptic digest was mixed on the stainless steel MALDI plate with 1 µL of alpha-cyano-4-hydroxycinnamic acid (Sigma-Aldrich; 5 mg/mL in ACN/0.3% TFA v/v) and dried at room temperature. Spectra were recorded in positive reflector mode with 20 kV as accelerating voltage, a delayed extraction time of 120 ns, and a 63% grid voltage. Internal mass calibration was performed using trypsin autolysis products $[M + H]^+ = 842.51$ and $[M + H]^+ = 2211.11$. Mass spectra were treated by Data Explorer 4.2 (Applied Biosystems). Spectral profiles were collected in the $m/z = 800$ –3000 Da. Proteins were identified among Swiss-Prot (2012.03.21 release) and NCBIInr (2011.10.10 release) databases using the MS-Fit software in the Protein Prospector package (<http://prospector.ucsf.edu/prospector/mshome.htm>) with the following parameters: trypsin specificity, one missed cleavage allowed, 50 ppm mass accuracy, carbamidomethylation (cysteine), and oxidation (methionine) as fixed and variable modifications, respectively.

2.5 De novo sequencing of the α-lactalbumin protein

The SDS-PAGE band corresponding to the α-lactalbumin protein in the camel and llama species was cut and digested with 100 ng trypsin as described above. HPLC was performed on a Dionex RSLCnano (Thermo Fisher Scientific). A 4 µL sample was loaded at 20 µL/min on a precolumn cartridge (stationary phase: C18 PepMap 100, 5 µm; column: 300 µm id, 5 mm; Dionex) and desalted with 0.08% TFA and 2% ACN. After 4 min, the precolumn cartridge was connected to the separating PepMap C18 column (stationary phase: C18 PepMap 100, 3 µm; column: 75 µm id, 150 mm; Dionex). Buffers were 0.1% formic acid ($HCOOH$) and 2% ACN (A), and 0.1% $HCOOH$ and 80% ACN (B). The peptide separation was achieved with a linear gradient from 0 to 36% B for 14 min at 300 nL/min. Including the regeneration step at 100% B and the equilibration step at 100% A, one run took 22 min. Eluted peptides were analyzed online with an LTQ-Orbitrap mass spectrometer (Thermo Electron) using a nanoelectrospray interface. Ionization (1.3 kV ionization potential) was performed with liquid junction and a capillary probe (10 µm id; New Objective). Peptide ions were analyzed using Xcalibur 2.07 with the following data-dependent acquisition steps: (i) full MS scan in orbitrap ($m/z = 300$ –1600, profile mode) and (ii) MS/MS in linear trap ($Q = 0.25$, activation time = 30 ms, and collision energy = 45%; centroid mode). Step (ii) was

repeated for the four major ions detected in step (i). Dynamic exclusion time was set to 40 s.

The *C. dromedarius* NCBI database (2010, 1044 protein entries) was used. This database, in conjunction with protein to peptide reverse database (created by Database Manager version 2.0), and contaminant databases were searched by algorithm X!Tandem (version 2012.10.01.1) with the software X!Tandem pipeline (version 3.3.1) developed by PAPPISO platform (<http://pappso.inra.fr/bioinfo/>). The precursor mass tolerance was 10 ppm and the fragment mass tolerance was 0.5 Da. One missed cleavage was allowed for trypsin. The carbamidomethylation of cysteine residues was set as a fixed modification and the oxidation of methionine residues as a variable modification. Results were filtered using inbuilt X!Tandem parser with peptide *E*-value of 0.05, a protein log (*E*-value) of -4.0 , and a minimum of two peptides. The unmatched peptides were exported by X!Tandem pipeline and used by DeNovo pipeline (version 1.4.2) developed by PAPPISO platform (<http://pappso.inra.fr/bioinfo/>) to perform de novo sequencing from IT mass spectrometers. The value of quality score was 0.8: This de novo pipeline performs automated analysis by connection to Fasts, a homology search in an iterative mode to identify proteins from peptide sequences. Only sequence with pepNovo score higher than 70 (corresponding to percentage of homology) was kept.

2.6 Identification of milk proteins by 1D-LC-MS/MS analysis

Each lane of gel was cut in 24 pieces (1 mm^2) and transferred into 96-well polypropylene plate that has been pierced by laser (4titude, Framestar, 0750/Las). Gel pieces were washed for 15 min with an ACN/100 mM NH_4HCO_3 mixture (1:1). Digestion was performed in 50 mM NH_4HCO_3 pH 8.0 and the quantity of modified trypsin (Promega, sequencing grade) was $0.1\text{ }\mu\text{g}$ per sample. Digestion was achieved for 6 h at 37°C . The supernatant was transferred by centrifugation in another 96-well plate without holes (4titude, Framestar, 0960/C). Peptides were extracted by 5% HCOOH in water/ACN v/v. Tryptic peptides were dried and resuspended in $20\text{ }\mu\text{L}$ of 0.1% v/v TFA. Few samples have been pooled for injection.

A QExactive (Thermo Fisher Scientific) coupled with Eksigent nano-LC (AB-sciex) was used for the nano-LC-MS/MS analysis of peptides. A total of $4\text{ }\mu\text{L}$ were injected on the nano-LC-Ultra system (Eksigent) chain. Sample was loaded at $7.5\text{ }\mu\text{L}/\text{min}$ on the precolumn cartridge (PepMap 100 C18, $5\text{ }\mu\text{m}$, $120\text{ }\text{\AA}$, 5 mm ; Dionex) and desalted with 0.1% HCOOH. Then, peptides were separated with a gradient of ACN on the reverse-phase column (C18 Biosphere, $3\text{ }\mu\text{m}$, $120\text{ }\text{\AA}$, $75\text{ }\mu\text{m}$ id, 15 cm).

Buffers were 0.1% HCOOH in water (A) and 0.1% HCOOH in ACN (B). The peptide separation was achieved with a linear gradient from 5 to 30% B for 40 min at $300\text{ nL}/\text{min}$. Eluted peptides were analyzed online with a QExactive mass spectrometer (Thermo Electron) using a nanoelectrospray interface. Ionization (1.8 kV ionization poten-

tial) was performed with stainless steel emitters ($30\text{ }\mu\text{m}$ id; Thermo Electron). Peptide ions were analyzed using Xcalibur 2.1 with the following data-dependent acquisition steps: (i) full MS scan ($m/z = 400\text{--}1400$) and (ii) MS/MS. Step (ii) was repeated for the eight major ions detected in step (i). Dynamic exclusion was set to 40 s. Lock mass option was chosen “best,” MS resolution was 70 000 at $m/z = 400$, autogain control was 1×10^6 , and maximum injection time was 250 ms. For MS2, the resolution was 17 500 at $m/z = 400$, autogain control was 2×10^5 , maximum injection time was 120 ms, isolation window $m/z = 3$, normalized collision energy = 30%, underfill ratio = 0.5%, intensity threshold = 8.3×10^3 , charge state = 2, 3, 4. Dynamic exclusion was set to 40 s.

The *Camelidae* database was downloaded from the NCBI website (<http://ncbi.nlm.gov>, November 2013, 23 501 protein entries). This database (merged in conjunction with reverse and contaminant databases) was searched by X!Tandem (version 2013.06.15, <http://www.thegpm.org/tandem/>) using the X!Tandempipeline (version 3.3.1) developed by the PAPPISO platform (<http://pappso.inra.fr/bioinfo/>). Enzymatic cleavage was declared as a trypsin digestion with one possible miscleavage. Cysteine carbamidomethylation and methionine oxidation were set to static and possible modifications, respectively. Precursor mass was 10 ppm and fragment mass tolerance was 0.02 Th. Only peptides with an *E*-value smaller than 0.1 were reported. Identified proteins were filtered and grouped using XTandem Pipeline (<http://pappso.inra.fr/bioinfo/xtandempipeline/>) according to (i) a minimum of two different peptides were required with an *E*-value smaller than 0.05, (ii) a protein *E*-value (calculated as the product of unique peptide *E*-values) smaller than 10^{-2} . To take redundancy into account, proteins with at least one peptide in common were grouped. This allowed to group proteins of similar function. Within each group, proteins with at least one specific peptide relatively to other members of the group were reported as subgroups. Detailed results are available in the publicly available PROTEICdb at http://pappso.inra.fr/protic/llama_cebo.

