Amino Acids
DOI 10.1007/s00726-016-2315-z

ORIGINAL ARTICLE

Site-specific glycosylation of donkey milk lactoferrin investigated by high-resolution mass spectrometry

Serafina Gallina[1] · Rosaria Saletti[1] · Vincenzo Cunsolo[1] · Vera Muccilli[1] · Salvatore Foti[1] · Peter Roepstorff[2] · Morten I. Rasmussen[2]

Received: 27 April 2016 / Accepted: 16 August 2016
© Springer-Verlag Wien 2016

Abstract A comprehensive monosaccharide composition of the *N*-glycans of donkey milk lactoferrin, isolated by ion exchange chromatography from an individual milk sample, was obtained by means of chymotryptic digestion, TiO_2 and HILIC enrichment, reversed-phase high-performance liquid chromatography, electrospray mass spectrometry, and high collision dissociation fragmentation. The results obtained allowed identifying 26 different glycan structures, including high mannose, complex and hybrid *N*-glycans, linked to the protein backbone via an amide bond to asparagine residues located at the positions 137, 281 and 476. Altogether, the *N*-glycan structures determined revealed that most of the *N*-glycans identified in donkey milk lactoferrin are neutral complex/hybrid. Indeed, 10 neutral non-fucosylated complex/hybrid *N*-glycans and 4 neutral fucosylated complex/hybrid *N*-glycans were found. In addition, two high mannose *N*-glycans, four sialylated fucosylated complex *N*-glycans and six sialylated non-fucosylated complex *N*-glycans, one of which containing *N*-glycolylneuraminic acid (Neu5Gc), were found. A comparison of the monosaccharide composition of the *N*-glycans of donkey milk lactoferrin with respect to that of human, bovine and goat milk lactoferrin is reported. Data are available via ProteomeXchange with identifier PXD004289.

Keywords Lactoferrin glycosylation · Donkey milk · *N*-Glycan structures · TiO_2 and HILIC enrichment · Mass spectrometry

Handling Editor: P. R. Jungblut.

Electronic supplementary material The online version of this article (doi:10.1007/s00726-016-2315-z) contains supplementary material, which is available to authorized users.

✉ Salvatore Foti
sfoti@unict.it

[1] Organic Mass Spectrometry Laboratory, Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy

[2] Protein Research Group, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

Introduction

Milk contains a variety of components, including proteins, endogenous peptides, lipids, carbohydrates and minerals, all of which contribute to the growth and development of newborns. In the human diet, beyond the simple nutritional value of milk compounds, other components, including glycoproteins, antibodies and oligosaccharides, also protect infants by reducing the number of pathogen infections and promoting the development of the intestinal epithelium (de Wit et al. 1998; Coppa et al. 2006; Yekta et al. 2010; Zivkovic et al. 2011). Milk is, therefore, more than a simple source of essential nutrients. On the other hand, among allergenic foods, milk represents the main source of allergens in early childhood: approximately 2–3 % of infants younger than 1 year of age are allergic to cow's milk proteins (CMPA). Considering that breast feeding is not always possible, indicated or sufficient, an alternative supply then becomes indispensable for infants allergic to cow's milk proteins.

Many clinical investigations have demonstrated that donkey milk (DM) is often a good natural breast milk substitute during early infancy. In vivo, donkey milk has shown to be well tolerated by a large number (83–88 %) of children affected by cow milk protein allergy (CMPA) in terms of clinical tolerability (Restani et al. 2002; Monti et al.

2007; Swar 2011; Monti et al. 2012). In vitro, monoclonal and polyclonal antibodies produced against cow milk proteins showed a very mild cross-reaction with donkey milk (Restani et al. 2009; El-Agamy et al. 2009). Furthermore, donkey milk is known to possess natural protective antimicrobial factors and a specific epidermal growth factor (EGF) that suggest its beneficial impact on gastrointestinal mucosa health and integrity; a claim particularly valuable for children, the elderly, and convalescents, who have a reduced immune defense system (Scafizzari et al. 2009). Consequently, DM has recently aroused scientific and clinical interest, and over the last decade many investigations, carried out by different proteomic approaches (Cunsolo et al. 2014), have been focused on the characterization of both major (Cunsolo et al. 2007a, b, 2009a; Bertino et al. 2010; Chianese et al. 2010) and low-abundance proteins (Barello et al. 2008; Cunsolo et al. 2011b). Although the mechanism of this tolerance has not yet been fully clarified, it is reasonable to hypothesize that the reduced allergenic properties of DM can be related to the low-sequence similarity of its protein components with respect to bovine milk, as recently reported (Cunsolo et al. 2009b; Saletti et al. 2012).

Milk proteins can be grouped into three classes, according to their different solubilities: caseins, whey proteins and milk fat globule membrane (MFGM) proteins. Lactoferrin (LF), member of the transferrin family, is one of the most important glycoproteins (Hutchens et al. 1994) present in the whey protein fraction of milk. Its function is modulated by both the polypeptide chain and its glycosylation (Barboza et al. 2012). The *N*-glycan components of LF whose glycosylation pattern has been studied so far have a highly heterogeneous structure, containing *N*-acetylglucosamine (GlcNAc), galactose(Gal), fucose (Fuc), mannose (Man), and *N*-acetylneuraminic acid (Neu5Ac). *N*-glycolylneuraminic acid (Neu5Gc) has been found in both goat milk LF and bovine milk (Le Parc et al. 2014; Nwosu et al. 2012). *N*-Glycans vary widely in composition and structure, even within a single site of glycosylation. *N*-Glycans are divided into three main classes: high mannose, complex and hybrid (Stanley et al. 2009). Identifying the structure and glycosylation site represents a significant analytical challenge because, unlike the expression of nucleic acid and peptide polymers, glycan biosynthesis is not a transcription of a coded structure. The non-templated nature of glycan biosynthesis results in the presence of heterogeneous elongated branches and isomers. Glycosylation can modify the structural conformation of the protein and consequently its biological activity (Marth and Grewal 2008; Shental-Bechor and Levy 2009). Milk glycans can interfere with pathogen adhesion to intestinal epithelial cells (Bode 2012), which reinforces the idea that glycosylation can be involved in protecting the host against microbial and viral attacks. LF is also known to exert bifidogenic effects on some bifidobacteria. One of the possible mechanisms underlying the bifidogenic activity of LF is the provision of sugar chains (Oda et al. 2014) bound to LF. The addition of human LF or bovine LF to a sugar-restricted medium induced gene expression associated with sugar metabolism in *Bifidobacteria infantis*, and deglycosylated LF was detected in the medium (Garrido et al. 2012). These findings suggested that sugar chains bound to LF could be utilized as a carbon source by bifidobacteria.

