DATASET BRIEF

A proteomic characterization of water buffalo milk fractions describing PTM of major species and the identification of minor components involved in nutrient delivery and defense against pathogens

Chiara D'Ambrosio, Simona Arena, Anna Maria Salzano*, Giovanni Renzone*, Luigi Ledda and Andrea Scaloni*

Proteomics & Mass Spectrometry Laboratory, ISPAAM, National Research Council, Naples, Italy

Received: December 13, 2007
Revised: April 14, 2008
Accepted: May 13, 2008

Water buffalo has been studied in relation to the exclusive use of its milk for the manufacture of high-quality dairy products. Buffalo milk presents physicochemical features different from that of other ruminant species, such as a higher content of fatty acids and proteins. We report here a detailed proteomic analysis of buffalo skim milk, whey and milk fat globule membrane fractions. Notwithstanding the poor information available on buffalo genome, identification of protein isoforms corresponding to 72 genes was achieved by a combined approach based on 2-DE/MALDI-TOF PMF and 1-DE/μLC-ESI-IT-MS-MS. Major protein components, *i.e.* α_{S1}-, α_{S2}-, β-, κ-caseins, α-lactalbumin and β-lactoglobulin, were characterized for PTM, providing a scientific basis to coagulation/cheese making processes used in dairy productions. Minor proteins detected emphasized the multiple functions of milk, which besides affording nutrition to the newborn through its major components, also promotes development and digestive tract protection in the neonate, and ensures optimal mammary gland function in the mother. Defense against pathogens is guaranteed by an arsenal of antimicrobial/immunomodulatory proteins, which are directly released in milk or occur on the surface of secreted milk-lipid droplets. Proteins associated with cell signaling or membrane/protein trafficking functions were also identified, providing putative insights into major secretory pathways in mammary epithelial cells.

Keywords:
Glycosylation / Milk-fat globule membrane / Milk proteins / Phosphorylation / Water buffalo

Milk is the most important food for young mammals and a common source of proteins and microelements for adult people. Its main polypeptide components, namely α_{S1}-casein (α_{S1}-CN), α_{S2}-casein (α_{S2}-CN), β-casein (β-CN), κ-casein (κ-CN), β-lactoglobulin (β-LGB) and α-lactalbumin B (α-LAB), have specific biological functions that have already been reviewed [1]. The in-depth knowledge of its chemistry has been used by the dairy manufacturing industry to develop/optimize modern technologies required to produce different milk products [2]. In fact, thorough information on the physicochemical properties of its numerous components has been essential to design processing equipment and conditions required for the manufacture/distribution of dairy products. Knowledge of milk protein components chemistry

Correspondence: Dr. Andrea Scaloni, Proteomics and Mass Spectrometry Laboratory, ISPAAM, National Research Council, via Argine 1085, 80147 Naples, Italy
E-mail: andrea.scaloni@ispaam.cnr.it
Fax: +39-81-596-5291

Abbreviations: **α-LAB**, α-lactalbumin; **β-LGB**, β-lactoglobulin; **BTN**, butyrophilin; **CD**, cluster of differentiation; **CN**, casein; **MFGM**, milk fat globule membrane; **TLR**, Toll-like receptor

* These authors contributed equally to this work.

has been also indispensable for detection of adulteration in raw materials used for production of high-quality products [3]. For these reasons and for the complexity of this biological matrix, powerful analytical techniques such as 2-DE and MS have found a widespread use as key tools to investigate dairy products [4–7].

Of the main buffalo species, the Italian one *(Bubalus bubalis)* has been widely studied in relation to the exclusive use of its milk for the manufacture of high-quality dairy products, among which Mozzarella cheese with Protected Designation of Origin mark is the most famous in the world. In 2006, this cheese and other similar products have been associated to a global market of 500 million and 2–3 billion dollars, respectively. Water buffalo milk presents physicochemical features different from that of closely related ruminant species, such as a higher content of fat and proteins. Since milk's appropriateness for cheese making is strictly dependent on both its casein composition and structure [8, 9], particularly on κ-CN content and protein phosphorylation, definitive data on buffalo milk proteome might help in understanding milk behavior at clotting and how Ca^{2+}-sensitive proteins aggregate through calcium phosphate bridges into micelles. Moreover, proteomic data could have important applications in the development of novel assays directed to the detection of adulterations due to occurrence of bovine material into buffalo milk.

Some proteomic studies on milk fractions have been reported for cow, sheep and horse [5–7, 10–14]. Despite descriptive data on the composition of its main constituents by chromatographic/electrophoretic approaches [15] and sequence information on α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, β-LGB and α-LAB genes (Swiss-Prot code O62823, O62825, Q9TSI0, P11840, P02755 and Q9TSN6), poor data however are available on the complete structural characterization of the buffalo milk proteins [15, 16]. MS studies reported in literature focused only on PTM analysis of α_{S1}- and β-CN [15]. Since proteomic analysis of bovine milk fractions was also proved effective in describing less abundant host defense proteins protecting neonate/mammary gland against infection or diagnostic markers of mastitis in affected animals [10, 11, 13], an extended use of proteomic technologies to description of buffalo milk protein components should be strongly encouraged in view of its potential impact on food chemistry and animal welfare applications.

To identify proteins present in buffalo milk, characterizing its distribution in skim milk, whey and milk fat globule membrane (MFGM) portions, samples from five animals were separated into these different fractions (see Supporting Information), which were subjected to further proteomic analysis. Figure 1 shows the proteomic map of water buffalo skimmed milk within a p*I* range of 4–7, obtained using a colloidal CBB G250 (panel A) or ammoniacal silver (panel B) staining. Globally, almost 200 protein spots were resolved on the gels of which 99 were identified by PMF experiments (Supporting Information Table S1). No significant gel differences were observed between animal individuals, suggesting the absence of genetic variants. These findings were in agreement with the reduced number of polymorphisms reported in literature for water buffalo milk proteins [17]. As expected, gels were dominated by six trains of spots identified as α_{S1}-CN (spot 8-13), α_{S2}-CN (spot 24-28), β-CN (spot 18-23), κ-CN A (spot 17, 34-41, 63, 65-67), β-LGB (spot 49-52, 80, 81) and α-LAB B (spots 47, 48, 53, 54, 83). Minor amounts of casein degradation products, important for the animal immunological response [11, 18], were also observed. MS data proved spot nature as the expression products of buffalo genes mentioned above and excluded the occurrence of other genetic variants. Globally, spots related to α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, β-LGB and α-LAB accounted for 24.8, 11.9, 24.9, 10.8, 7.7 and 4.8% of the total gel absorbance, respectively. These values are quantitatively different from those reported for bovine/ovine milk [19]. In addition, proteins associated with transport (albumin, mammary-derived growth inhibitor), structure (neurofilament triplet M protein) and defense/immunity (polymeric immunoglobulin receptor) functions were also identified (Table 1). The latter protein binds polymeric IgA and IgM at the basolateral surface of epithelial cells.