3 Results

3.1 Identification of major proteins from *Camelidae* milk

Proven proteomic tools [12] were used to characterize the protein fraction of *Camelidae* milk. This approach couples RP-HPLC in the first dimension with SDS-PAGE in the second dimension. The described protocol allowed an effective and reproducible separation of the skimmed milk protein fraction. An RP-HPLC elution profile of llama skimmed milk proteins is shown in Fig. 1A. Up to nine peaks (labeled a–i) were identified, and the protein content of each peak was further analyzed by SDS-PAGE (Fig. 1B). Most of chromatographic peaks contained a major protein component, with the exception of peak d, which was noticeably constituted by two proteins, identified by MALDI-TOF PMF as α_{s1} - and

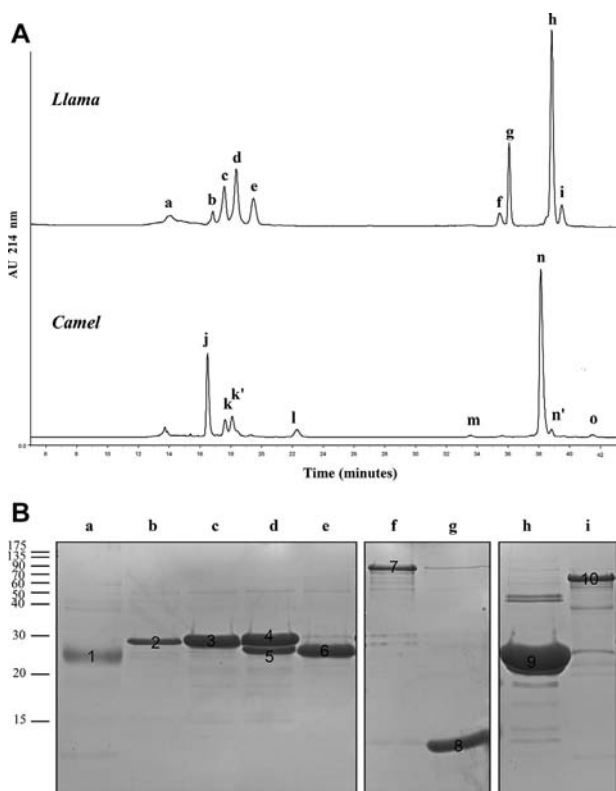


Figure 1. Bidimensional RP-HPLC/SDS-PAGE analysis of *Lama glama* (llama) and *Camelus dromedarius* (camel) milk proteins fractions. (A) Elution profile of proteins from individual milk samples. Eluted fractions were then collected and subsequently analyzed by SDS-PAGE. (B) SDS-PAGE analysis of peaks a–i identified in (A). Molecular weight markers (kDa) are indicated at the left of the gel. Main electrophoretic bands were identified by a PMF analysis as κ -casein (band 1), α_{s1} -casein (bands 2–4), α_{s2} -casein (bands 5–6), lactoferrin (band 7), α -lactalbumin (band 8), β -casein (band 9), and serum albumin (band 10). See Table 1 for protein identification details.

α_{s2} -caseins (Table 1). Thus, identified proteins eluted from the reverse-phase column were in the following order: κ -casein (peak a), α_{s1} -casein (peaks b–d), α_{s2} -casein (peaks d–e), lactoferrin (peak f), α -lactalbumin (peak g), β -casein (peak h), and serum albumin (peak i). Interestingly, the α_{s1} -casein fraction was resolved in two or three peaks, depending on the individual. Indeed, peak b, which was found to contain a α_{s1} -casein-related protein (Table 1), was present in only two of the four llama milk samples analyzed.

Camel milk samples were analyzed using RP-HPLC under the same chromatographic conditions. A representative elution profile of skimmed proteins from camel milk is shown in Fig. 1A. Major peaks were also collected and analyzed by SDS-PAGE. Protein identification was performed by PMF (data not shown). As for llama milk proteins, most of chromatographic peaks contained a major protein component, with the exception of peaks k–k', which were constituted by two proteins, identified by PMF as α_{s1} - and α_{s2} -caseins. PMF-identified proteins in camel milk were α_{s1} -casein (peaks j–k'), α_{s2} -casein (peaks k–k'), α -lactalbumin (peak l), the peptido-

glycan recognition protein (PGRP; peak m), β -casein (peaks n–n'), and lactoferrin (peak o). Remarkably, retention times for α -lactalbumin were dramatically different between the llama and the camel species (36 and 22 min after the beginning of the run for llama and camel α -lactalbumin, respectively). Thus, α -lactalbumin from llama milk was eluted at a later time than its camel counterpart. Since the hydrophobicity of the protein is the main parameter affecting the retention time in reverse-phase chromatography, it may be concluded from these observations that the α -lactalbumin from llama milk is more hydrophobic than its camel counterpart.

3.2 Identification of minor proteins from llama milk

Our SDS-PAGE analyses clearly showed that, besides major milk proteins, some low-abundance proteins were present in llama milk samples. In order to get a comprehensive view of all the proteins present in llama milk, we performed LC-MS/MS experiments on whole llama milk samples. One microliter nondefatted llama milk (diluted 1:5 with water) was analyzed by 15% SDS-PAGE. The lane was cut into 24 bands and then bands were digested by trypsin. Resulting peptides were analyzed by MS/MS. A total of 81 functional groups of proteins were identified in llama milk (Supporting Information Table 1). We calculated the false discovery rate (FDR) performing identification with the *Camelidae* database and the protein to peptide reverse database and the FDR was 0.1%. Among these proteins were previously identified proteins (mostly caseins and major whey proteins). A high number of immune defense related proteins (e.g. immunoglobulins and antibacterial proteins) were identified in llama milk. Noticeably, the PGRP, which is highly expressed in camel milk, was also identified in llama milk by MS/MS.

3.3 Determination of the molecular masses of major proteins from camel and llama milk

Alternatively, we used an online protocol coupling reverse-phase chromatography with ESI-MS (LC-ESI-MS) to determine the molecular masses (M_r) of previously identified proteins in camel and llama milk. Chromatographic elution profiles were highly comparable with those previously obtained in our RP-HPLC/SDS-PAGE approach (Fig. 2), thus permitting the attribution of a molecular mass (gained from LC-ESI-MS experiments) to a previously MS-identified milk protein (gained from RP-HPLC/SDS-PAGE combined with MALDI-TOF-MS). Regarding the camel species, the comparison between measured molecular masses and the theoretical values obtained from the amino acid sequences of most of camel milk proteins allowed the accurate identification of the protein corresponding to each chromatographic fraction, once the PTMs of caseins (essentially phosphorylation) were taken into account (Table 2). We identified the long isoform of α_{s1} -casein (UniprotKB O97943–1), the protein with six phosphate groups being the major posttranslational variant observed. A posttranscriptional variant (UniprotKB O97943–2)

