UNIVERSITY OF CATANIA

DEPARTMENT OF BIOLOGICAL, GEOLOGICAL AND ENVIRONMENTAL SCIENCES

DOCTORATE IN BIOTECNOLOGY

XXVIII cycle

Structural Characterization of Donkey Lactoferrin: Amino Acid Sequence and Glycan Compositions

SERAFINA GALLINA

PhD thesis

coordinator: Prof. **Vito De Pinto**

supervisor: Prof. **Salvatore Foti**

correlators: Prof. **Rosaria Saletti**

ACADEMIC YEAR 2012-2015

INTRODUCTION

MILK

The importance of milk in the human diet and in the world economy is well known and it is largely due to its unique nutritive quality, complexity and richness. Milk is the key element for infant nutrition as it represents the only complete source of all essential nutrients for newborns and infants, providing sugars, proteins, fats, vitamins and minerals for healthy growth. Milk is also consumed in various forms, including milk powders, infant formulae, yogurt, cheeses and cream. Proteins represent one of the most important milk components from nutritional and physiological viewpoints. The high nutritional value of the milk protein fraction is related to the high content of essential amino acids.

Milk proteins

Milk proteins can be grouped into three classes according to their different solubilities, which reflect their different structures and functional roles: caseins, whey proteins and milk fat globule membrane (MFGM) proteins.

Caseins (CNs), which are a group of proteins coded by four tightly linked autosomal genes (*CSN1S1*, *CSN1S2*, *CSN2* and *CSN3*), are classically subdivided into four families: α_{s1}-, α_{s2}-, β- and κ-CN.[1] Almost all caseins in milk are organized as macromolecular aggregates of proteins and minerals, known as the casein micelles. The structure of the micelles is not yet well recognized, but several studies have demonstrated a predominantly surface location for κ-CN, which probably plays a fundamental role of stabilizing the micelle structure. The amount of CNs shows remarkable species variation; indeed while they account for 80% (w/w) of all bovine milk proteins, in human[2] and equine[3,4] mature milk, caseins represent only 35% and 50% of total protein content, respectively. Several studies have shown that caseins are quite small molecules with molecular masses of 18-25 kDa, with a great heterogeneity generated by post-translational processing, alternative splicing of the gene product or genetic polymorphisms, that affect the primary structure features and the quantity of each protein.[5,6,7] All caseins are phosphoproteins, showing different levels of phosphorylation, which occurs at serine or threonine residues located in the Ser/Thr-

Xxx-Glu/pSer motif (where Xxx is any amino acid residue and pSer is a phosphoseryl residue). On the contrary, κ-CNs are the only group of casein that also contain carbohydrate moieties. The carbohydrate groups are attached to the κ-CN via O-glycosidic bonds to serine and threonine residues present in the C-terminal region of the protein. The β-CN appears to be the most susceptible to the action of the endogenous milk protease plasmin. [8, 9] Plasmin cleaves at specific sites of β-CNs producing a series of complementary C-terminal and N-terminal polypeptide fragments known as γ-caseins and proteose peptone (PP) components, respectively. [10, 11] These minor components are soluble in acidic condition and therefore in the past were incorrectly classified as whey proteins, a term that is traditionally used to indicate the broad group of milk proteins soluble at pH ~4.6.

The main components of whey proteins are represented by β-lactoglobulin (β-LG), α-lactalbumin (α-LA), serum albumin (SA), lactoferrin (LF) and immunoglobulins (Ig$_s$), but numerous minor proteins (i.e. low-abundance proteins), including enzymes, enzyme inhibitors, metal-binding proteins etc., are also present. [12] Whey protein fraction, with its associated bioactive properties, is considered a functional milk fraction having positive effects on health; therefore whey protein-based hydrolyzed formulae are often used as substitute of common infant formulae to prevent gastrointestinal intolerance to whole bovine milk. [13, 14]

Finally, the MFGM proteins is the third group of milk proteins, having very low nutritional values, but playing important roles in different cell processes and in the defense mechanism for the infants. MFGM proteins represent only 1–4% of total protein content in mature milk; on the contrary in colostrum (i.e. the milk produced in the first week after birth), where the caseins are untraceable, they constitute, together with the immunoglobulins, the principal proteins. Major MFGM proteins include mucin 1, xanthine oxidase, butyrophilin, adipophilin, PAS6/7 (lactadherin) and fatty acid binding protein. [15]

Milk proteins allergy

Proteins represent one of the most important milk components from nutritional and physiological viewpoints. The high nutritional value of the milk protein fraction is related to the high content of essential amino acids. Moreover, milk proteins have native or latent biological functionality, being an important source of bioactive peptides. [16]