Heterogeneity in α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, β-LGB and α-LAB electrophoretic forms represented the occurrence of PTM described in previous studies [5, 7, 10, 12]. Based on the very high sequence homology between ruminant milk protein sequences (Supporting Information Figs. S1–S3) and previous reports in literature, a series of technological procedures were successfully applied to the complete characterization of the most abundant buffalo proteins, *i.e.* α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, β-LGB and α-LAB, including PTM. In particular, using of TiO_2-based microcolumns for selective enrichment of phosphopeptides [20], resolving glycopeptides by dedicated micro-fractionation procedures [21], performing MALDI-TOF-MS experiments with the instrument in linear modality and used either in positive and negative polarity [5, 7, 10, 12], and parallel performing MALDI-TOF-MS experiments on samples subjected to treatment with alkaline phosphatase or PNGase F [7, 21] proved particularly effective in PTM description. Since Cys residues within buffalo α_{S2}-CN, κ-CN, β-LGB and α-LAB sequence were totally conserved with respect to other ruminants (see Supporting Information Figs. S1–S3), no analyses on disulfide arrangements were performed.

After trapping phosphopeptides from α_{S1}-CN tryptic digests by TiO_2 resin, MALDI-TOF-MS signals changing their relative intensity between samples allowed the identification of peptides with different degrees of phosphorylation among spots 9-11 (Supporting Information Figs. S4A–C and Table S2). Combining all spectra recorded, the sequence coverage was 94.5%. These data proved that phosphorylation occurred at Ser41, 46, 48, 64, 66, 67, 68 and 75, confirming previous observations on α_{S1}-CN (Supporting Information Fig. S1) [1, 7, 15, 20, 22, 23]. Thus, we concluded that spot 9, 10 and 11 corresponded to 199 residues long α_{S1}-CN forms, each having eight, seven and six phosphate groups, respec-

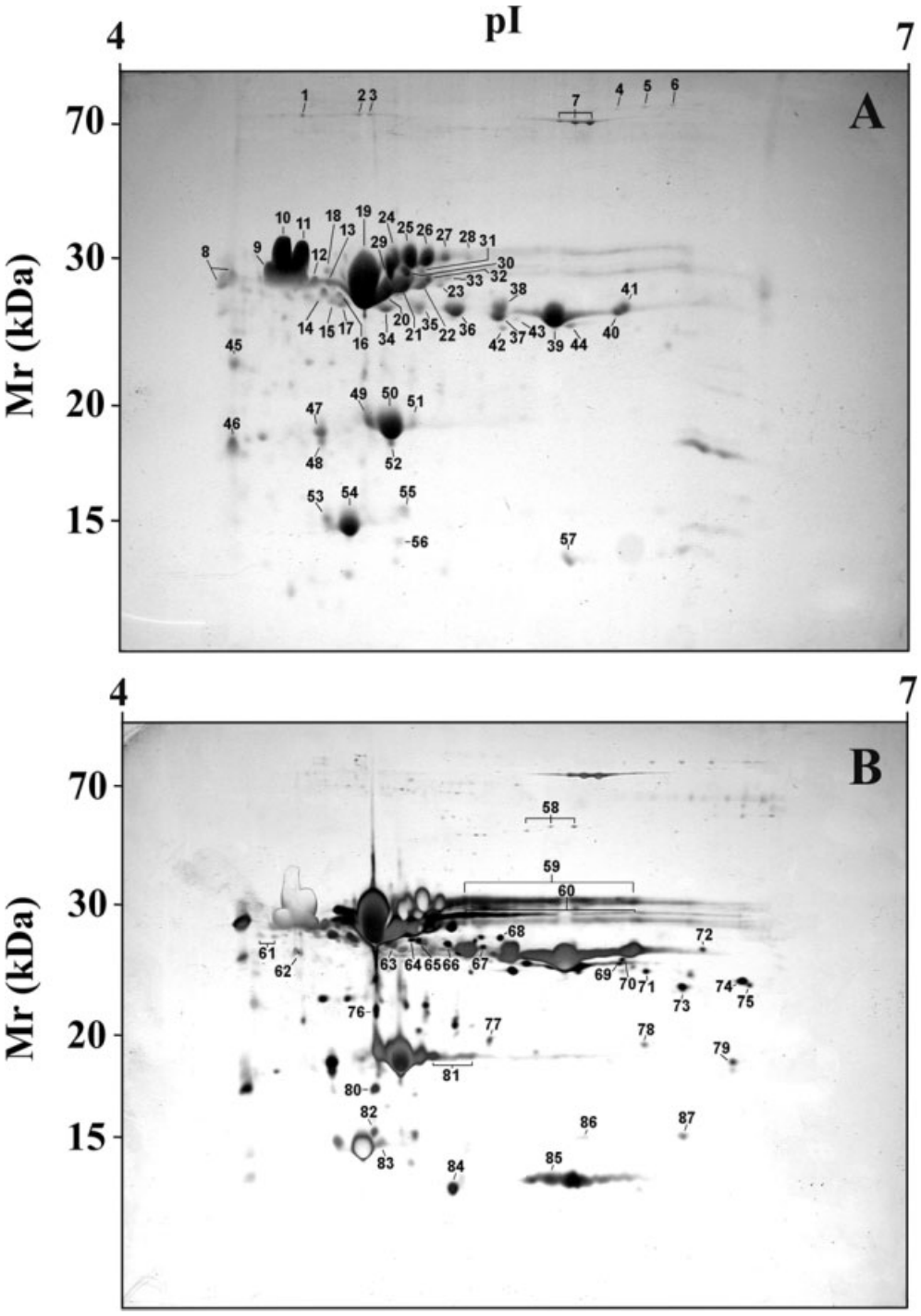

Figure 1. 2-DE of buffalo skim milk proteins. Milk proteins (500 μg) were separated in the first dimension using 18-cm pH 4-7 IPG strips and in the second dimension using 14% SDS-PAGE. Proteins were visualized by colloidal CBB G250 (panel A) or ammoniacal silver (panel B) staining. Identified protein spots are numbered and reported in Supporting Information Table S1.

tively. As result of the lacking of a series of phosphorylation sites otherwise present in bovine and ovine protein, buffalo α_{S1}-CN showed a reduced phosphorylation degree with respect to other ruminants [1].