Table 1. MALDI-TOF PMF identification results of major proteins from llama milk

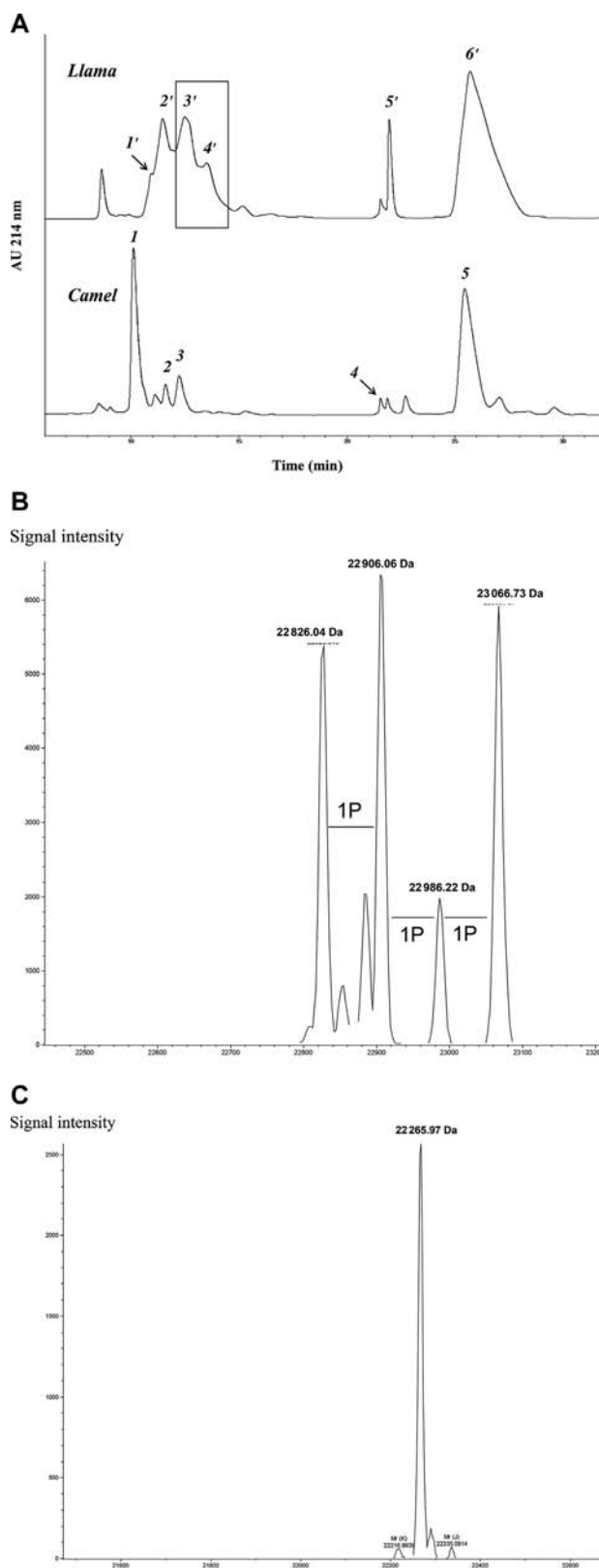
HPLC fraction ^{a)}	SDS-PAGE band ^{b)}	Identified protein (UniprotKB accession) <i>Species</i>	Mr (Da) ^{c)}	Mr of tryptic peptides			Peptide sequence ^{d)}
				<i>m/z</i> submitted	<i>MH</i> ⁺ matched	<i>Delta</i> ppm	
a	1	κ -Casein (P79139) <i>Camelus dromedarius</i>	18 254	1284.7091	1284.6741	27.2578	(K)YFPIQFVQSR(Y)
				1560.7620	1560.7236	24.6347	(R)YPSYGINYYQHR(L)
b	2	α_{s1} -Casein (O97943) <i>Camelus dromedarius</i>	25 308	959.4934	959.4945	−1.15	(R)YLEELHR(L)
				1314.7019	1314.7165	−11.1	(R)YLEELHRLNK(Y)
c	3	α_{s1} -Casein (O97943) <i>Camelus dromedarius</i>	25 308	1360.6820	1360.6703	8.60	(R)QENIDELKDTR(N)
				862.3731	862.3723	0.852	(K)EDMPSQR(Y)
				878.3499	878.3673	−19.8	(K)EDMPSQR(Y)
				955.5932	955.5935	−0.265	(K)LLQLEAIR(D)
d	4	α_{s1} -Casein (O97943) <i>Camelus dromedarius</i>	25 308	1331.6444	1331.6624	−13.5	(K)DILKEDMPSQR(Y)
				1360.6441	1360.6703	−19.2	(R)QENIDELKDTR(N)
				862.3804	862.3723	9.33	(K)EDMPSQR(Y)
				878.3747	878.3673	8.43	(K)EDMPSQR(Y)
	5	α_{s2} -Casein (O97944) <i>Camelus dromedarius</i>	21 266	955.5929	955.5935	−0.640	(K)LLQLEAIR(D)
				1331.6372	1331.6624	−18.9	(K)DILKEDMPSQR(Y)
				879.5385	879.5411	−2.92	(K)KVAIHPSK(E)
				913.4723	913.4778	−5.98	(K)ISQFYQK(W)
				929.4667	929.4727	−6.52	(K)TVYQYQK(T)
				1127.5560	1127.5666	−9.47	(K)TMTWPNHIK(R)
e	6	α_{s2} -Casein (O97944) <i>Camelus dromedarius</i>	21 266	1143.5457	1143.5615	−13.8	(K)TMTWPNHIK(R)
				1299.6759	1299.6627	10.2	(K)TMTWPNHIK(R)
				879.5365	879.5411	−5.16	(K)KVAIHPSK(E)
				913.4692	913.4778	−9.36	(K)ISQFYQK(W)
				929.4624	929.4727	−11.1	(K)TVYQYQK(T)
				1127.5492	1127.5666	−15.4	(K)TMTWPNHIK(R)
	7	Lactoferrin (Q9TUM0) <i>Camelus dromedarius</i>	75 251	1143.5457	1143.5615	−13.9	(K)TMTWPNHIK(R)
				1615.6488	1615.6727	−14.8	(K)EDICSTFC EEAVR(N)
				844.4977	844.4887	10.7	(K)DSALGLLR(I)
				848.3680	848.3832	−17.8	(K)CAQWQR(R)
f	7	<i>Camelus dromedarius</i>	75 251	865.3630	865.3621	0.976	(K)CQEWQR(Q)
				876.4111	876.4210	−11.3	(K)NTEQWAK(D)
				889.4887	889.4890	−0.328	(K)ITWNSLR(G)
				929.4548	929.4550	−0.181	(K)FCLFQSK(T)
				935.5425	935.5421	0.409	(R)VPSHAVVAR(S)
				945.4202	945.4207	−0.532	(K)SCHTAVDR(T)
				947.5631	947.5673	−4.40	(K)GYLAVAVVR(K)
				948.4192	948.4204	−1.25	(K)CVPNSSER(Y)
				1017.5167	1017.5211	−4.35	(R)ETAAEVELR(R)
				1073.5000	1073.5157	−14.6	(K)KSCHTAVDR(T)
g	8	α -Lactalbumin (P00710) <i>Camelus dromedarius</i>	14 431	1097.4994	1097.5051	−5.14	(R)YYGYTGAFR(C)
				1325.6153	1325.6406	−19.1	(R)FECIQAI STEK(A)
				1421.6873	1421.7093	−15.5	(R)CLAENVGDVAFVK(D)
				811.4062	811.4131	−8.50	(K)QFTKCK(L)
				1120.5012	1120.5092	−7.10	(K)LEQWQCEK(W)
h	9	β -Casein (Q9TVD0) <i>Camelus dromedarius</i>	24 651	1124.4690	1124.4711	−1.85	(R)NICDISCDK(F)
				1225.6071	1225.6212	−11.5	(R)EYGLFQINNK(I)
				856.5266	856.5251	1.82	(K)ETIIPKR(K)
				1436.8256	1436.8624	−25.6	(R)GLHPVPQPLVPVIA(-)
				1651.8477	1651.8989	−30.9	(K)VLPVPQQMVPYPQR(A)
i	10	Serum albumin (P49064) <i>Felis catus</i>	65 846	1667.8312	1667.8938	−37.5	(K)VLPVPQQMVPYPQR(A)
				898.4749	898.4815	−7.39	(R)LCVLHEK(T)
				919.5035	919.5030	0.554	(K)LC TVASLR(D)
				927.4952	927.4934	1.87	(K)YLYEIR(R)
				1075.4997	1075.4989	0.747	(R)NECFLOHK(D)
				1138.5032	1138.4980	4.57	(K)CCTESLVNR(R)
				1149.6067	1149.6150	−7.21	(K)LVNEVTEFAK(G)
				1749.6213	1749.6625	−23.6	(K)ECCHGDLLECADDR(A)

a) Letters refer to HPLC fractions as depicted in Fig. 1A.

b) Numbers correspond to SDS-PAGE bands as shown in Fig. 1B.

c) Molecular mass of the mature protein without any PTMs.

d) Methionine residues modified by oxidation are indicated in italics. Cysteine residues modified by acrylamide are indicated in bold.



resulting from the out-splicing of exon 16 of the α_{s1} -casein transcript was also identified (Table 2). Other caseins identified in camel milk were α_{s2} -casein (UniprotKB O97944) displaying ten phosphate groups and β -casein (UniprotKB Q9TVD0), the protein with four phosphate groups being the major posttranslational variant identified (Table 2). Other molecular masses corresponding to caseins were phosphorylation isoforms (± 79.98 Da) or cryptic splice variants resulting from the deletion of a nucleotide triplet encoding a glutamine residue (-128 Da; Table 2).