These peptides can be inactive as long as they are "hidden" in the primary structure of the proteins, but can become active upon hydrolysis by proteolytic enzymes during processing or gastrointestinal digestion. On the other hand, the wide use of cow's milk in human diet has shown that a considerable percentage of subjects are allergic to its proteins components. Particularly, cow's milk, used as substitute of breast feeding when mother's milk is not available or advisable, represents the main source of allergens in infants, commonly known as cow milk protein allergy (CMPA). Approximately 2-3% of infants younger than one years of age are allergic to cow's milk proteins. This allergy is normally outgrown in the first year of life but 15% of allergic children remain allergic. The proteins most frequently and most intensively recognized by IgE are the caseins and β-LG, even if lower abundant (i.e. lactoferrin, IgG and bovine serum albumin) and trace components appear to be potential allergens. [17, 18]

MS-based methods have been largely employed as confirmatory tools for unambiguous identification and/or characterization of milk allergens. In a general approach, allergens of milk protein extracts can be detected by two-dimensional electrophoresis (2D-PAGE) separation, electro-transferring onto a nitrocellulose membrane and Ig-E immunoblotting analysis with sera from allergic patients. Subsequently, identification of the candidate allergens can be easily performed by their in-gel enzymatic digestion, MS analysis and database searching. This approach has been employed to identify the most abundant cow's milk IgE-reactive protein isoforms in twenty paediatric patients with documented IgE-mediated CMPA. [19] The authors found that the prevalence of cow's milk protein allergens was: 95% IgG heavy chain, 90% α_{s2}-casein, 55% α_{s1}-casein, 50% κ-casein, 50% lactoferrin, 45% β-lactoglobulin, 45% serum albumin, 15% β-casein and 0% α-lactalbumin. As highlighted by the authors, the 2D-immunoblot experiments were not in good agreement with the radioallergosorbent test (RAST), showing an evident discrepancy particularly for α-lactalbumin and α_{s2}-casein. Indeed, while in the immunoblot experiments identification of α-lactalbumin as allergen was not achieved for all subjects, the RAST tests classified this allergen at least as class 2 (moderate level of allergen specific IgE) in seven patients. On the contrary, the 2D-immunoblot of six patients with negative RAST results for total caseins, identified α_{s2}-casein as IgE-immunoreactive protein. The difference observed for α-lactalbumin could be related to the prevalent presence of conformational epitopes of this allergen, that are destroyed in the denaturing conditions of the immunoblot experiments. Instead, the different results obtained for α_{s2}-casein could be explained by the lack of detection of some α_{s2}-casein

epitopes in the RAST analysis, which therefore may lead to false negative results and appears to not always be reliable for diagnosis purposes.

Altenative milk

Taking into account the high incidence of cow milk protein allergy (CMPA) in infants and considering that breast feeding is not always possible, indicated or sufficient, alternative supply becomes indispensable. Therefore, one of the major objective of the pharmaceutical industries is the production of milk and milk-based foods (i.e. infant formulae) close to breast milk. Usually, infant formulas are products based on bovine milk, which is modified by enzymatic and/or thermal treatments, and represent the preferred choices in the treatment of CMPA. However, it should be considered that allergy in these products is reduced, but never completely suppressed, and adverse reactions have been experienced also with these preparations. [20, 21]

Donkey's milk (DM)

Many investigations have reported real benefits from using non-bovine milks as alternative in cases of cow milk protein allergy (CMPA). However, allergies to non-bovine milk proteins have also been documented. [22] On this respect, many clinical trials have shown that equine milk, and in particular, donkey's milk (DM) has special nutritional and therapeutic properties, and may represent a safe and alternative food in CMPA, providing dietary adequacy and good palatability. [23, 24] Although the mechanism of this tolerance has not yet been fully clarified, the reduced allergenic properties of DM are probably related to its unique functional properties and distinctive composition, which is the most similar to that of human milk (HM), with comparable amount of lactose, minerals, fatty acid and proteins. In particular, the relationship between hypo-allergenicity of DM and its proteome fraction has been recently explored by Cunsolo et al. In an investigation of the minor protein components of DM they have identified 106 unique gene products, [25] among which 10% were obtained from the scanty database of *Equus asinus*, whereas the largest proportion (70%) were homologous to those of *Equus caballus*, which is the closest related specie to donkey. On the contrary, only just about 3% could be attributed to bovine milk because the genetic distance between cow and donkey. Indeed, as reported earlier by the same authors, [26, 27, 28, 29] milk proteins from donkey and *Bos taurus* share a low-sequence