Supporting Information Fig. S5 and Table S3 describe the MALDI-TOF mass spectra of the enriched tryptic digests from buffalo α_{S2}-CN isoforms (spots 24-27). Combining all spectra, the sequence coverage was 95.2% and information

Table 1. Buffalo milk proteins identified by 2-DE plus MALDI-TOF PMF or 1-DE plus μLC-ESI-IT-MS-MS analysis[a)]

Protein	Accession number	Organism	Identification method	Milk fraction	Localization	Function
Actin	P60712	*B. taurus*	PMF, MS-MS	W, MFGM	C	S
Aldose 1-epimerase	Q8K157	*M. musculus*	MS-MS	MFGM	C	E
Adipophilin, adipocyte differentiation-related protein	Q9TUM6	*B. taurus*	PMF, MS-MS	W, MFGM	M	FTM
ADP-ribosylation factor (Arf)-like protein 15, Arf-related protein 2	Q5EA19	*B. taurus*	PMF	MFGM	M	MPT
Albumin	P02769	*B. taurus*	PMF, MS-MS	S, W	S	FTM
BRCA1-associated protein BRAP2	Q99MP8	*M. musculus*	MS-MS	MFGM	C	CS
Breast cancer type 1 susceptibility protein BRCA1	Q864U1	*B. taurus*	MS-MS	MFGM	S	CS
Butyrophilin subfamily 1 member A1	P18892	*B. taurus*	PMF, MS-MS	MFGM	M	FTM
Calcium-binding protein p22	Q99653	*H. sapiens*	PMF	MFGM	M	MPT
Cargo selection protein, TIP47	Q3SX32	*B. taurus*	MS-MS	MFGM	M	MPT
α_{S1}-Casein	O62823	*B. bubalis*	PMF, MS-MS	S, W, MFGM	S	GT
α_{S2}-Casein	Q3Y443	*B. bubalis*	PMF, MS-MS	S, W	S	GT
β-Casein	Q9TSI0	*B. bubalis*	PMF	S, W	S	GT
κ-Casein	P11840	*B. bubalis*	PMF, MS-MS	S, W, MFGM	S	GT
Cell death activator CIDE-A	A4FUX1	*B. taurus*	MS-MS	MFGM	C	CS
Coiled-coil domain-containing protein 116	Q4V8B5	*R. norvegicus*	MS-MS	MFGM	C	U
Complement C3	Q2UVX4	*B. taurus*	MS-MS	W	S	DI
Cyclophilin A, peptidyl-prolyl cis-trans isomerase	Q861V5	*B. taurus*	PMF	MFGM	C	PSBF
Cysteine sulfinic acid decarboxylase	Q9Y600	*H. sapiens*	MS-MS	MFGM	C	E
Dymeclin	Q8CHY3	*M. musculus*	MS-MS	MFGM	M	U
EH-domain containing protein 4	Q1MWP8	*M. musculus*	MS-MS	MFGM	M, S	U
Eukaryotic translation initiation factor 5A-1*	Q6EWQ7	*B. taurus*	PMF	MFGM	C	PSBF
Factor XIIa inhibitor	P50448	*B. taurus*	MS-MS	W	S	DI
Fatty acid synthase	Q71SP7	*B. taurus*	MS-MS	MFGM	C	FTM
Folate-binding protein	P02702	*B. taurus*	MS-MS	W	M, S	FTM
β-1,4-Galactosyltransferase 1	P08037	*B. taurus*	MS-MS	W	M, S	E
Growth/differentiation factor 8, myostatin	O18836	*B. taurus*	MS-MS	W	S	CS
Guanine nucleotide-binding protein Gi, subunit α1	P63096	*H. sapiens*	MS-MS	MFGM	M	CS
Guanine nucleotide-binding protein Gl/GS/GT subunit β1	P62871	*B. taurus*	PMF, MS-MS	MFGM	M	CS
Guanine nucleotide-binding protein Gl/GS/GT subunit β2	P11017	*B. taurus*	PMF, MS-MS	MFGM	M	CS
IgM heavy chain	Q1RMK2	*B. taurus*	PMF, MS-MS	W, MFGM	S	DI
Ig λ-like polypeptide	Q2KIF5	*B. taurus*	MS-MS	W	S	DI
Ig λ-like polypeptide	Q1RMN8	*B. taurus*	MS-MS	W	S	DI
Inter-α-trypsin inhibitor heavy chain H4	Q3T052	*B. taurus*	MS-MS	W	S	PSBF, DI
Keratin intermediate filament 16a	Q9EQD6	*M. musculus*	MS-MS	MFGM	C	S
Lactadherin, PAS6/7	Q95114	*B. taurus*	PMF, MS-MS	W, MFGM	M	FTM
α-Lactalbumin	Q9TSN6	*B. bubalis*	PMF, MS-MS	S, W	S	E
β-Lactoglobulin	P02755	*B. bubalis*	PMF, MS-MS	S, W, MFGM	S	GT
Lactoperoxidase	A5JUY9	*B. bubalis*	MS-MS	W	S	E, DI
Lactophorin	P80195	*B. taurus*	PMF, MS-MS	S, W, MFGM	S	U, DI
Lactotransferrin	O77698	*B. bubalis*	MS-MS	W	S	DI
Leucine-rich α2-glycoprotein 1	Q2KIF2	*B. taurus*	MS-MS	W	S	U
α2-Macroglobulin MUG1	P28665	*M. musculus*	MS-MS	MFGM	S	DI
Mammary-derived growth inhibitor, fatty acid-binding protein	Q5XLB1	*B. bubalis*	PMF, MS-MS	S, W, MFGM	S	FTM
Monocyte differentiation antigen CD14	Q1PBD0	*B. taurus*	MS-MS	W, MFGM	M, S	DI
Mucin 1, PAS1	Q8WML4	*B. taurus*	MS-MS	MFGM	M	U
Mucin 15, PAS3	Q8MI01	*B. taurus*	MS-MS	MFGM	M	U
Multidrug transporter breast cancer resistance protein BCRP/ABCG2, ATP-binding cassette	Q32PJ1	*B. taurus*	MS-MS	MFGM	M	GT
Neurofilament triplet M protein	O77788	*B. taurus*	PMF	S	C	S
Nucleobindin 1	Q02819	*M. musculus*	PMF	W	M, S	DI
Nucleobindin 2	Q0IIH5	*M. musculus*	PMF	W	M, S	DI
Pancreatic elastase inhibitor	Q9TS74	*B. taurus*	MS-MS	W	S	PSBF
Pigment epithelium-derived factor	Q95121	*B. taurus*	MS-MS	W	S	PSBF