Caseins were also precipitated from camel milk and then they were subjected to alkaline phosphatase hydrolysis. Dephosphorylated samples were subsequently online analyzed with LC-ESI-MS as performed for native camel milk samples. Again, dromedary camel caseins were undoubtedly identified by means of comparison between observed molecular masses and theoretical molecular masses (calculated from the amino acid sequence of the protein by using the PeptideMass tool at the ExPASy website (http://web.expasy.org/peptide_mass/); Table 3). Results were consistent with our previous observations with regard to the level of phosphorylation of α_{s1} , α_{s2} , and β -caseins from camel milk (Table 2). Indeed, differences between observed molecular masses of fully processed (native) caseins (Table 2) and dephosphorylated α_{s1} , α_{s2} , and β -caseins from camel milk (Table 3) were ca. 480 Da (+6 phosphate groups), 799 Da (+10 phosphate groups), and 320 Da (+4 phosphate groups), respectively. In addition, because observed molecular masses of camel caseins were very close to the theoretical molecular masses calculated from the amino acid sequence of the protein (i.e. molecular masses of caseins displaying no PTMs), the dephosphorylation of caseins by alkaline phosphatase was assumed to be complete. Similar analyses were performed on caseins precipitated from llama milk samples. Table 3 reports observed molecular masses for dephosphorylated caseins from llama milk. Since the dephosphorylation process was assumed to be complete (as observed for camel caseins), these molecular masses probably correspond to the molecular masses of llama milk proteins without any phosphate moieties.

Again, chromatographic elution profiles of llama milk proteins were highly comparable with those previously obtained in our RP-HPLC/SDS-PAGE approach (Fig. 2), therefore permitting the attribution of a molecular mass (gained from LC-ESI-MS experiments) to a previously MALDI-MS-identified llama milk protein (Fig. 1, Table 1). Thus, two polymorphic forms (variants A and B) of α_{s1} -casein were found,

Figure 2. Online LC-ESI-MS analysis of llama and camel milk samples. (A) RP-HPLC profiles of llama and camel milk samples. Molecular masses corresponding to major eluted peaks are indicated in Tables 2 (camel milk) and 4 (llama milk). (B) Deconvoluted spectrum of chromatographic peaks corresponding to the α_{s2} -casein fraction (squared) from llama milk ($M_r = 22\,906.48$ Da) together with its phosphorylation isoforms (mass difference = 79.98 Da). (C) Deconvoluted spectrum of the chromatographic region containing the α_{s2} -casein protein after dephosphorylation of the casein fraction by the calf intestinal alkaline phosphatase.

Table 2. Determination of molecular masses of camel milk proteins using LC-ESI-MS

Peak number ^{a)}	Identified protein	UniprotKB accession	Observed molecular mass (Da)	Theoretical molecular mass (Da)	Delta ppm
1	α_{s1} -Casein, short isoform, 6P ^{b)}	O97943–2	24 769.29	24 769.12	+7.1
	α_{s1} -Casein, short isoform, 6P, splice variant ^{c)}	n/a	24 641.13	24 640.98	+6.1
	α_{s1} -Casein, short isoform, 5P	O97943–2	24 689.48	24 689.14	+14.0
	α_{s1} -Casein, short isoform, 5P, splice variant ^{c)}	n/a	24 561.47	24 561.01	+18.9
	α_{s1} -Casein, short isoform, 7P	O97943–2	24 849.22	24 849.09	+5.0
	α_{s1} -Casein, short isoform, 7P, splice variant ^{c)}	n/a	24 721.65	24 720.96	+28.2
2	α_{s1} -Casein, long isoform, 6P	O97943–1	25 787.25	25 787.21	+1.7
	α_{s1} -Casein, long isoform, 6P, splice variant ^{c)}	n/a	25 659.03	25 659.08	–1.9
	α_{s1} -Casein, long isoform, 5P	O97943–1	25 707.04	25 707.23	–7.3
	α_{s2} -Casein, 10P	O97944	22 065.86	22 065.69	+7.4
3	α -Lactalbumin	P00710	14 430.25	14 430.35	–7.4
4	Peptidoglycan recognition protein 1	Q9GK12	19 143.36	19 143.64	–14.6
5	β -Casein, 4P	Q9TVDO	24 971.03	24 970.68	+13.8
	β -Casein, 4P, splice variant ^{c)}	n/a	24 842.64	24 842.55	+3.5
	β -Casein, 3P	Q9TVDO	24 890.87	24 890.70	+6.8
	β -Casein, 2P	Q9TVDO	24 811.47	24 811.70	–9.5

a) Peak number refers to Fig. 2.

b) P: phosphate group.

c) Molecular mass corresponding to a cryptic splice variant resulting from the deletion of a nucleotide triplet encoding a glutamine residue (mass difference: 128 Da).

Table 3. Determination of molecular masses of dephosphorylated camel and llama milk caseins using LC-ESI-MS

Species	Identified protein	UniprotKB accession	Observed molecular mass (Da)	Theoretical molecular mass (Da) ^{a)}	Delta ppm
Camel	α_{s1} -Casein, long isoform	O97943	25 307.26	25 307.33	–2.6
	α_{s1} -Casein, short isoform	O97943–2	24 289.20	24 289.23	–1.2
	α_{s2} -Casein	O97944	21 266.11	21 265.89	+10.1
	β -Casein	Q9TVDO	24 650.89	24 650.76	+5.3
Llama	α_{s1} -Casein, long isoform A	n/a	25 184.92	n/a	–
	α_{s1} -Casein, long isoform B	n/a	25 230.05	n/a	–
	α_{s2} -Casein	n/a	22 265.97	n/a	–
	β -Casein	n/a	24 650.88	n/a	–

a) Theoretical molecular masses of dromedary camel caseins were calculated from the amino acid sequence of the protein by using the PeptideMass tool at the ExPASy website (http://web.expasy.org/peptide_mass/). No valid amino acid sequences are currently available for llama milk proteins. Theoretical values for molecular masses of llama milk proteins are thus not available.

featuring a mean molecular mass of 25 585.23 Da (peak 2', Fig. 2A) and 25 630.39 Da (peak 3', Fig. 2A), respectively (Table 4). Considering molecular masses of nonphosphorylated proteins (25 184.92 and 25 230.05 Da for α_{s1} -casein variants A and B, respectively; Table 3), these molecular masses may correspond to α_{s1} -casein variants displaying five phosphate groups (+399.9 Da). Remarkably, α_{s1} -casein variant A was present in only two of the four llama milk samples analyzed, whereas α_{s1} -casein variant B was detected in all llama milk samples. A molecular mass of 24 565.99 Da was also attributed to a α_{s1} -casein-related protein (peak 1', Fig. 2A, Table 4). According to our previous RP-HPLC/SDS-PAGE experiments, this peak, which corresponds very likely to a short isoform of α_{s1} -casein A (peak b, Fig. 1), was present in only two of the four llama milk samples analyzed by LC-ESI-MS. A mean molecular mass of 22 906.10 Da was observed for the major posttranslational variant of a