Within the frame of our research on investigating donkey milk proteins (Cunsolo et al. 2011a), we have previously reported the sequence characterization and glycosylation sites identification of donkey milk LF (Gallina et al. 2016). The aim of this study was to determine the site-specific *N*-glycans structure in donkey milk LF, isolated by ion exchange chromatography from an individual milk sample. The combined use of chymotryptic digestion, TiO_2 and HILIC enrichment, reversed-phase high-performance liquid chromatography, electrospray mass spectrometry and high collision dissociation fragmentation resulted in the identification of 26 different glycan structures.

Materials and methods

Materials

Acetic acid (AA), formic acid (FA), glycolic acid (GA), α-chymotrypsin from bovine pancreas, albumin from chicken egg white, fetuin from fetal bovine serum were obtained from Sigma-Aldrich (Milan, Italy). Ammonium bicarbonate, sodium acetate, HPLC-grade H_2O and acetonitrile (ACN) were provided by Carlo Erba (Milan, Italy). Dithiothreitol (DTT), iodoacetamide (IA), ammonium acetate and sodium chloride were purchased from Fischer-Scientific (Milan, Italy). Micro-Spin regenerated cellulose filters (0.45 μm pore size) and nylon membranes (0.22 μm pore size) used for filtration of samples and ion exchange chromatography (IEC) solvents, respectively, were from Alltech (Milan, Italy). Spectra-Por Float-A-Lyzer dialysis tubes (cut off 3.5–5 kDa) were obtained from Sigma-Aldrich (Milan, Italy). TiO_2 beads were obtained from GL Sciences (Japan). Polyhydroxyethyl A was provided by PolyLCINC (Columbia, MD, USA). POROS Oligo R3 was purchased from Applied Biosystems (Framingham, MA, USA). GELoader tips were obtained from Eppendorf (Hamburg, Germany). All the materials had a high degree of purity and were used without further purification. Donkey milk LF was isolated from an individual milk sample as previously reported (Gallina et al. 2016).

Enzymatic digestion of lactoferrin and glycopeptide enrichment

To obtain low molecular mass glycopeptides, the chromatographic fraction corresponding to donkey LF was reduced, alkylated and digested using α-chymotrypsin. The enzyme was dissolved in 100 mM NH_4HCO_3, pH 8.3, added at a molar enzyme/substrate ratio 1:50 and the solution was incubated at 37 °C overnight. The digestion reaction was stopped by cooling in liquid nitrogen. The digested mixture, containing non-glycosylated peptides and glycopeptides, was dried down and diluted in 100 μL of loading buffer (80 % ACN, 5 % TFA and 1 M glycolic acid) and added to 0.3 mg TiO_2-beads. After stirring, TiO_2-beads, which interact selectivity with sialylated glycopeptides (Larsen et al. 2007), were separated by centrifugation. The supernatant, containing only neutral glycopeptides and non-glycosylated peptides, was removed and stored. The beads were washed first with 50 μL of loading buffer, second with 50 μL of washing buffer 1 (80 % ACN, 1 % TFA), and third with 50 μL of washing buffer 2 (10 % ACN, 0.5 % TFA). In each step the supernatant was stored. Then the beads with sialylated glycopeptides were dried and the sialylated glycopeptides were eluted from the beads with 50 μL of elution buffer (60 μL ammonia solution (28 %) in 940 μL H_2O, pH 11.3). To enrich for neutral glycopeptides, a home-made column packed with polyhydroxyethyl A 3 μm as stationary phase was used (Hagglund et al. 2004). In brief, a D10 pipette tip (Gilson) was plugged with a C18 empore disc in the bottom and the HILIC material in ACN was then applied on the top. The previous supernatant stored from TiO_2 enrichment, containing only neutral glycopeptides and non-glycosylated peptides, was dried down, re-suspended in 20 μl of 80 % ACN and 2 % aqueous FA, loaded in the home-made column. The flow-through contained only non-glycosylated peptides, whereas the neutral glycopeptides, trapped in the column were eluted with 2 % aqueous FA.