Table 1. Continued

Protein	Accession number	Organism	Identification method	Milk fraction	Localization	Function
Platelet glycoprotein 4, PAS4, CD36	P26201	*B. taurus*	MS-MS	W, MFGM	M	U, FTM
Polymeric immunoglobulin receptor	P81265	*B. taurus*	PMF, MS-MS	S, W	M, S	DI
Ras-related protein Rab1A	P62820	*H. sapiens*	MS-MS	MFGM	M	MPT
Ras-related protein Rab3A	P11023	*B. taurus*	MS-MS	MFGM	M	MPT
Ras-related protein Rab15	P59190	*H. sapiens*	MS-MS	MFGM	M	MPT
Ras-related protein Rab18	Q9NP72	*H. sapiens*	PMF, MS-MS	MFGM	M	MPT
Responsive to centrifugal force and shear	Q6QRN7	*B. taurus*	MS-MS	MFGM	M	U
RING finger protein 10	Q08E13	*B. taurus*	MS-MS	MFGM	C	U
Signal transducer and activator of transcription 5B, STAT5	Q9TUM3	*B. taurus*	MS-MS	MFGM	C	CS
Sodium-dependent phosphate transport protein 2B	Q27960	*B. taurus*	MS-MS	MFGM	M	GT
α-Soluble NSF attachment protein	P81125	*B. taurus*	PMF	MFGM	M	MPT
Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	Q15738	*H. sapiens*	MS-MS	MFGM	M	FTM
Synaptosomal-associated protein 23, SNAP23	O00161	*H. sapiens*	MS-MS	MFGM	M	MPT
Synaptobrevin 3, Vesicle-associated membrane protein 3	Q15836	*H. sapiens*	MS-MS	MFGM	M	MPT
Syntaxin 3	A6QLH3	*B. taurus*	MS-MS	MFGM	M	MPT
Toll-like receptor 2, TLR2	Q95LA9	*B. taurus*	MS-MS	MFGM	M	DI
Ubiquitin	P62988	*H. sapiens*	MS-MS	MFGM	C	CS
Xanthine dehydrogenase/oxidase	P80457	*B. taurus*	PMF, MS-MS	W, MFGM	M	FTM
Zinc-α2-glycoprotein	Q3ZCH5	*B. taurus*	PMF, MS-MS	W	S	DI

a) Accession number: Swiss-Prot. Milk fraction where the protein was identified is denoted; skim milk (S), whey (W) and milk fat globule membrane (MFGM). Protein localization is indicated as cytoplasmic (C), membrane (M) and secreted (S). Functional categories listed are cell signaling (CS), defense/immunity (DI), enzyme (E), fat transport/metabolism (FTM), general transport (GT), membrane/protein trafficking (MPT), protein synthesis/binding/folding (PSBF), structural (S) and unknown function (U).

on all Ser residues described as phosphorylated in bovine/ovine α_{S2}-CN was obtained (Supporting Information Fig. S1) [1, 7, 20, 22, 23]. PSD experiments on selected phosphopeptides confirmed in some cases that phosphorylation occurred at the same sites of bovine counterpart (Ser129, 131 and 143) (Supporting Information Fig. S6). Based on the very high sequence homology between buffalo and bovine protein, we should hypothesize that phosphorylation occurs also at Ser8, 9, 10, 16, 31, 53, 56, 57, 58 and 61. Spots 24, 25, 26 and 27 corresponded to 207 residues long α_{S2}-CN forms having 13, 12, 11 and 10 phosphate groups, respectively. Since spots 29–31 showed a p*I* value and panel of phosphopeptides similar to those reported for spots 24–26, but a lower *M*r value, they were tentatively associated to truncated forms of α_{S2}-CN. Similar isoforms have been already characterized in bovine and ovine and have been associated to exon skipping caused by a single-base mutation that affects splicing of the pre-mRNA [7].

β-CN isoforms corresponding to spots 19-23 were similarly analyzed by MALDI-TOF-MS (Supporting Information Fig. S7 and Supporting Information Table S4). Recorded data ensured a sequence coverage of 96.2% and confirmed previous observations on phosphorylation sites (Ser15, 17, 18, 19 and 35) in the buffalo [15] and the bovine/ovine proteins [1, 7, 20, 22, 23] (Supporting Information Fig. S2). Thus, we conclude that spot 19, 20, 21, 22 and 23 correspond to 209 residues long β-CN forms each having five, four, three, two and one phosphate groups, respectively. The lack of a putative phosphorylation site within buffalo β-CN B sequence determined a reduced phosphorylation degree of this protein with respect to what observed in other ruminants [1, 7, 20, 22, 23].

Heterogeneity observed for bovine/ovine κ-CN [5, 7, 12] found a good equivalence in buffalo protein. No biotin-based κ-CN enrichment procedures were used and then the number of isoforms detected was reduced with respect to bovine milk [12]. PTM analysis was limited to seven major protein components, namely spots 34–40. Since the known sites of modification in bovine/ovine κ-CN are all located in the C-terminal portion of the protein corresponding to a large tryptic peptide [1, 5, 12] (Supporting Information Fig. S2), using of TiO_2 proved particularly effective in trapping these phosphorylated-sialylated species [20, 24]. Additional experiments on whole digests, providing a sequence coverage of 94.5%, confirmed that no additional PTM occurred in other protein regions (data not shown). Supporting Information Fig. S8 shows the MALDI-TOF-MS analysis of the enriched tryptic digests from spots 34–40; signals associated to modified peptides (113-169) and (112-169), generated as result of missed cleavage after Lys116 and Lys113, were evident for all samples. Spectra increased its complexity moving from 40 to 34 spot, with major peaks shifting to higher mass values and