α_{s2} -casein protein (peak 3', Fig. 2A and B). However, another molecular mass ($22\,825.98 \pm 0.26$) probably corresponding to a α_{s2} -casein phosphorylation isoform was observed in all llama milk samples analyzed (peak 4', Fig. 2A and B, Table 4). Interestingly, after dephosphorylation of the casein fraction by the calf intestinal alkaline phosphatase, a single peak ($M_r = 22\,265.97$ Da) was observed in the chromatographic region containing the α_{s2} -casein protein (Fig. 2C). This molecular mass may correspond to the α_{s2} -casein without any PTMs (i.e. phosphorylation), provided that the dephosphorylation process was assumed to be complete. In one llama milk sample, an intense ion (as detected by ESI-MS) was attributed to a β -casein protein with a molecular mass of 24 970.88 Da, a mass very close to that (24 971.03 Da) observed for the major camel counterpart displaying four phosphate groups (Table 2). Other molecular masses attributed to β -casein-related proteins from llama milk were phosphorylation

Table 4. Determination of molecular masses of llama milk proteins using LC-ESI-MS

Peak number ^{a)}	Putative identification of the protein	Observed molecular mass (Da)
1'	α_{s1} -Casein, short isoform A, 5P ^{b)}	24 565.99 $\pm$ 0.42 ($n = 2$)
2'	α_{s1} -Casein, long isoform A, 5P	25 585.23 $\pm$ 0.23 ($n = 2$)
3'	α_{s1} -Casein, long isoform B, 5P	25 630.39 $\pm$ 0.18 ($n = 4$)
	α_{s2} -Casein, 8P	22 906.10 $\pm$ 0.16 ($n = 4$)
4'	α_{s1} -Casein, long isoform B, 6P	25 709.94 $\pm$ 0.44 ($n = 4$)
	α_{s2} -Casein, 7P	22 825.98 $\pm$ 0.26 ($n = 4$)
5'	α -Lactalbumin	14 314.17 $\pm$ 0.04 ($n = 4$)
6'	β -Casein, 4P	24 970.88 ($n = 1$)
	β -Casein, 4P, splice variant ^{c)}	24 842.62 $\pm$ 0.10 ($n = 4$)
	β -Casein, 3P	24 890.91 $\pm$ 0.11 ($n = 4$)
	β -Casein, 3P, splice variant ^{c)}	24 763.92 $\pm$ 1.15 ($n = 2$)
	β -Casein, 2P	24 811.48 $\pm$ 0.58 ($n = 3$)
	β -Casein, 2P, splice variant ^{c)}	24 682.66 ($n = 1$)
	β -Casein, 1P	24 731.05 $\pm$ 0.22 ($n = 2$)
	β -Casein, 0P	24 651.35 ($n = 1$)

a) Peak number refers to Fig. 2.

b) P: phosphate group (mass difference: 79.98 Da). The number of phosphate moieties on each protein species is deduced from molecular masses observed by LC-ESI-MS after dephosphorylation of precipitated llama caseins (see Table 3).

c) Molecular mass corresponding to a cryptic splice variant resulting from the deletion of a nucleotide triplet encoding a glutamine residue (mass difference: 128 Da).

isoforms (24 890.91, 24 811.48, 24 731.05, and 24 651.35 Da, the latter molecular mass probably corresponding to the β -casein protein bearing no phosphate group). Interestingly, after dephosphorylation of the casein fraction by the alkaline phosphatase, a molecular mass of 24 650.88 Da was attributed to the major peak corresponding to β -casein (Table 3). Taken together, these data confirm the occurrence of four phosphate groups (+319.9 Da) on the β -casein protein in both species. Finally, a mean molecular mass of 24 842.62 was attributed to a β -casein isoform in all llama milk samples analyzed by LC-ESI-MS (peak 6', Fig. 2). This molecular mass corresponds very likely to a β -casein cryptic splice variant resulting from the deletion of a nucleotide triplet encoding a glutamine residue (−128 Da) and bearing four phosphate groups (Table 4).

Interestingly, observed molecular masses for α -lactalbumin were 14 430.25 and 14 314.17 Da in the camel and llama species, respectively. Therefore, amino acid substitutions in the protein sequence of α -lactalbumin from llama milk should be associated with a loss of around 116 Da compared with the molecular mass of its camel counterpart.

3.4 De novo sequencing of the α -lactalbumin protein from camel and llama milk

Whatever the chromatographic method we used to resolve skimmed milk protein separation (RP-HPLC/SDS-PAGE or LC-ESI-MS), α -lactalbumin from llama milk was systematically eluted at a later time than camel α -lactalbumin, thus suggesting that α -lactalbumin from llama milk contain one or more hydrophobic amino acid residues than α -lactalbumin protein from camel milk. Since the llama genome is not sequenced, identification of peptides by a classical database search approach was not feasible. We therefore decided to perform a de novo sequencing approach to detect putative amino acid mutations in the protein sequence for α -lactalbumin from llama milk. Indeed, the de novo sequencing extracts protein sequence information directly from the spectrum peaks derived from peptide fragment ions without recourse to any protein database. Skimmed milk proteins from camel and llama milk were thus separated by SDS-PAGE and the band corresponding to the α -lactalbumin protein in each species was picked and digested with trypsin. Peptides resulting from the enzymatic digestion were extracted, and then they were separated by nano-LC coupled with MS/MS (LC-MS/MS). We used an *in-house* integrative approach combining direct peptide sequencing from tandem spectra with blast/FASTA searches against the *C. dromedarius* NCBI database. Results were then validated by manual curation of CID spectra corresponding to peptides displaying an amino acid mutation as compared with the α -lactalbumin protein sequence from camel species (UniprotBK accession number P00710). Figure 3A shows representative total ion current chromatograms of peptides arising from the tryptic digestion of the SDS-PAGE band corresponding to the α -lactalbumin protein from camel (upper panel) and llama (lower panel) milk, respectively. Significant differences were found among samples (Fig. 3A, arrows). Figure 3B shows the CID mass spectra of peptides from llama α -lactalbumin corresponding to the sequence IWCL/IDNENLQSR (corresponding to amino acid residues 59–70 in α -lactalbumin from camel milk) and FLDDLTDDL/IMCAK (corresponding to amino acid residues 79–93 in α -lactalbumin from camel milk). The peptide IWCLDNENLQSR was seen twice, at $m/z = 774.36664$ and 745.85468 corresponding to experimental $[M + 2H]^{2+}$ with and without carbamidomethylation of the cysteine residue (theoretical $[M + 2H]^{2+} = 774.36755$ and 745.85705 , respectively). Two peptides corresponding to the amino acid sequence FLDDLTDDLIMCAK were also observed, one with carbamidomethylation of the cysteine residue (observed $[M + 2H]^{2+} = 836.36524$; theoretical $[M + 2H]^{2+} = 836.36602$) and one with carbamidomethylation of the cysteine residue plus an oxidized methionine residue (observed $[M + 2H]^{2+} = 844.36310$; theoretical $[M + 2H]^{2+} = 844.36302$).