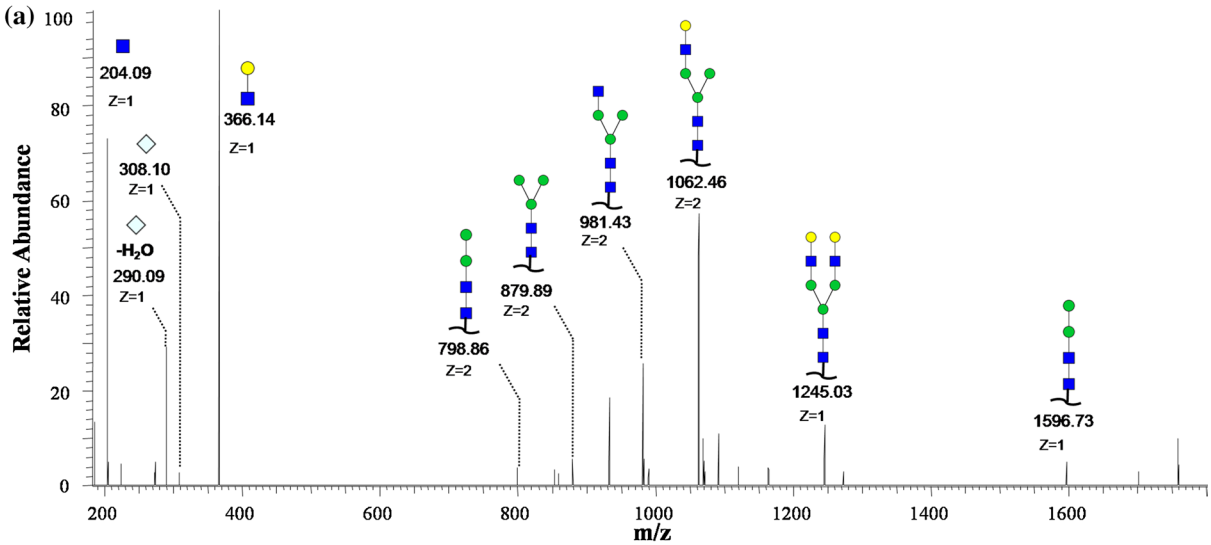
Nano-LC MS/MS

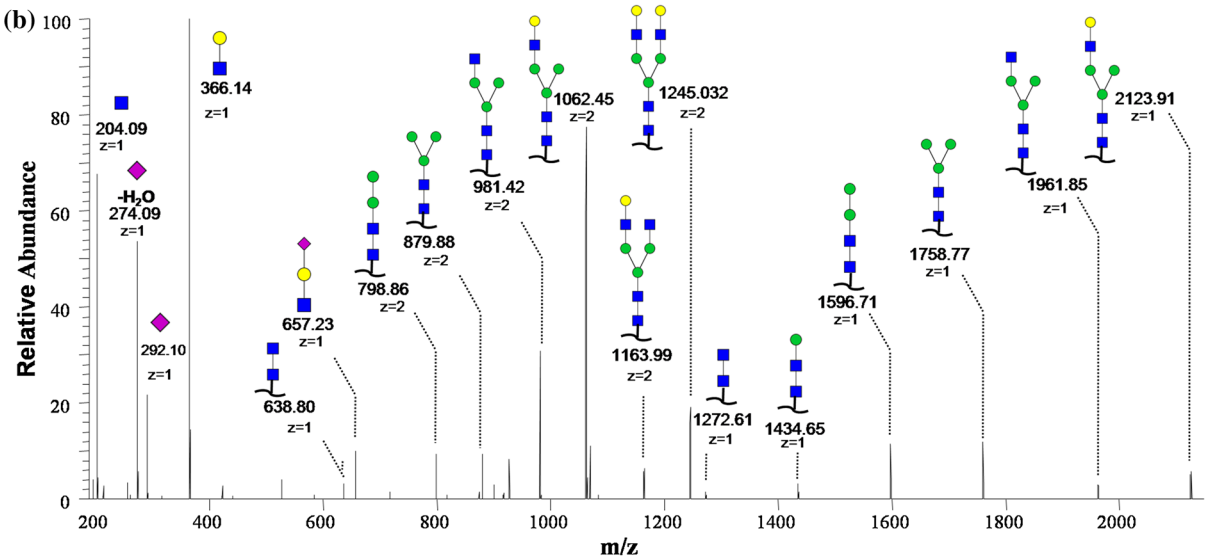
Mass spectrometry data were acquired on a Hybrid Quadrupole-Orbitrap mass spectrometer (Q-Exactive Plus, Thermo Fisher Scientific, Bremen, Germany). Liquid chromatography was carried out using a Thermo Scientific Dionex UltiMate 3000 nano HPLC (Thermo Fisher Scientific). Five μL of the reconstituted samples in 0.1 % FA were loaded onto a home-made μ-pre-column (100 μm × 2 cm, 5 μm ReproSil-Pur C18). After flooding the trapping column with solvent A (0.1 % FA) at a flow rate of 5 μL/min for 5 min, the eluent was switched from the trapping column onto a home-made reversed-phase C18 column (75 μm × 17 cm, 3 μm ReproSil-Pur C18). Peptides were separated by elution at a flow rate of 250 nL/min at room temperature with a linear gradient of solvent B (ACN + 0.1 % FA) in A from 1 % to 35 % in 35 min. Eluting peptide cations were converted to gas-phase ions by the Thermo Scientific Nanospray Ion Source using a source voltage of 2.5 kV and introduced into the mass spectrometer through a heated ion transfer tube (275 °C). The mass spectrometer was operated in data-dependent mode acquisition as follows: (1) survey scans of peptide precursors from 700 to 2000 *m/z*, performed at 70 K resolution (@ 200 *m/z*); (2) MS/MS analysis, in the *m/z* range of 200–2000, of the eight most intense ions. Automatic gain control (AGC) target for full MS acquisitions was set to 1×10^6 with a maximum ion injection time of 120 ms. Microscans were set to 1 for both the MS and MS/MS. Dynamic exclusion was set to 15 s. Peptides of two and higher charges were fragmented in the HCD collision cell using a normalized collision energy (NCE) of 20. Subsequent MS/MS were acquired using an AGC target value of 2×10^4, a maximum injection time of 100 ms and a resolution of 17.5 K. MS data acquisition was performed using the *Xcalibur* v. 3.0 software (Thermo Fisher Scientific).

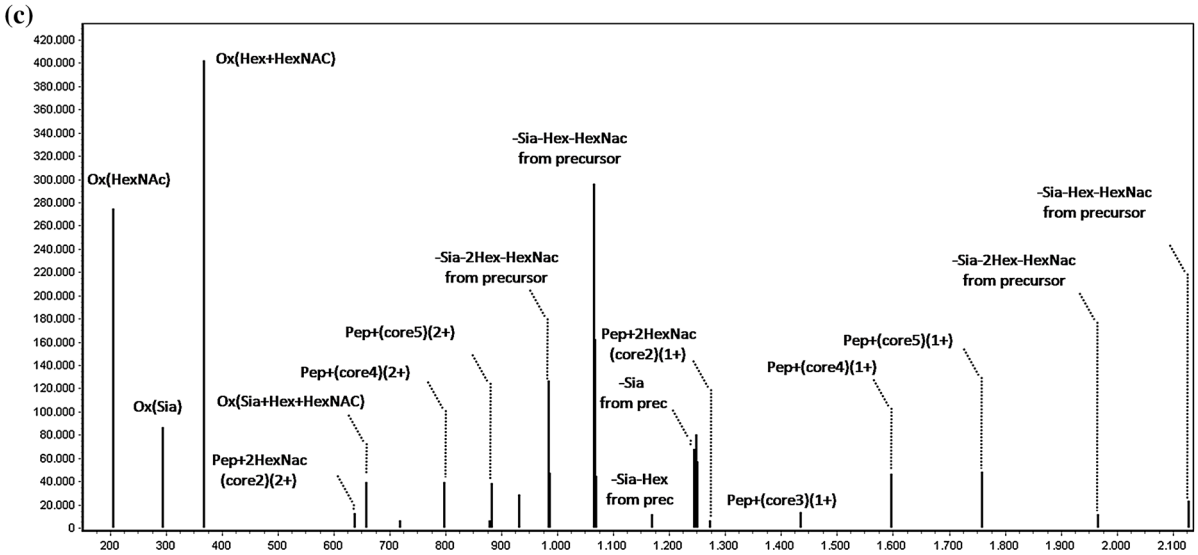
Glycopeptides were identified with *MassAI*, a freely available, in-house-developed software tool (http://www.massai.dk). The software compares the fragmentation patterns of MS/MS scans (mgf format), against a protein database and a flexible library of glycans (Li et al. 2015). The *N*-glycans in the library comprised hexoses (Hex), *N*-acetylhexoseamine (HexNAc), Fuc, Neu5Ac and Neu5Gc. The search was performed with the August 2015 version of *MassAI*, using default settings. This includes: chymotryptic digest with three missed cleavages, carbamidomethylation of Cys as fixed modification, oxidation of Met and deamidation of Gln as variable modifications, mass tolerance of 10 ppm at precursor level and 0.1 Da tolerance at ms/ms peak level.