a number of extra peaks observed. They were associated to peptides with an increased phosphorylation and O-glycosylation degree, which correspond to κ-CN forms with more acidic p*I* values, resulting from the occurrence of an increasing number of phosphate and N-acetylneuraminic acid (NeuAc) units. The complex pattern of peaks reported in Supporting Information Figs. S8D–G originated from the well-known partial loss of carbohydrate groups, particularly loss of NeuAc (Supporting Information Table S5) [5]. However, this phenomenon did not prevent assignment of the parent forms by MS data alone. Our findings were supported by the number of core neutral disaccharides present within the spectrum of each spot and the total number of NeuAc moieties inferred from the distinctive pattern of p*I* shifts on the 2-DE map (Fig. 1A). Accordingly, unmodified (spot 40), phosphorylated (spot 39), diphosphorylated (spot 37), phosphorylated-glycosylated (spot 38), phosphorylated-glycosylated (spot 36), phosphorylated-diglycosylated (spot 35) and phosphorylated-diglycosylated κ-CN (spot 34) were identified (Supporting Information Table S5). Similar κ-CN forms have been reported in bovine milk proteome [1, 5, 12], elucidating the complex pattern of progressive phosphorylation and O-glycosylation and identifying modified residues. Based on the very high sequence homology between buffalo and bovine protein as well as the nature of κ-CN isoforms characterized, we should hypothesize that the modification sites are highly conserved between the two animals (Supporting Information Fig. S2). This hypothesis found a positive confirmation in MS analysis of the chymotryptic digest of peptides (113-169)P/(112-169)P and (113-169)2P/(112-169)2P (Supporting Information Figs. S8B and C). In the first case, the spectrum showed MH^+ signals at *m/z* 3629.9, 3758.3 and 2458.3, associated to aspecific peptides (113-146), (112-146) and (147-169)P. In the second case, the spectrum showed MH^+ signals at *m/z* 3710.2, 3838.5 and 2458.2, assigned to phosphopeptides (113-146)P, (112-146)P and (147-169)P. These results suggest that the main and secondary site of phosphorylation in buffalo κ-CN are Ser149 and Ser127 [12], similar to what is observed for bovine protein.

A detailed characterization of the PTM occurring in α-LAB was obtained by comparative MALDI-TOF-MS analysis of all fractions resulting from chromatographic resolution of the tryptic digest from spot 54 and 47; these spots showed a Δ*M*r of 3–4 kDa (Fig. 1A). In both cases, acquired mass spectra allowed obtaining a sequence coverage of 91.1%. No peptides bearing PTM were identified in the spectra recorded for spot 54; in contrast, the fraction from spot 47 eluting at 24.7 min showed the typical pattern of adjacent signals differing by 162, 203, and 291 mass units, disappearing after PNGase F treatment, which made glycopeptides immediately recognizable (Supporting Information Figs. S9A and B). In fact, signals in the *m/z* range of 4231.2–5011.0 were associated with glycosylated forms of peptide (32-53), generated as result of aspecific cleavage of the expected large tryptic peptide (17-58) at Phe31 and Phe53, as also confirmed by PSD analysis (Supporting Information Fig. S9C). Based on the known biosynthetic pathway of N-linked oligosaccharides and the molecular mass of the peptide moiety, these signals were assigned to fragment (32-53) bearing complex type N-linked glycans bound to Asn45. A mixture of bi- and tri-antennary structures containing also N-acetylgalactosaminylated sugar chains was observed (Supporting Information Fig. S9A). The latter N-linked oligosaccharides have been already reported in glycosylated bovine α-LAB [25]; they have been considered as molecular moieties typical of glycoproteins from mammary gland/epithelial cells [26]. In contrast, the non-glycosylated peptide (32-53) was not detected in the spot 47 digest. On the other hand, ascertained occurrence of the peptides (63-79) and (59-79) as non-glycosylated species (MH^+ signals at *m/z* 2005.1 and 2592.8, respectively), demonstrated that no other glycosylation sites were present in α-LAB. By using a similar approach, β-LGB isoforms (spot 49-52) were comparatively evaluated for PTM analysis; although a sequence coverage of 93.2% was obtained for all spots, no difference was detected between MS spectra.

To identify the maximum number of buffalo proteins present in milk whey fraction, we used a combined strategy comprising analysis with 2-DE plus MALDI-TOF PMF and 1-DE plus μLC-ESI-IT-MS-MS. In the first case, whey proteins were separated on 2-DE gels covering pH 3–10 and *M*r 6–100 kDa ranges, generating maps similar to that reported in Fig. 2A. Among 84 spots detected within gels, 56 were identified by PMF experiments and associated to 11 mammalian genes (Supporting Information Table S6). Expression products of nine genes were present as multiple protein isoforms, generating trains of electrophoretic spots. Whey was also loaded onto a 12% SDS-PAGE and resolved (Fig. 2B); each gel lane was cut into different slices that were digested with trypsin and subjected to μLC-ESI-IT-MS-MS analysis. Each sample was analyzed three times to ensure data reproducibility (almost 96.9%). Polypeptide species corresponding to 29 genes were identified (Supporting Information Table S7), among which 22 not detected by 2-DE plus PMF analysis, thus confirming that complementary approaches generally ensures higher identification rates. Out of the protein products identified as corresponding to 33 genes, 29 corresponded to secreted and four to membrane-associated components; in the latter group, nicked forms with an increased solubility were observed (Supporting Information).

As expected, most abundant proteins on the gels corresponded to β-LGB and α-LAB [1, 10, 13]. Additional components associated to defense/immunity (lactotransferrin, IgM heavy chain, Ig λ chains, polymeric immunoglobulin receptor, complement C3, cluster of differentiation (CD)14, Zn-α2-glycoprotein, factor XIIa inhibitor, inter α-trypsin inhibitor, nucleobindin 1 and 2), transport (albumin, folate-binding protein, mammary-derived growth inhibitor) and metabolic (lactoperoxidase, β-1,4-galactosyltransferase) functions were also identified (Table 1 and Supporting Information). Some of these proteins are important for ensuring essential nutri-