Thus, our de novo approach allowed us to identify the mutations R62L (or the isobaric residue I) and K89L (or I) in the protein sequence for llama α -lactalbumin relative to its camel counterpart (Fig. 4). Regarding the molecular mass of llama

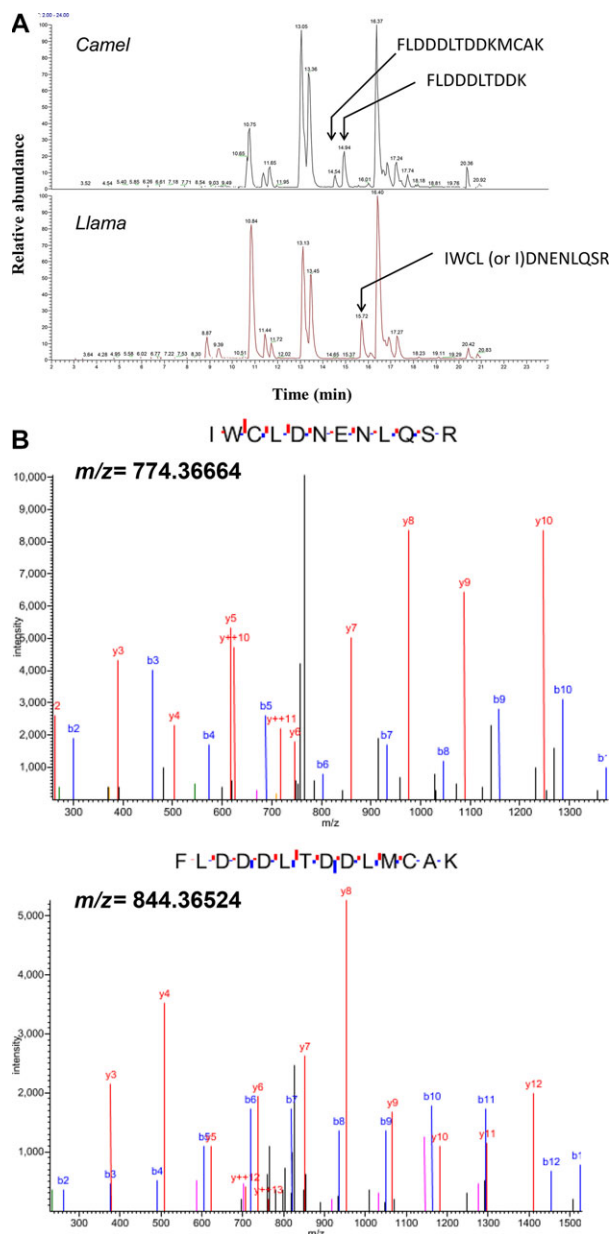


Figure 3. De novo sequencing of the α -lactalbumin protein from llama milk by LC-MS/MS. (A) Total ion current chromatograms of peptides arising from the tryptic digestion of the α -lactalbumin protein from camel and llama milk. Peptide sequences corresponding to the fragment ion detected in each mass spectrometer scan (y-axis) are indicated. (B) CID tandem mass spectra of peptides IWCLDNENLQSR (observed $[M + 2H]^{2+} = 774.36664$, with carbamidomethylation of the cysteine residue; theoretical $[M + 2H]^{2+} = 774.36755$) and FLDDDLTDDLMCAK (observed $[M + 2H]^{2+} = 836.36524$, with carbamidomethylation of the cysteine residue; theoretical $[M + 2H]^{2+} = 836.36602$) arising from the tryptic digestion of the α -lactalbumin protein from llama milk. The sequence of the peptide is shown with the predicted y and b fragments.

α -lactalbumin, these two amino acid mutations resulted in a loss of exactly 58 Da compared with the molecular mass of its camel counterpart. On the other hand, previous studies reporting the N-terminal sequencing of the first 23 amino acid

residues of α -lactalbumin from llama milk showed one amino acid exchange at position 14 when compared to the known sequence of its camel counterpart [5]. Likewise, this amino acid exchange between the camel and llama α -lactalbumin protein sequences is associated with an additional loss of 58 Da. Therefore, the three amino acid substitutions in the α -lactalbumin from llama milk relative to its camel counterpart, namely D14G, R62L/I, and K89L/I, should be associated with a loss of roughly 116 Da, which fits perfectly with our experimental data showing a molecular mass of around 14 430 and 14 314 Da for camel and llama α -lactalbumins, respectively (Tables 2 and 4). Because Leu and Ile residues are isobaric, MS analysis did not allow the distinction between these two residues. However, the multiple alignment of the α -lactalbumin protein from 13 mammalian species (Fig. 4) shows the occurrence of an isoleucyl residue at position 89 in most species, including cattle and humans. Then, it may be hypothesized that an Ile residue is preferred at this position. In addition, a partial amino acid sequence available for α -lactalbumin from guanaco milk (GenBank accession number CAA07693) was compatible with our MS results. However, based on these amino acid sequence data, the GRAVY value for camel and llama α -lactalbumin can be calculated as -0.678 and -0.523 by the ProtParam tool hosted at the Expasy website (<http://web.expasy.org/protparam/>). Thus, it can be assumed that llama α -lactalbumin is more hydrophobic than its camel counterpart [13], which is in good accordance with our RP-HPLC experiments showing a higher retention time for llama α -lactalbumin than for camel α -lactalbumin (Figs. 1 and 2).

3.5 Identification of major MFGM proteins from llama milk

A representative pattern of llama MFGM proteins resolved in 10% SDS-PAGE gels after Coomassie staining is shown in Fig. 5. The lane corresponding to MFGM proteins from llama milk was cut into 24 bands and then bands were digested by trypsin. Resulting peptides were analyzed by MS/MS. Using this approach, a total of 187 functional groups of proteins were identified in the MFGM from llama milk (Supporting Information Table 2). The calculated FDR was 0.2%. The protein profile observed for llama MFGM proteins was very close to the protein profile observed for dromedary camel MFGM proteins [14]. From the top of the gel, major MFGM proteins from llama milk identified by MS were xanthine oxidase, butyrophilin, and lactadherin (LDH). Remarkably, LDH, a major MFGM protein, appears as two equally expressed polypeptides, as previously described for LDH from camel milk [14, 15].

4 Discussion

Although comprehensive sequences for most of llama (*L. glama*) milk proteins are still missing, major milk proteins were identified in this study by comparing the observed

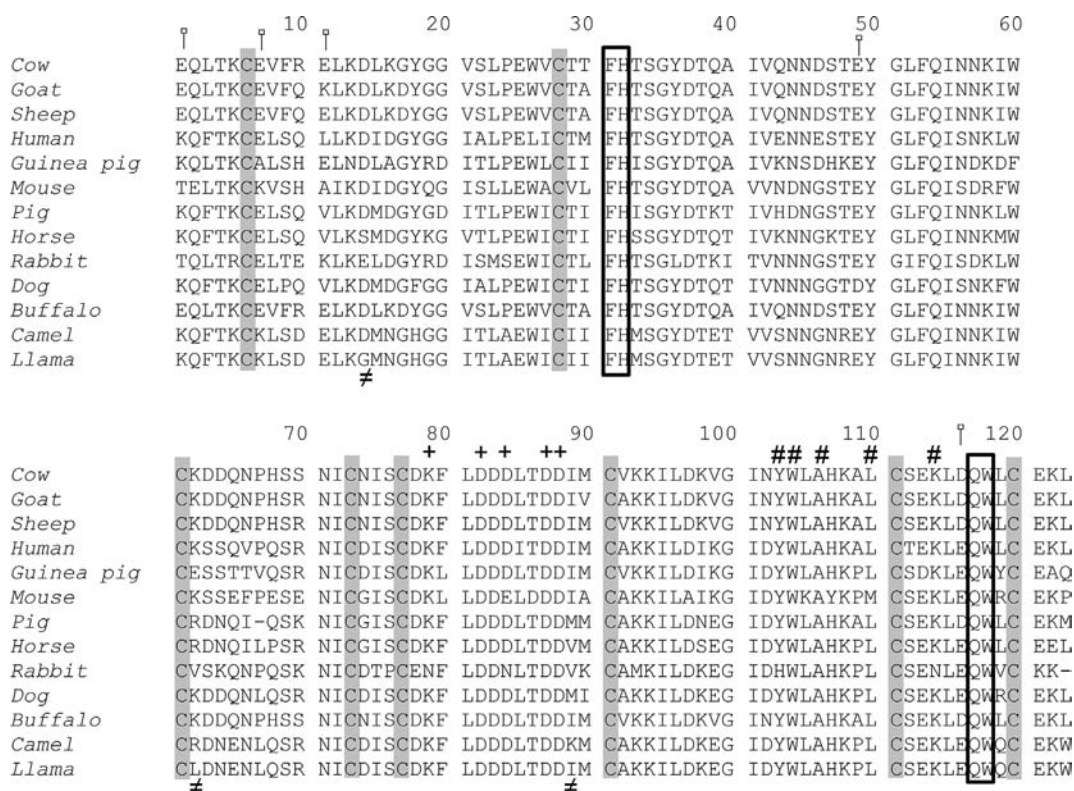


Figure 4. Multiple alignment of the amino acid sequence of α-lactalbumin from cow (P00711), goat (P00712), sheep (P09462), human (P00709), guinea pig (Q9TVD0), mouse (P29752), pig (P18137), horse (P08334), rabbit (P00716), dog (Q9N2G9), water buffalo (Q9TNS6), camel (P00710), and llama (this study) species. Accession numbers are from the UniprotKB database. Amino acid sequences were aligned using the clustal Omega tool hosted at the European Bioinformatics Institute (EBI) web server (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). The signal peptide has been omitted. Conserved cysteine residues are shaded. Amino acid residues located in the aromatic cluster I, which are crucial for α-lactalbumin activity, are boxed. Amino acid residues that are important for the structure of the protein are indicated by a '#' symbol. Calcium-binding sites of α-lactalbumin are indicated by the '+' sign, whereas zinc-binding sites are indicated by the '†' symbol. Amino acid residues that differ in the α-lactalbumin protein from camel and llama milk are highlighted by a '≠' symbol.