Results and discussion

Donkey milk LF was isolated from the whey fraction of an individual donkey milk sample as previously described (Gallina et al. 2016). To obtain a comprehensive site-specific monosaccharide composition of the *N*-glycans of donkey milk LF, the isolated protein was dialyzed, reduced, alkylated and digested with α-chymotrypsin, which produces shorter glycopeptides than trypsin. RP-HPLC/nESI–MS/MS was selected for glycopeptide analysis. Despite recent advances in mass spectrometry, analysis of protein glycosylation is still very challenging, because the signals of glycopeptides are often suppressed in the presence of other peptides (Geyer and Geyer 2006; Wohlgemuth et al. 2009). For this reason, glycopeptide enrichment was


(a)
Relative Abundance
m/z
204.09
Z=1
366.14
Z=1
308.10
Z=1
-H2O
290.09
Z=1
798.86
Z=2
879.89
Z=2
981.43
Z=2
1062.46
Z=2
1245.03
Z=1
1596.73
Z=1


(b)
Relative Abundance
m/z
204.09
z=1
366.14
z=1
-H2O
274.09
z=1
292.10
z=1
638.80
z=1
657.23
z=1
798.86
z=2
879.88
z=2
981.42
z=2
1062.45
z=2
1163.99
z=2
1245.032
z=2
1272.61
z=1
1434.65
z=1
1596.71
z=1
1758.77
z=1
1961.85
z=1
2123.91
z=1


(c)
Ox(HexNAc)
Ox(Sia)
Ox(Hex+HexNAC)
Pep+2HexNac
(core2)(2+)
Ox(Sia+Hex+HexNAC)
Pep+(core4)(2+)
Pep+(core5)(2+)
-Sia-2Hex-HexNac
from precursor
-Sia-Hex-HexNac
from precursor
-Sia-Hex
from prec
-Sia
from prec
Pep+2HexNac
(core2)(1+)
Pep+(core3)(1+)
Pep+(core4)(1+)
Pep+(core5)(1+)
-Sia-2Hex-HexNac
from precursor
-Sia-Hex-HexNac
from precursor

◂**Fig. 1** **a** MS/MS spectrum of triply charged molecular ions of the glycopeptide $HexNAc_4Hex_5Neu5Gc_1$ + GRNKSSAF in the RP-HPLC/ESI–MS/MS analysis of the α-chymotryptic digest of donkey LF. Experimentally determined molecular mass of the glycopeptide is 2795.113, which corresponds to the theoretical one 2795.111, with an error of 0.002 Da (1 ppm); **b** MS/MS spectrum of triply charged molecular ions of the glycopeptide $HexNAc_4Hex_5Neu5Ac_1$ + GRNKSSAF. Experimentally determined molecular mass of the glycopeptide is 2779.116 Da, which corresponds to the theoretical one 2779.118, with an error of 0.002 Da (1 ppm); **c** *MassAI* identification of the glycopeptide $HexNAc_4Hex_5Neu5Ac_1$+GRNKSSAF. The freely available *MassAI* software compares the fragmentation patterns of ms/ms scans (mgf format), against a protein database and a flexible library of glycans

performed prior to RP-HPLC/nESI–MS/MS analysis by TiO_2 and HILIC enrichments as described in the experimental section.

An additional complication in the full characterization of the *N*-glycans of animal milk proteins is the presence of Neu5Gc, which is generally not found in human milk. Neu5Gc prevents determination of the composition based exclusively on accurate mass because combinations of Fuc and Neu5Gc yield masses isobaric with oligosaccharides containing Neu5Ac and Hex. For example, the neutral mass 1931.69 Da may not only correspond to $GlcNAc_4Hex_5Neu5Ac_1$ but can also correspond to $GlcNAc_4Hex_4Fuc_1Neu5Gc_1$. Tandem MS is required to discriminate between the two compositions, because MS/MS spectra of glycans or glycopeptides are characterized by the presence of carbohydrate-specific oxonium fragment ions.

Therefore, the enriched mixtures of glycopeptides were investigated by RP-HPLC/nESI–MS/MS coupled online with a hybrid ESI–MS LTQ/Orbitrap mass spectrometer (Q-Exactive Plus) using HCD fragmentation and MS/MS data were used for database searching.

In the spectra obtained, diagnostic ions at *m/z* 308.10 $[Neu5Gc+H]^+$ and 290.10 $[Neu5Gc–H_2O+H]^+$ indicate the presence of Neu5Gc, while ions at *m/z* 292.10 $[Neu5Ac+H]^+$, *m/z* 274.10 $[Neu5Ac–H_2O+H]^+$ indicate the presence of Neu5Ac. In addition, ions at *m/z* 163.06 $[Hex+H]^+$, 204.09 $[HexNAc+H]^+$, 325.11 $[2Hex+H]^+$, 366.14 $[HexNAc+Hex+H]^+$, 657.23 $[HexNAc+Hex+Neu5Ac+H]^+$ and 407.17 $[2HexNAc+H]^+$, are indicative of the presence of glycopeptides. Figure 1a shows the ESI–MS/MS spectrum of the triply charged molecular ion of the glycopeptide $HexNAc_4Hex_5Neu5Gc_1$ + GRNKSSAF. The experimentally determined molecular mass of the glycopeptide is 2795.113 Da, which corresponds to the theoretical one of 2795.111 Da, with an error of 0.002 Da (1 ppm). The carbohydrate-specific oxonium fragment ions at *m/z* 308.10 $[Neu5Gc+H]^+$ and 290.09 $[Neu5Gc–H_2O+H]^+$ unequivocally indicate the presence of $Neu5Gc_1$. Figure 1b shows the ESI–MS/MS spectrum of the triply charged molecular ion of the glycopeptide $HexNAc_4Hex_5Neu5Ac_1$+GRNKSSAF. The experimentally determined molecular mass of the glycopeptide is 2779.116 Da, which corresponds to the theoretical one of 2779.118 Da, with an error of 0.002 Da (1 ppm). Based on the determined molecular mass, two possible glycan compositions can be suggested for the GRNKSSAF glycosylated peptide: $GlcNAc_4Hex_5Neu5Ac_1$ or $GlcNAc_4Hex_4Fuc_1Neu5Gc_1$. However, the MS/MS spectrum confirms the $GlcNAc_4Hex_5Neu5Ac_1$ compositions due to the presence of *m/z* 292.10 $[Neu5Ac+H]^+$ and *m/z* 274.09 $[Neu5Ac–H_2O+H]^+$ ions.