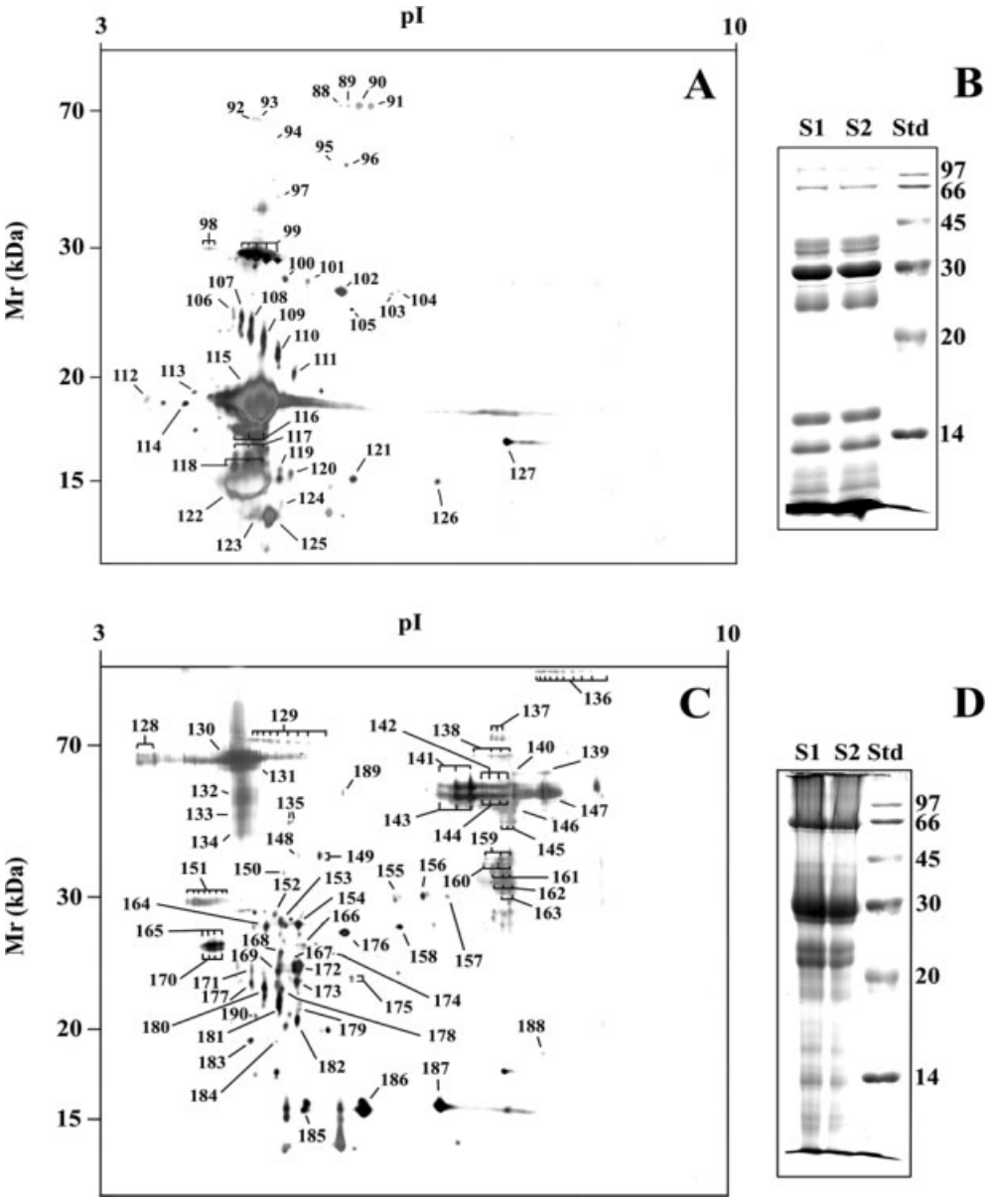

Figure 2. 2-DE of buffalo whey (panel A) and MFGM (panel C) proteins. Proteins (150 μg) were separated in the first dimension using 18-cm pH 3-10 IPG strips and in the second dimension using 12.5% SDS-PAGE. Proteins were visualized by ammoniacal silver staining. Identified protein spots are numbered and reported in Supporting Information Tables S6 (whey) and S8 (MFGM). 1-DE of buffalo whey (panel B) and MFGM (panel D) proteins. Proteins (100 and 50 μg) were separated by using 12% SDS-PAGE and visualized by colloidal CBB G250 and ammoniacal silver staining, respectively. S1 and S2 correspond to milk samples from different animals. Identified proteins are reported in Supporting Information Tables S7 (whey) and S9 (MFGM).

ents to the newborn and promoting its development (caseins, folate-binding protein, lactotransferrin, β-1,4-galactosyltransferase, miostatin).

A combined strategy was also used to investigate proteins present within buffalo MFGM fraction; this approach was necessary because unique studies already published on MFGM, referring to human and bovine membranes, emphasized the problematic issue of analyzing its hydrophobic proteins [10, 11, 27, 28]. In the case of 2-DE, the use of analytical gels ensured optimal spot resolution. Figure 2C shows the electrophoretic map obtained for buffalo MFGM proteins, using a gel covering pH 3–10 and 10–100 kDa ranges. Approximately 196 reproducible spots were detected. All proteins occurring within this 2-DE map were analyzed by MALDI-TOF PMF procedures; 118 spots were identified and associated to 19 mammalian genes (Supporting Information

Table S8). Proteins corresponding to 12 genes were found as multiple electrophoretic species generating trains of spots; heterogeneity was mainly due to variability in apparent pI values or a combination of p*I*/*M*r parameters. Some of these proteins have been already reported containing phosphate groups and/or N-linked glycans, generating distinct 2-DE isoforms [27]; uncommon glycosylation branching patterns have been described for these proteins synthesized in mammary grand, which may be influential in maternal protection of the newborn against bacterial/viral pathogenic attack [26, 27]. On the other hand, spots associated with dominant xanthine dehydrogenase/oxidase, adipophilin, lactadherin and butyrophilin (BTN) isoforms also appeared at various *M*r values (Fig. 2C). Their occurrence was the result of limited proteolysis events yielding minor species with reduced *M*r values, which have a different relative solubility with respect to native proteins; observation of these components has been already described in dedicated studies on human/bovine MFGM [10, 11, 27–29]. In particular, almost 28 spots were identified as BTN1A1 proteolysis products, which probably corresponded to the various truncated forms containing exo-IgI, exo-IgC1 and cyto-B30.2 domains [28]. No other spots were associated to other BTN-related genes, as result of the poor information on buffalo genome.

To increase protein recovery, MFGM polypeptides were also solubilized in an SDS-containing buffer and resolved by glycine-based SDS-PAGE (Fig. 2D). A whole-gel lane was cut into different slices, which were digested and directly analyzed by automated μLC-ESI-IT-MS-MS for identification of protein components. Also in this case, each sample was analyzed three times to ensure data reproducibility (almost 97%). By using this approach, polypeptide species corresponding to 42 different genes were identified (Supporting Information Table S9). A larger number of membrane proteins were identified, suggesting that using of loading buffers with higher solubilizing properties or absence of experimental steps involving IPG [11] strongly improved analysis of hydrophobic proteins. In conclusion, out of the polypeptide products ascertained in MFGM fraction and associated to 50 genes, 11 were identified by both proteomic approaches, eight were uniquely identified by combined 2-DE/MALDI-TOF-MS and 31 were uniquely identified by combined 1-DE/μLC-ESI-IT-MS-MS. Out of these 50 proteins, 65% are membrane-associated proteins, with the remainder being either cytoplasmic (25%) or secreted (10%) polypeptides.