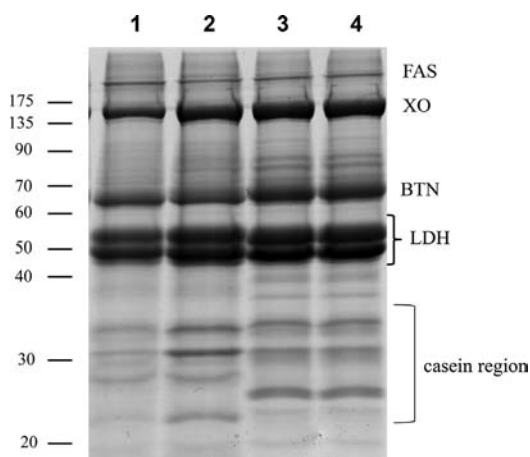


Figure 5. Representative pattern of MFGM proteins in SDS-PAGE. A total of 20 μg MFGM proteins were separated on 10% SDS-PAGE and stained with Bio-Safe (Bio-Rad). MFGM proteins are from camel (lanes 1–2) and llama (lanes 3–4) milk. Positions of protein standards (kDa) are indicated at the left of the panel. Positions of major MFGM proteins identified by MS/MS are indicated at the right of the gel. FAS: fatty acid synthase; XO: xanthine oxidase; BTN: butyrophilin; LDH: lactadherin.

masses of the peptides of unknown proteins to the theoretical peptide masses of each protein encoded in the genome of a closely related species, the dromedary camel (*C. dromedarius*), for which well-annotated sequences of most milk proteins are already available [9]. Using this approach, most of llama milk proteins including the four caseins and some whey proteins (namely α-lactalbumin, lactoferrin, and serum albumin) can be identified. β-Lactoglobulin, the major whey protein found in ruminant milk, was apparently absent in llama milk as previously reported by others [16]. The absence of β-lactoglobulin protein has also been reported in camel milk [17]. Thus, *Camelidae* milk might be a promising substitute for children allergic to cow milk proteins, as β-lactoglobulin is a highly immunogenic protein [18]. Also, we were not able to identify the whey acidic protein, a minor whey protein found in camel milk [19], probably because of its low concentration relative to the other whey proteins. We cannot therefore conclude on the presence of a whey acidic protein related protein in llama milk. The PGRP, an immune-defense protein whose expression has been previously reported in the lactating mammary gland from camel [20], was easily identified by RP-HPLC coupled with MS in camel, but not in llama milk. However, a

PGRP-related protein was identified by LC-MS/MS in llama milk. This protein is probably present in lower concentrations in llama milk than in camel milk. Regarding MFGM proteins, protein profiles observed for extracted proteins from llama milk were very close to those obtained for MFGM proteins extracted from camel milk. Particularly, LDH, a major MFGM protein and highly polymorphic protein [15], appears as two equally expressed polypeptides, as previously described for LDH from camel milk [14]. On the other hand, LDH from goat milk appears as a single polypeptide chain, whereas bovine LDH is expressed as two proteins corresponding to glycosylation variants of a single polypeptide chain [10, 21]. The biological significance of the LDH polymorphism in the mammary gland is currently unknown [14].

4.1 Camel and llama casein fractions: a comparative analysis

The MALDI-TOF PMF approach allowed us to undoubtedly identify skimmed milk proteins along the chromatographic run. Based on these data, molecular masses were attributed to most of milk proteins by LC-ESI-MS analyses.

Molecular masses for the long and short isoforms of a α_{s1} -casein protein (variant A) from llama milk were determined as 25 585.23 and 24 565.99 Da, respectively. Since the out-spliced exonic sequence contains no phosphorylation sites, the molecular mass of the peptide encoded by the skipped exon can be calculated as roughly 1037 Da. This is very close to the molecular mass of nonapeptide EQAYFHLE encoded by the out-spliced exon in α_{s1} -casein (1036.10 Da) from camel milk, thus suggesting a high level of homology between skipped exons in the two species. Remarkably, based on the identification of major chromatographic peaks by LC-ESI-MS, it can be noticed that the short α_{s1} -casein isoform is the main expressed protein in the camel, whereas long α_{s1} -casein isoforms are prevailing isoforms of α_{s1} -casein in llama milk.

Observed molecular masses of β -caseins were comparable in the two species (24 970.88 and 24 971.03 Da in the llama and camel species, respectively), thus suggesting a high level of sequence homology. The most intense ion as detected by LC-ESI-MS of camel skimmed milk proteins was attributed to a β -casein isoform displaying four phosphate residues. This is not in agreement with previous studies by Kappeler et al., who report that the most frequent form of β -casein had only three phosphate groups bound instead of the four groups predicted from the amino acid sequence [9]. However, the molecular mass of the most intense ion attributed to a β -casein isoform in one llama milk sample was 24 970.88, which may correspond to an isoform displaying four phosphate groups, as evidenced in the camel species. Indeed, molecular masses for other isoforms were 24 890.91 Da (–1 phosphate group), 24 811.48 Da (–2 phosphate groups), and 24 731.05 Da (–3 phosphate groups). A molecular mass (24 651.35 Da) corresponding to a putative β -casein protein bearing no phosphate group was also characterized in llama milk. Additionally, molecular masses corresponding to cryptic splice variants resulting from the deletion of a nucleotide

triplet encoding a glutamine residue (–128 Da) were also characterized in camel and llama milk. This phenomenon, which may result from errors of the splicing machinery, was already evidenced in equine β -casein [22]. Accordingly, in the camel species, the presence of a glutamine residue at the beginning of exon 6 of the β -casein gene supports this hypothesis [23].

The most striking difference in size between the llama and camel species was observed for α_{s2} -casein. A mean molecular mass of 22 906.10 Da was attributed to the main isoform of α_{s2} -casein in three of the four llama milk samples analyzed, whereas a molecular mass of 22 065.86 Da was observed for its camel counterpart, the isoform with ten phosphate groups being the main isoform as detected by LC-ESI-MS analysis of camel milk. In the camel species, based on amino acid sequence analysis of α_{s2} -casein, it is assumed that only nine motifs for posttranslational phosphorylation on serine residue can occur [9]. However, phosphorylation of caseins can take place on threonine residues as well (in tripeptide sequences Thr-X-A, where X represents any amino acid residue and A represents an acidic residue) [24]. Accordingly, a consistent probability for phosphorylation at Thr₃₉ ($p = 0.726$), Thr₁₁₄ ($p = 0.949$), Thr₁₂₆ ($p = 0.534$), and Thr₁₃₂ ($p = 0.932$) occurs in the sequence for α_{s2} -casein (UniprotKB accession number O97944) when using the predictive software NetPhos 2.0 server at the ExPASy Bioinformatics Resource Portal (<http://www.cbs.dtu.dk/services/NetPhos/>).