The glycopeptides were determined by accurate mass measurements of the parent ions and from the MS/MS spectra, using *MassAI*. The *MassAI* identification of the glycopeptide $HexNAc_4Hex_5Neu5Ac_1$+GRNKSSAF is shown in Fig. 1c as a typical example. All results were manually checked to validate the accurate mass of the potential glycopeptides and the compatibility of the fragment ions observed in the MS/MS spectra.

N-linked glycosylation in mammals occurs via the amide group of asparagine in the consensus tripeptide sequence Asn-X-Ser/Thr, where X can be any amino acid except proline (Bause and Hettkamp 1979). Occasionally, *N*-glycans are found at Asn-X-Cys, providing that the cysteine residue is in the reduced form (Stanley et al. 2009). The primary structure of donkey LF presents three Asn residues in position 137, 281 and 476 that satisfy the consensus tripeptide sequence Asn-X-Ser/Thr and two Asn in position 168 and 513 that satisfy the consensus tripeptide Asn-X-Cys. Among the five potential sites of glycosylation, only Asn in position 137, 281 and 476, involved in the consensus tripeptide sequence Asn-X-Ser/Thr, was found glycosylated. In this study, the distribution of *N*-glycans among the three glycosylation sites was identified, as shown by the representative glycopeptides reported in Table 1, thus providing for the first time the determination of the most comprehensive monosaccharide composition of the *N*-glycans of donkey milk LF. The results summarized in Table 2 show that Asn 281 and Asn 476 are the sites bearing the largest number of different glycan structures in donkey LF, with 21 and 17 different *N*-glycan structures, respectively. It should be noted that the *N*-glycan compositions identified can be related to a typical glycan structure, although from our data it is not possible to differentiate between isomers.

Conversely, Asn 137 shows a minor number of *N*-glycan compositions (6 glycopeptides). Moreover, five of the six *N*-glycan compositions identified were also found linked to Asn in position 281 and 476. 13 *N*-glycan compositions were found in common between Asn in positions 281 and 476, while 12 *N*-glycan compositions were found distributed differently between Asn in position 281 and 476.

Altogether the *N*-glycan compositions determined in donkey milk LF revealed that most of the *N*-glycans

Table 1 Identified donkey milk lactoferrin *N*-glycopeptides

Monosaccharide composition of *N*-glycans	*N*-Glycan residue ($-H_2O$) mass	Peptide sequence	Molecular mass		Mass error	
			Theor.	Exp.	ppm	Da
Site 137						
$HexNAc_4Hex_5$	1622.58	LNWTGPPEPLQKAVANF	3503.561	3503.578	5	0.017
$HexNAc_4Hex_5Neu5Ac_1$	1913.68	LNW	2344.894	2344.889	2	0.005
$HexNAc_5Hex_4$	1663.61	LNWTGPPEPLQKAVANFF	3691.655	3691.646	2	0.009
$HexNAc_5Hex_4Fuc_1Neu5Ac_1$	2100.76	LNW	2531.978	2532.015	15	0.037
$HexNAc_5Hex_5Fuc_1$	1971.72	LNWTGPPEPLQKAVANFF	3999.763	3999.741	6	0.022
$HexNAc_6Hex_5$	2028.74	LNWTGPPEPLQKAVANFF	4056.787	4056.765	5	0.022
Site 281						
$HexNAc_2Hex_5$	1216.42	GRNKSSAF	2081.864	2081.866	1	0.002
$HexNAc_3Hex_3$	1095.40	GRNKSSAF	1960.838	1960.805	16	0.032
$HexNAc_3Hex_4$	1257.40	GRNKSSAF	2122.891	2122.891	0	0.000
$HexNAc_3Hex_4Neu5Ac_1$	1548.545	GRNKSSAF	2413.986	2413.985	0	0.001
$HexNAc_4Hex_4$	1460.53	GRNKSSAF	2325.970	2325.969	0	0.001
$HexNAc_4Hex_4Neu5Ac_1$	1751.62	GRNKSSAF	2617.065	2617.053	5	0.012
$HexNAc_4Hex_5$	1622.58	GRNKSSAF	2488.023	2488.023	0	0.000
$HexNAc_4Hex_5Neu5Ac_1$	1913.68	GRNKSSAF	2779.118	2779.116	1	0.002
$HexNAc_4Hex_5Neu5Gc_1$	1929.67	GRNKSSAF	2795.113	2795.111	1	0.002
$HexNAc_4Hex_5Fuc_1$	1768.64	GRNKSSAF	2634.081	2634.088	3	0.007
$HexNAc_4Hex_5Fuc_1Neu5Ac_1$	2059.735	GRNKSSAFQLF	3313.387	3313.341	4	0.046
$HexNAc_5Hex_3$	1501.555	GRNKSSAF	2366.996	2366.995	0	0.001
$HexNAc_5Hex_4$	1663.61	GRNKSSAF	2529.049	2529.049	0	0.000
$HexNAc_5Hex_4Neu5Ac_1$	1954.70	GRNKSSAF	2820.145	2820.145	0	0.000
$HexNAc_5Hex_5Fuc_1$	1971.72	GRNKSSAF	2837.160	2837.126	13	0.036
$HexNAc_5Hex_6$	1987.71	GRNKSSAF	2853.155	2853.161	2	0.006
$HexNAc_5Hex_6Fuc_1Neu5Ac_1$	2424.87	GRNKSSAFQLF	3678.519	3678.529	3	0.010
$HexNAc_6Hex_5$	2028.74	GRNKSSAF	2894.181	2894.143	13	0.038
$HexNAc_6Hex_5Fuc_1$	2174.80	GRNKSSAFQLF	3428.450	3428.465	4	0.015
$HexNAc_6Hex_6$	2190.79	GRNKSSAFQLF	3444.445	3444.442	1	0.003
$HexNAc_6Hex_7Fuc_1Neu5Ac_1$	2790.00	GRNKSSAFQLF	3655.440	3655.520	23	0.080
Site 476						
$HexNAc_2Hex_5$	1216.42	NQTGSCKF	2156.829	2156.827	0	0.002
$HexNAc_2Hex_6$	1378.48	NQTGSCKF	2318.882	2318.887	2	0.005
$HexNAc_3Hex_4$	1257.40	NQTGSCKF	2197.856	2197.856	0	0.000
$HexNAc_3Hex_4Neu5Ac_1$	1548.545	NQTGSCKF	2488.951	2488.957	2	0.006
$HexNAc_4Hex_4$	1460.53	NQTGSCKFDKFF	2938.193	2938.181	5	0.012
$HexNAc_4Hex_4Neu5Ac_1$	1751.62	NQTGSCKFDKFF	3229.288	3229.285	1	0.003
$HexNAc_4Hex_5$	1622.58	NQTGSCKF	2562.988	2562.987	0	0.001
$HexNAc_4Hex_5Neu5Ac_1$	1913.68	NQTGSCKF	2854.083	2854.082	0	0.001
$HexNAc_4Hex_5Fuc_1Neu5Ac_1$	2059.735	NQTGSCKF(Q → E)	3001.125	3001.151	9	0.026
$HexNAc_4Hex_5Neu5Ac_2$	2204.77	NQTGSCKF	3145.178	3145.176	1	0.002
$HexNAc_5Hex_4$	1663.608	NQTGSCKF	2604.014	2604.034	8	0.020
$HexNAc_5Hex_4Neu5Ac_1$	1954.70	NQTGSCKF	2895.110	2895.110	0	0.000
$HexNAc_5Hex_5Fuc_1$	1971.72	NQTGSCKF	2912.125	2912.093	11	0.032
$HexNAc_5Hex_6Fuc_1$	2204.77	NQTGSCKF(Q → E)	3116.188	3116.187	0	0.001
$HexNAc_6Hex_3$	1704.635	NQTGSCKF	2645.041	2645.062	8	0.021
$HexNAc_6Hex_5$	2028.74	NQTGSCKF	2969.146	2969.168	7	0.022
$HexNAc_6Hex_5Fuc_1$	2178.80	NQTGSCKF	3145.178	3145.176	1	0.002