Identified proteins represented all functional categories (Table 1 and Supporting Information); those involved in membrane/protein trafficking (21%), fat transport/metabolism (17%), cell signaling (15%) and defense (9%) functions accounted for little more than a half of the proteins detected in MFGM. This classification, however, did not fully consider the multiple functions of some proteins. For example, xanthine dehydrogenase/oxidase is not classified as host defense-related, despite clearly having a role in production of reactive-oxygen species, which modulate immune response [30]. A similar immunoprotective/defense function can also be ascribed to (i) lactadherin, which resists being digested in the stomach of milk-fed newborn, and prevents symptomatic rotavirus infection in mammal-fed neonates [31]; (ii) cyclophilin A, which may be secreted in response to inflammatory stimuli and acts as an effective neutrophil/eosinophil chemoattractant [32]; (iii) mucins, which have been associated to immunoprotective roles on the basis of their binding/sequestering activity on pathogenic microrganisms and inhibitory effects on rota/pox-viruses replication [27]. Thus, all proteins mentioned above may be considered together with immunoglobulins, membrane-associated CD14, Toll-like receptor 2 (TLR2) and α2-macroglobulin, thus increasing the number of defense-related polypeptide species [1, 10, 11, 27]. In particular, CD14 cooperates with TLRs to mediate the innate immune response to bacterial LPS, thus determining NF-kB-regulated production of a variety of antibiotic peptides [33]. TLRs are mainly present on antigen presenting cells and signal the occurrence of pathogens to the host. Similarly, inflammation-associated α2-macroglobulin is generally overexpressed in mammary gland by activation of STAT5 [34]; it has a function of binding foreign particles, thereby serving as a humoral defense barrier against pathogens in milk during mastitis. The functional significance of all these proteins in host defense has to be tested by further investigations on their direct effect on microbes or in modulating the responsiveness of immune cells to their presence. Moreover, the diversity of proteins identified in MFGM fraction reflected the different functional roles of this specialized membrane [27, 28, 35] and provided preliminary information on lipid droplet's secretory pathways within mammary gland and its eventual controlling processes (see Supporting Information). In this section, additional information on proteins belonging to other functional classes is also reported.

In conclusion, this study firstly provides a detailed description of buffalo milk proteome, including minor protein components. Dominant polypeptide species, *i.e.* α_{S1}-, α_{S2}-, β-, κ-CN, β-LGB and α-LAB, were fully characterized, describing relative quantitative levels and PTM. In particular, an increased content in α_{S2}- and κ-CN as well as α_{S1}- and β-CN with a reduced phosphorylation degree were ascertained in buffalo milk with respect to other ruminants. By describing some of the main characteristics affecting milk cheese-making aptitude, together with total fat/protein content, pH and Ca^{2+} concentration, this study provides a scientific basis to the coagulation processes used in buffalo milk-based dairy productions. In fact, elevated κ-CN content in milk decreases casein micelle size, reducing gelation/coagulation and increasing digestibility; similarly, low α-LAB levels limit lactose concentration in milk and consequently diminish water relative concentration [8, 9, 35]. On the other hand, increased phosphorylation on caseins improves milk emulsification properties [35]. Proteins regulating pathways of milk fat secretion or involved in stabilizing lipid droplets after secretion were also characterized. These results may be of major importance to the dairy manufacturing because the nature of

the cream fraction influences the built-up properties and organoleptic qualities of milk and dairy products [35]. Since the official method for the detection of adulteration in buffalo milk is still based on 1-DE (Italian Gazzetta Ufficiale n.160, 11/07/1994), we also believe that the resolutive 2-DE maps here reported will be useful for the development of new certified protocols for the detection of illegal addition of milk from other ruminants to buffalo one.

The number of minor protein species identified in this work emphasizes the multiple functions of milk, which besides affording nutrition to the newborn through its major components, also promotes neonate development and protection of the digestive tract from pathogens, as well as helps maintaining optimal mammary gland function in the mother. In fact, this gland has a cavous architecture, which is inherently susceptible to microbial infection; abundant nutrients in residual milk and body temperature ensure optimal growing conditions for microbes. Thus, protection against pathogens in neonate and mother organs is guaranteed by the arsenal of antimicrobial and immunomodulatory proteins, which are secreted into milk. The protective role of some of these components in milk has been underlined in dedicated reviews [1, 27, 28, 35, 36] and, recently, found its definitive scientific foundation in two comparative proteomic investigations made on animals affected by mastitis [10, 13]. These studies and our results underline that MFGM and whey fractions seem to be particularly rich in these antimicrobial/immunomodulatory proteins. Moreover, our findings suggest that molecules generally expressed on antigen presenting cells/general secretory vesicles and involved in signaling the presence of pathogens, such as CD14 and TLRs, also occur in MFGM and, directly, on the apical membrane of the secretory cells; thus, microbes entering the mammary gland could be signaled to the host directly through this gland.

This work was supported by grants from National Research Council (AG.P04.015 and RSTL 862).

The authors have declared no conflict of interest.

References

[1] Farrell, H. M., Jimenez, R., Bleck, G. T., Brown, E. M. *et al.*, Nomenclature of the proteins of cows' milk–sixth revision. *J. Dairy Sci.* 2004, *87*, 1641–1674.

[2] Guy, P. A., Fenaille, F. Contribution of mass spectrometry to assess quality of milk-based products. *Mass Spectrom. Rev.* 2006, *25*, 290–326.

[3] Chen, R. K., Chang, L. W., Chung, Y. Y., Lee, M. H. *et al.*, Quantification of cow milk adulteration in goat milk using high-performance liquid chromatography with electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2004, *18*, 1167–1171.

[4] Scaloni, A., Perillo, V., Franco, P., Fedele, E. *et al.*, Characterization of heat-induced lactosylation products in caseins by immunoenzymatic and mass spectrometric methodologies. *Biochim. Biophys. Acta* 2002, *1598*, 30–39.

[5] Holland, J. W., Deeth, H. C., Alewood, P. F., Proteomic analysis of kappa-casein micro-heterogeneity. *Proteomics* 2004, *4*, 743–752.

[6] Fenaille, F., Parisod, V., Tabet, J. C., Guy, P. A., Carbonylation of milk powder proteins as a consequence of processing conditions. *Proteomics* 2005, *5*, 3097–3104.