Multiple alignment of the amino acid sequence of α_{s2} -casein from ten mammalian species [23] reveals that several exonic sequences are constitutively spliced in the camel (exons 8, 10, and 13). Since then, it can be hypothesized that the size difference for α_{s2} -casein from llama (ca. 1000 Da; Table 3) corresponds to an exonic sequence that is spliced out in the camel, but not in the llama species. However, additional experiments involving genomic tools are mandatory to conclude precisely about the splicing mechanisms that occur on the α_{s2} -casein encoding gene in the llama species.

4.2 Amino acid substitutions in the α -lactalbumin protein between the camel and llama species: implications for structure to activity relationships

The milk whey protein, α -lactalbumin, is the regulatory component of the lactose synthase enzymatic complex, which synthesizes and regulates the biosynthesis of lactose in milk. α -Lactalbumin, whose expression is restricted to the mammary gland, binds reversibly with the catalytic component of the lactose synthase complex triggering a subsequent 1000-fold decrease in the K_m for glucose, thereby allowing efficient production of lactose under physiological conditions. Site-directed mutagenesis studies have highlighted the role for structural components in the bovine α -lactalbumin protein. Substitutions for Q117 or W118 in the aromatic cluster I of bovine α -lactalbumin were found to reduce the affinity for the catalytic component of the lactose synthase complex, whereas changes at F31 or H32 have major effects on the ability to promote glucose binding. Other substitutions located in the

aromatic cluster II were found to affect folding rather than activity [25, 26]. Calcium-binding sites as well as other divalent cations binding sites were also identified in α -lactalbumin protein [27]. The multiple alignment of α -lactalbumin protein from 13 mammalian species (Fig. 4) shows that these amino acid residues are highly conserved across species, thus suggesting a similar role for them in the regulation of lactose synthesis by *Camelidae* α -lactalbumins. Nevertheless, amino acid substitutions that occur between the camel and the llama α -lactalbumin proteins are not located within any of these sites. Because significant differences with regard to lactose levels in camel and llama milk have been already reported [4], additional experiments are needed to decipher the impact of amino acid mutations reported here in the llama α -lactalbumin with regard to its activity.

4.3 Concluding remarks

While camel milk has been largely investigated, only little attention has been paid to milk proteins from another camelid, the llama (*L. glama*). We provide here the first characterization of llama milk proteins by combining different proteomic approaches, including peptide mass matching, LC-ESI-MS, and de novo sequencing. We achieved the identification of the four caseins in llama milk as well as some whey proteins and we provided experimental molecular masses for most of milk proteins identified in *Camelidae* milk. In addition, a full-length amino acid sequence can be reconstructed for llama α -lactalbumin based on experimental data gained from MALDI-TOF, LC-ESI-MS, and de novo sequencing by MS/MS. We hope the data provided here will help the annotation of llama milk proteins encoding genes. We also think that our results can be useful for understanding biodiversity as well as for gaining insights into the biology and evolution of milk proteins.

The authors warmly acknowledge Dr. André Richard and Mrs. Sophie Vialis for providing llama milk samples. They are also grateful to the Experimental station of the Arid Land Institute of Medenine (Medenine, Tunisia) and the INRA Experimental farm "Domaine de Galle" (F-18390, Osmoy) for providing camel and goat milk samples, respectively.

The authors have declared no conflict of interest.

5 References

- [1] Kappeler, S., 1998, <http://dx.doi.org/10.3929/ethz-a-002025594>.
- [2] Morin, D. E., Rowan, L. L., Hurley, W. L., Braselton, W. E., *J. Dairy Sci.* 1995, 78, 1713–1720.
- [3] Riek, A., Gerken, M., *J. Dairy Sci.* 2006, 89, 3484–3493.
- [4] Medhammar, E., Wijesinha-Bettoni, R., Stadlmayr, B., Nilsson, E., Charrondiere, U. R., Burlingame, B., *J. Sci. Food Agric.* 2012, 92, 445–474.
- [5] Cantisani, A., Napolitano, L., Giuffrida, M. G., Conti, A., *J. Biochem. Biophys. Methods* 1990, 21, 227–236.
- [6] Fernandez, F. M., Oliver, G., *Small Rumin. Res.* 1992, 8, 97–106.
- [7] Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., Ward, L. D., Lowe, C. B., Holloway, A. K., Clamp, M., Gnerre, S., Alföldi, J., Beal, K., Chang, J., Clawson, H., Cuff, J., Di Palma, F., Fitzgerald, S., Flicek, P., Guttman, M., Hubisz, M. J., Jaffe, D. B., Jungreis, I., Kent, W. J., Kostka, D., Lara, M., Martins, A. L., Massingham, T., Moltke, I., Raney, B. J., Rasmussen, M. D., Robinson, J., Stark, A., Vilella, A. J., Wen, J., Xie, X., Zody, M. C., Worley, K. C., Kovar, C. L., Muzny, D. M., Gibbs, R. A., Warren, W. C., Mardis, E. R., Weinstock, G. M., Wilson, R. K., Birney, E., Margulies, E. H., Herrero, J., Green, E. D., Haussler, D., Siepel, A., Goldman, N., Pollard, K. S., Pedersen, J. S., Lander, E. S., Kellis, M., *Nature* 2011, 478, 476–482.
- [8] Martin, P., Leroux, C., *Biochem. Biophys. Res. Commun.* 1992, 183, 750–757.
- [9] Kappeler, S., Farah, Z., Puhani, Z., *J. Dairy Res.* 1998, 65, 209–222.
- [10] Cebo, C., Caillat, H., Bouvier, F., Martin, P., *J. Dairy Sci.* 2010, 93, 868–876.
- [11] Aschaffenburg, R., Drewry, J., *XVth International Dairy Congress*, International Dairy Federation, London, UK 1959, pp. 1631–1637.
- [12] Miranda, G., Mahé, M., Leroux, C., Martin, P., *Proteomics* 2004, 4, 2496–2509.
- [13] Kyte, J., Doolittle, R. F., *J. Mol. Biol.* 1982, 157, 105–132.
- [14] Saadaoui, B., Henry, C., Khorchani, T., Mars, M., Martin, P., Cebo, C., *Proteomics* 2013, 13, 1180–1184.
- [15] Cebo, C., Martin, P., *Int. Dairy J.* 2012, 24, 70–77.
- [16] Morin, D. E., Rowan, L. L., Hurley, W. L., *Small Rumin. Res.* 1995, 17, 255–261.
- [17] Kappeler, S. R., Farah, Z., Puhani, Z., *J. Dairy Sci.* 2003, 86, 498–508.
- [18] Ehlayel, M. S., Hazeima, K. A., Al-Mesaifri, F., Bener, A., *Allergy Asthma Proc.* 2011, 32, 255–258.
- [19] Beg, O. U., von Bahr-Lindstrom, H., Zaidi, Z. H., JÖRnvall, H., *Eur. J. Biochem.* 1986, 159, 195–201.
- [20] Kappeler, S. R., Heuberger, C., Farah, Z., Puhani, Z., *J. Dairy Sci.* 2004, 87, 2660–2668.
- [21] Hvarregaard, J., Andersen Lars Berglund, M. H., Rasmussen, J. T., Petersen, T. E., *Eur. J. Biochem.* 1996, 240, 628–636.
- [22] Matéos, A., Girardet, J., Mollé, D., Corbier, C., Gaillard, J. L., Miclo, L., *Rapid Commun. Mass Spectrom.* 2010, 24, 1533–1542.
- [23] Martin, P., Cebo, C., Miranda, G., in: McSweeney, P. L. H., Fox, P. F. (Eds.), *Advanced Dairy Chemistry*, Springer, New York 2013, pp. 387–429.
- [24] Mercier, J. C., *Biochimie* 1981, 63, 1–17.
- [25] Malinovsky, V. A., Tian, J., Grobler, J. A., Brew, K., *Biochemistry* 1996, 35, 9710–9715.
- [26] Grobler, J. A., Wang, M., Pike, A. C., Brew, K., *J. Biol. Chem.* 1994, 269, 5106–5114.
- [27] Permyakov, E., Berliner, L., *FEBS Lett.* 2000, 473, 269–273.