Table 2 Identified donkey milk lactoferrin *N*-glycans at Asparagine 137, 281 and 476

Monosaccharide composition of *N*-glycans	*N*-Glycan structure	*N*-Glycan residue ($-H_2O$)	Site 137	Site 281	Site 476
$HexNAc_2Hex_5$	High mannose	1216.42		*	*
$HexNAc_2Hex_6$	High mannose	1378.48			*
$HexNAc_3Hex_3$	Complex	1095.40		*	
$HexNAc_3Hex_4$	Complex/hybrid	1257.40		*	*
$HexNAc_3Hex_4Neu5Ac_1$	Complex	1548.545		*	*
$HexNAc_4Hex_4$	Complex/hybrid	1460.53		*	*
$HexNAc_4Hex_4Neu5Ac_1$	Complex	1751.62		*	*
$HexNAc_4Hex_5$	Complex/hybrid	1622.58	*	*	*
$HexNAc_4Hex_5Fuc_1$	Complex/hybrid	1768.64		*	
$HexNAc_4Hex_5Neu5Ac_1$	Complex	1913.68	*	*	*
$HexNAc_4Hex_5Neu5Gc_1$	Complex	1929.67		*	
$HexNAc_4Hex_5Fuc_1Neu5Ac_1$	Complex	2059.735		*	*
$HexNAc_4Hex_5Neu5Ac_2$	Complex	2204.77			*
$HexNAc_5Hex_3$	Complex	1501.555		*	
$HexNAc_5Hex_4$	Complex/hybrid	1663.61	*	*	*
$HexNAc_5Hex_4Neu5Ac_1$	Complex	1954.70		*	*
$HexNAc_5Hex_4Fuc_1Neu5Ac_1$	Complex	2100.76	*		
$HexNAc_5Hex_5Fuc_1$	Complex/hybrid	1971.72	*	*	*
$HexNAc_5Hex_6$	Complex/hybrid	1987.71		*	
$HexNAc_5Hex_6Fuc_1$	Complex	2133.77			*
$HexNAc_5Hex_6Fuc_1Neu5Ac_1$	Complex	2424.87		*	
$HexNAc_6Hex_3$	Complex	1704.635			*
$HexNAc_6Hex_5$	Complex	2028.74	*	*	*
$HexNAc_6Hex_5Fuc_1$	Complex	2174.79		*	*
$HexNAc_6Hex_6$	Complex	2190.79		*	
$HexNAc_6Hex_7Fuc_1Neu5Ac_1$	Complex	2790.00		*	

* indicates the compositions identified at each specific site

identified are neutral complex/hybrid. Indeed, ten neutral non-fucosylated complex/hybrid *N*-glycans and four neutral fucosylated complex/hybrid *N*-glycans were found. In addition, two high mannose *N*-glycans, four sialylated fucosylated complex/hybrid *N*-glycans and six sialylated non-fucosylated *N*-glycans, one of which with Neu5Gc, are present (Table 2).

The glycosylation profile of LF in human, bovine and goat has been described (Yu et al.2011; Barboza et al. 2012; Nwosu et al. 2012; Le Parc et al. 2014). The human LF (hLF) sequence contains three potential glycosylation sites that correspond to the consensus tripeptide sequence Asn-X-Ser/Thr, at position Asn 138, Asn 479 and Asn 624. All these three sites were found to be glycosylated, even if the third site (Asn 624) is mostly not glycosylated (van-Berkel et al. 1996). In addition, Yu et al. (2011) showed that the glycan profile at Asn 138 is more complex in comparison to that at Asn 479. In bovine lactoferrin (bLF), all the five potential glycosylation sites, which correspond to the consensus tripeptide sequence Asn-X-Ser/Thr, at Asn 252, Asn 300, Asn 387, Asn 495 and Asn 564 were found to be glycosylated, with Asn 252, Asn 387 and Asn 495 most differently glycosylated. Conversely, Asn 564 and Asn 300 were less differently glycosylated with significantly fewer glycopeptide products. To the best of our knowledge, there are no published reports about the occupancy of goat lactoferrin glycosylation sites. Sequence analysis reveals the presence of five potential sites (Asn 233, Asn 281, Asn 368, Asn 476 and Asn 545) involved in the consensus tripeptide sequence Asn-X-Ser/Thr just like in bovine lactoferrin. From the comparison of the *N*-glycan types (Tables 3, Supplementary Table S1), it appears that high mannose *N*-glycans were found only in bovine, goat and donkey milk LF, whereas this structure is not present in human milk LF. The majority of the *N*-glycans in donkey milk LF are neutral *N*-glycans (Table 3). A similar ratio is found in bovine milk LF and in goat milk LF, while in human milk LF a higher number of sialylated *N*-glycans is observed. In addition,