[7] Mamone, G., Caira, S., Garro, G., Nicolai, A. *et al.*, Casein phosphoproteome: identification of phosphoproteins by combined mass spectrometry and two-dimensional gel electrophoresis. *Electrophoresis* 2003, *24*, 2824–2837.

[8] Ageitos, J. M., Vallejo, J. A., Poza, M., Villa, T. G., Fluorescein thiocarbamoyl-kappa-casein assay for the specific testing of milk-clotting proteases. *J. Dairy Sci.* 2006, *89*, 3770–3777.

[9] Vasbinder, A. J., Rollema, H. S., Bot, A., de Kruif, C. G., Gelation mechanism of milk as influenced by temperature and pH; studied by the use of transglutaminase cross-linked casein micelles. *J. Dairy Sci.* 2003, *86*, 1556–1563.

[10] Smolenski, G., Haines, S., Kwan, F. Y., Bond, J. *et al.*, Characterisation of host defence proteins in milk using a proteomic approach. *J. Proteome Res.* 2007, *6*, 207–215.

[11] Reinhardt, T. A., Lippolis, J. D., Bovine milk fat globule membrane proteome. *J. Dairy Res.* 2006, *73*, 406–416.

[12] Holland, J. W., Deeth, H. C., Alewood, P. F., Resolution and characterisation of multiple isoforms of bovine kappa-casein by 2-DE following a reversible cysteine-tagging enrichment strategy. *Proteomics* 2006, *6*, 3087–3095.

[13] Hogarth, C. J., Fitzpatrick, J. L., Nolan, A. M., Young, F. J. *et al.*, Differential protein composition of bovine whey: a comparison of whey from healthy animals and from those with clinical mastitis. *Proteomics* 2004, *4*, 2094–2100.

[14] Miranda, G., Mahe, M. F., Leroux, C., Martin, P., Proteomic tools to characterize the protein fraction of Equidae milk. *Proteomics* 2004, *4*, 2496–2509.

[15] Ferranti, P., Scaloni, A., Caira, S., Chianese, L. *et al.*, The primary structure of water buffalo alpha(s1)- and beta-casein identification of phosphorylation sites and characterization of a novel beta-casein variant. *J. Protein Chem.* 1998, *17*, 835–844.

[16] Chianese, L., Caira, S., Lilla, S., Pizzolongo, F. *et al.*, Primary structure of water buffalo alpha-lactalbumin variants A and B. *J. Dairy Res.* 2004, *71*, 14–19.

[17] Mitra, A., Schlee, P., Krause, I., Blusch, J. *et al.*, Kappa-casein polymorphisms in Indian dairy cattle and buffalo: a new genetic variant in buffalo. *Anim. Biotechnol.* 1998, *9*, 81–87.

[18] Clare, D. A., Swaisgood, H. E. Bioactive milk peptides: a prospectus. *J. Dairy Sci.* 2000, *83*, 1187–1195.

[19] Galvani, M., Hamdan, M., Righetti, P. G., Two-dimensional gel electrophoresis/matrix-assisted laser desorption/ionisation mass spectrometry of commercial bovine milk. *Rapid Commun. Mass Spectrom.* 2001, *15*, 258–264.

[20] Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. *et al.*, Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* 2005, *4*, 873–886.

[21] Morelle, W., Canis, K., Chirat, F., Faid, V. *et al.*, The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* 2006, *6*, 3993–4015.

[22] Kweon, H. K., Hakansson, K. Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* 2006, *78*, 1743–1749.

[23] Zhou, H., Xu, S., Ye, M., Feng, S. *et al.*, Zirconium phosphonate-modified porous silicon for highly specific capture of phosphopeptides and MALDI-TOF MS analysis. *J. Proteome Res.* 2006, *5*, 2431–2437.

[24] Larsen, M. R., Jensen, S. S., Jakobsen, L. A., Heegaard, N. H., Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. *Mol. Cell. Proteomics* 2007, *6*, 1778–1787.

[25] Slangen, C.J., Visser, S., Use of mass spectrometry to rapidly characterize the heterogeneity of bovine alpha-lactalbumin. *J. Agric. Food Chem.* 1999, *47*, 4549–4556.

[26] Sato, T., Taka, J., Aoki, N., Matsuda, T. *et al.*, Expression of beta-N-acetylgalactosaminylated N-linked sugar chains is associated with functional differentiation of bovine mammary gland. *J. Biochem.* 1997, *122*, 1068–1073.

[27] Mather, I. H., A review and proposed nomenclature for major proteins of the milk-fat globule membrane. *J. Dairy Sci.* 2000, *83*, 203–247.

[28] Cavaletto, M., Giuffrida, M. G., Conti, A., Milk fat globule membrane components - a proteomic approach. *Adv. Exp. Med. Biol.* 2008, *606*, 129–141.

[29] Häggqvist, B., Näslund, J., Sletten, K., Westermark, G. *et al.*, Medin: an integral fragment of aortic smooth muscle cell-produced lactadherin forms the most common human amyloid. *Proc. Natl. Acad. Sci. USA* 1999, *96*, 8669–8674.

[30] Martin, H. M., Hancock, J. T., Salisbury, V., Harrison, R., Role of xanthine oxidoreductase as an antimicrobial agent. *Infect. Immun.* 2004, *72*, 4933–4939.

[31] Kvistgaard, A. S., Pallesen, L. T., Arias, C. F., López, S. *et al.*, Inhibitory effects of human and bovine milk constituents on rotavirus infections. *J. Dairy Sci.* 2004, *87*, 4088–4096.

[32] Yurchenko, V., Zybarth, G., O'Connor, M., Dai, W. *et al.*, Active site residues of cyclophilin A are crucial for its signaling activity via CD147. *J. Biol. Chem.* 2002, *277*, 22959–22965.

[33] Labéta, M. O., Vidal, K., Nores, J. E., Arias, M. *et al.*, Innate recognition of bacteria in human milk is mediated by a milk-derived highly expressed pattern recognition receptor, soluble CD14. *J. Exp. Med.* 2000, *191*, 1807–1812.

[34] Borth, W., Alpha 2-macroglobulin, a multifunctional binding protein with targeting characteristics. *FASEB J.* 1992, *6*, 3345–3353.

[35] Bauman, D. E., Mather, I. H., Wall, R. J., Lock, A. L., Major advances associated with the biosynthesis of milk. *J. Dairy Sci.* 2006, *89*, 1235–1243.

[36] Spitsberg, V. L., Bovine milk fat globule membrane as a potential nutraceutical. *J. Dairy Sci.* 2005, *88*, 2289–2294.