Table 3 Comparison of *N*-glycan structures among human, bovine, goat and donkey milk lactoferrin

N-Glycan structures	Number of identified *N*-glycans			
	Human milk LF[a]	Bovine milk LF[b]	Goat milk LF[c]	Donkey milk LF
High mannose	0	6	5	2
Neutral complex/hybrid	2	11	9	10
Neutral fucosylated complex/hybrid	10	12	6	4
Sialylated complex/hybrid	1	2	7 (5[d])	6 (1[d])
Sialylated fucosylated complex/hybrid	13	0	5 (3[d])	4

[a] (Yu et al.2011; Barboza et al. 2012)

[b] (Nwosu et al. 2012)

[c] (Le Parc et al. 2014)

[d] *N*-Glycans containing Neu5Gc

the ratio of fucosylated types in donkey milk LF is similar to that of bovine milk LF and goat milk LF, while it is lower than in human milk LF. In addition, the number of sialylated types is similar to goat milk LF and higher than in bovine milk LF. Furthermore, among the sialylated types in goat milk LF, eight contain Neu5Gc, which is generally not present in humans. In contrast, only one type containing the Neu5Gc was found in donkey milk. Prior to this study, Neu5Gc was only found in goat milk LF (Table 3), whereas a recent work (Nwosu et al. 2012) suggested its presence in bovine milk (Supplementary Table S1). As a whole, the comparison shows that donkey LF glycosylation composition is markedly different with respect to human and bovine, while it is closer to that of goat, the most relevant difference being the higher presence of Neu5Gc in goat LF. It has been demonstrated that different glycan profiles result in marked differences in immunogenicity and allergenic properties (Almond et al. 2013) and, therefore, the knowledge of the monosaccharide composition of the *N*-glycans is a necessary prerequisite for investigating the allergenic properties of Lfs.

Conclusions

Current interest in LF is related to its biological properties, including antibacterial, antiviral, antioxidant activities, iron-binding and immunomodulation. However, the mechanisms of the antibacterial activity of lactoferrin, including the role of glycosylation in the protection against pathogen infection, have not been completely elucidated yet.

The present paper reports the most comprehensive monosaccharide composition of the *N*-glycans of donkey LF isolated from an individual donkey. The data obtained allowed identifying 26 different *N*-glycan structures linked at residues 137, 281 and 476. Most of the *N*-glycans identified are neutral complex/hybrid. In fact, ten neutral non-fucosylated complex/hybrid *N*-glycans and four neutral fucosylated complex/hybrid *N*-glycans were found. In addition, two high mannose *N*-glycans, four sialylated fucosylated complex *N*-glycans and six sialylated non-fucosylated complex *N*-glycans, one of which with Neu5Gc, are present. The site-specific monosaccharide composition of the *N*-glycans elucidated in this study could enable future investigations on the relationship between glycosylation pattern and protein function.

Acknowledgments This work was supported by Grants from PO FERS 2007/13 4.1.2.A, project "Piattaforma regionale di ricerca translazionale per la salute", CUP B65E12000570008 and from POR 2007/2013 project "BRIT" CUP E61D11000280007. The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD004289. We are grateful to Dr. D. Franchina and Dr. K. Torrisi (ASILAT srl farm, Milo, Catania) for the gift of the raw donkey milk.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest with regard to publication of this research work.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study. Additional informed consent was obtained from all individual participants for whom identifying information is included in this article.

References

Almond RJ, Flanagan BF, Antonopoulos A, Haslam SM, Dell A, Kimber I, Dearman RJ (2013) Differential immunogenicity and allergenicity of native and recombinant human lactoferrins: role of glycosylation. Eur J Immunol 43(1):170–181

Barboza M, Pinzon J, Wickramasinghe S, Froehlich JW, Moeller I, Smilowitz JT, Ruhaak LR, Huang J, Löennerdal B, German JB,

Medrano JF, Weimer BC, Lebrilla CB (2012) Glycosylation of human milk lactoferrin exhibits dynamic changes during early lactation enhancing its role in pathogenic bacteria–host interactions. Mol Cell Proteomics 11(6):M111.015248

Barello C, Perono Garoffo L, Montorfano G, Zava S, Berra B, Conti A, Giuffrida MG (2008) Analysis of major proteins and fat fractions associated with mare's milk fat globules. Mol Nutr Food Res 52(12):1448–1456

Bause E, Hettkamp H (1979) Primary structural requirements for *N*-glycosylation of peptides in rat liver. FEBS Lett 108(2):341–344

Bertino E, Gastaldi D, Monti G, Baro C, Fortunato D, Perono Garoffo L, Coscia A, Fabris C, Mussap M, Conti C (2010) Detailed proteomic analysis on DM: insight into its hypoallergenicity. Front Biosci 2:526–536

Bode L (2012) Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 22(9):1147–1162

Chianese L, Calabrese MG, Ferranti P, Mauriello R, Garro G, De Simone C, Quarto M, Addeo F, Cosenza G, Ramunno L (2010) Proteomic characterization of donkey milk caseome. J Chromatogr A 1217(29):4834–4840

Coppa GV, Zampini L, Galeazzi T, Gabrielli O (2006) Prebiotics in human milk: a review. Digest Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver 38(Suppl 2):S291–S294

Cunsolo V, Costa A, Saletti R, Muccilli V, Foti S (2007a) Detection and sequence determination of a new variant beta-lactoglobulin II from donkey. Rapid Commun Mass Spectrom 21(8):1438–1446

Cunsolo V, Saletti R, Muccilli V, Foti S (2007b) Characterization of the protein profile of donkey's milk whey fraction. J Mass Spectrom 42(9):1162–1174

Cunsolo V, Cairone E, Fontanini D, Criscione A, Muccilli V, Saletti R, Foti S (2009a) Sequence determination of αs1-casein isoforms from donkey by mass spectrometric methods. J Mass Spectrom 44(12):1742–1753

Cunsolo V, Cairone E, Saletti R, Muccilli V, Foti S (2009b) Sequence and phosphorylation level determination of two donkey beta-caseins by mass spectrometry. Rapid Commun Mass Spectrom 23(13):1907–1916

Cunsolo V, Muccilli V, Fasoli E, Saletti R, Righetti PG, Foti S (2011a) Poppea's bath liquor: the secret proteome of she-donkey's milk. J Proteomics 74(10):2083–2099

Cunsolo V, Muccilli V, Saletti R, Foti S (2011b) Applications of mass spectrometry techniques in the investigation of milk proteome. Eur J Mass Spectrom 17(4):305–320

Cunsolo V, Muccilli V, Saletti R, Foti (2014) Mass spectrometry in food proteomics: a tutorial. J Mass Spectrom 49:768–784

de Wit JN (1998) Nutritional and functional characteristics of whey proteins in food products. J Dairy Sci 81(3):597–608

El-Agamy EI, Nawar M, Shamsia SM, Awad S, Haenlein GFW (2009) Are camel milk proteins convenient to the nutrition of cow milk allergy children? Small Rumin Res 82:1–6

Gallina S, Cunsolo V, Saletti R, Muccilli V, Di Francesco A, Foti S, Lorentzen AM, Roepstorff P (2016) Sequence characterization and glycosylation sites identification of donkey milk lactoferrin by multiple enzyme digestions and mass spectrometry. Amino Acids. doi:10.1007/s00726-016-2209-0

Garrido D, Nwosu C, Ruiz-Moyano S, Aldredge D, German JB, Lebrilla CB, Mills DA (2012) Endo-beta-*N*-acetylglucosaminidases from infant gut-associated bifidobacteria release complex *N*-glycans from human milk glycoproteins. Mol Cell Proteomics 11(9):775–785

Geyer H, Geyer R (2006) Strategies for analysis of glycoprotein glycosylation. Biochim Biophys Acta Proteins Proteomics 1764(12):1853–1869

Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P (2004) A new strategy for identification of *N*-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. J Proteome Res 3(3):556–566

Hutchens TW, Lönnerdal B, Rumball S (1994) Lactoferrin structure and function—remaining questions, methodological considerations and future directions. In: Hutchens TW, Rumball SV, Lönnerdal B (eds) Lactoferrin: structure and function. Springer, New York, pp 287–291

Larsen MR, Jensen SS, Jakobsen LA, Heegaard NH (2007) Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. Mol Cell Proteomics 6(10):1778–1787

Le Parc A, Dallas DC, Duaut S, Leonil J, Martin P, Barile D (2014) Characterization of goat milk lactoferrin *N*-glycans and comparison with the *N*-glycomes of human and bovine milk. Electrophoresis 35(11):1560–1570

Li Haiying, Rasmussen Morten I, Larsen Martin R, Guo Yao, Jers Carsten, Palmisano Giuseppe, Mikkelsen Jørn D, Kirpekar Finn (2015) Automated *N*-glycan profiling of a mutant *Trypanosoma rangeli* sialidase expressed in *Pichia pastoris*, using tandem mass spectrometry and bioinformatics. Glycobiology 25(12):1350–1361

MassAI Bioinformatics MassAI. http://www.MassAI.dk

Marth JD, Grewal PK (2008) Mammalian glycosylation in immunity. Nat Rev Immunol 8(11):874–887

Monti G, Bertino E, Muratore MC, Coscia A, Cresi F, Silvestro L, Fabris C, Fortunato D, Giuffrida MG, Conti A (2007) Efficacy of donkey's milk in treating highly problematic cow's milk allergic children: an in vivo and in vitro study. Pediatr Allergy Immunol 18:258–264

Monti G, Viola S, Baro C, Cresi F, Tovo PA, Moro G, Ferrero MP, Conti A, Bertino E (2012) Tolerability of donkey's milk in 92 highly-problematic cow's milk allergic children. J Biol Regul Homeost Agents 26:75–82

Nwosu CC, Aldredge DL, Lee H, Lerno LA, Zivkovic AM, German JB, Lebrilla CB (2012) Comparison of the human and bovine milk *N*-glycome via high-performance microfluidic chip liquid chromatography and tandem mass spectrometry. J Proteome Res 11(5):2912–2924

Oda H, Wakabayashi H, Yamauchi K, Abe F (2014) Lactoferrin and bifidobacteria. Biometals 27(5):915–922

Restani P, Beretta B, Fiocchi A, Ballabio C, Galli C (2002) Cross-reactivity between mammalian proteins. Ann Allergy Asthma Immunol 89(6 Suppl 1):11–15

Restani P, Ballabio C, Di Lorenzo C (2009) Molecular aspects of milk allergens and their role in clinical events. Anal Bioanal Chem 395(1):47–56

Saletti R, Muccilli V, Cunsolo V, Fontanini D, Capocchi A, Foti S (2012) MS-based characterization of αs2-casein isoforms in donkey's milk. J Mass Spectrom 47(9):1150–1159

Scafizzari M, Giannico F, Potere O, Trani A, Colonna MA, Zezza L, Vonghia G, Caputi Jambreghi A (2009) Epidermal growth factor (EGF) in mare and ass milk: a preliminary investigation. Ital J Anim Sci 8(Suppl 2):737

Shental-Bechor D, Levy Y (2009) Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. Curr Opin Struc Biol 19(5):524–533

Stanley P, Schachter H, Taniguchi N (2009) *N*-Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds) Essentials of glycobiology, 2nd edn. Cold Spring Harbor Laboratory Press, New York **(chapter 8)**

Swar MO (2011) Donkey milk-based formula: a substitute for patients with cow's milk protein allergy. Sudan J Paediatr 11:21–24

vanBerkel PHC, vanVeen HA, Geerts MEJ, deBoer HA, Nuijens JH (1996) Heterogeneity in utilization of *N*-glycosylation sites

Asn(624) and Asn(138) in human lactoferrin: a study with glycosylation-site mutants. Biochem J 319:117–122

Wohlgemuth J, Karas M, Eichhorn T, Hendriks R, Andrecht S (2009) Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. Anal Biochem 395(2):178–188

Yekta MA, Verdonck F, Van Den Broeck W, Goddeeris BM, Cox E, Vanrompay D (2010) Lactoferrin inhibits *E. coli* O157:H7 growth and attachment to intestinal epithelial cells. Vet Med Czech 55(8):359–368

Yu T, Guo C, Wang J, Hao P, Sui S, Chen X, Zhang R, Wang P, Yu G, Zhang L, Dai Y, Li N (2011) Comprehensive characterization of the site-specific *N*-glycosylation of wild-type and recombinant human lactoferrin expressed in the milk of transgenic cloned cattle. Glycobiology 21(2):206–224

Zivkovic AM, German JB, Lebrilla CB, Mills DA (2011) Human milk glycobiome and its impact on the infant gastrointestinal microbiota. PNAS 108(Suppl 1):4653–4658