similarity. These differences are remarkable when comparing the IgE-binding linear epitopes of cow's α_s-CNs and the corresponding domains present in donkey's counterparts, and may help to explain the already demonstrated low allergenic properties of DM. On the other hand, some clinical trials, carried out using donkey bulk milk, have been reported that a minor number of subjects suffering of CMPA do not tolerate DM as well. It should be noted that bulk sampling strategy prevents the identification of the specific role of single proteins and the impact of the polymorphism on allergenic traits, and investigation of individual milk samples is advisable to highlight the composition-allergenic properties relationships. With this aim, the existence of individual donkey's milk samples lacking single protein components (e.g. β-LG and α_{s1}-CN) [30] might provide to be useful in explaining the unsolved cases observed in clinical trials. The good balance of casein and whey proteins may represent another important factor in determining the hypo-allergenicity of DM proteins. [31] Indeed, the DM protein fraction is particularly rich in whey proteins, representing 35–50% of the nitrogen fraction, whereas in cow's milk only 20%. Donkey whey proteins mainly consist of α-lactalbumin (α-LA), β-lactoglobulins (β-LGs), blood serum albumin (SA), lactoferrin (LF) and lysozyme (LYS). Only one α-LA genetic variant has been reported in donkey's milk, even if an apparent heterogeneity of this component was observed.[32] Donkey β-LG consists of two components, the major β-LG I and the minor β-LG II, each of them showing genetic polymorphism. [33, 34, 35] Two isoforms of LYS, [36] named A and B are reported in donkey's milk, whereas the previously cDNA deduced amino acid sequence of donkey's SA was recently characterized by mass spectrometry. [37] Notwithstanding the increasing knowledge of DM protein fraction, up to the present no information is available for the primary structure of LF, one of the most important glycoproteins, that confers high hygienic qualities to DM [38] and presents an array of biochemical properties, including immuno-modulation, iron-binding, antioxidant, antibacterial and antiviral activities. [39]

Therefore, a better knowledge of donkey's milk components, particularly proteins, is a priority for many researchers. Mass spectrometry-based techniques, which arguably represent the core tools in proteomic analyses, are widely used to characterize milk proteins not only in native fresh but also in processed milk.

In the frame of our research line oriented to the characterization of the donkey's milk proteins and its genetic polymorphism, in this project it is reported for the first time the

direct sequence determination of the primary structure and the N-glycan composition of donkey LF by coupling RP-HPLC, ion exchange chromatography (IEC), enzymatic digestions and mass spectrometric analysis

LACTOFERRIN

Lactoferrin (LF) is a glycosylated globular protein first identified in 1939 in bovine milk ,[40] and subsequently separated from human milk.[41] Since 1960 LF has attracted increasing scientific interest due to its high concentration in human breast milk. Human milk is rich in lactoferrin, with a concentration around 6-8 mg/ml and 2-4 mg/ml, in colostrum and mature milk, respectively.[42, 43, 44, 45] Lactoferrin concentration in ruminant milk is 10-100 times lower than in human milk (in the range of 0.02-0.2 mg/mL).[46, 47] LF is present in the whey protein fraction of human and ruminant milks[48] with a net positive charge and pI in the range of 8.0-8.5.[49, 50] Lactoferrin (also called lactotransferrin), found in the mammalian milk, constitute, with the serotransferrins present in serum of all vertebrates and with the ovotransferrins characterized in the egg white of birds, a homogenous family of proteins: the transferrins. Lf is not only confined to mammalian milks, but is also a component of many external secretions such as saliva, tears, semen and mucosal secretions[51], and is further an important constituent of the neutrophilic granules of leucocytes.[52]

Lactoferrin's structure

The LF amino acid sequences of nine species, human, bovine, camel, buffalo, sheep, goat, pig, horse, and mouse are currently known. An appreciable range in the LF pair-wise sequence identities has been emphasized[53] and summarized in Table 1.[54] Lfs from cow, buffalo, goat, and sheep share over 90% (green square in the table) sequence identity, suggesting a group of extremely closely related proteins. However, mouse LF is the most different from other proteins with its sequence identity in the 63-67% range (violet square). Next most divergent proteins are human and equine LFs with their sequence identity in the 69-72% range of (blue square). Finally, pig and camel LFs have the averaged sequence identities of 71.78% and 74.25%, respectively (see Table 1). All together these findings suggest that LFs represent a highly conserved family of proteins.[53]

Table 1: Pair-wise sequence identities of the lactoferrin from different species compared by the "Align" tool of the UniProt database (http://www.uniprot.org/align/)

	Mouse P08071	Human P02788	Camel Q9TUM0	Cow P24627	Buffalo O77698	Goat Q29477	Sheep Q5MJE8	Pig P14632	Horse O77811
Mouse P08071		70.42	66.52	63.84	63.98	63.42	63.56	64.27	64.41
Human P02788	70.42		73.80	69.44	69.72	70.56	70.70	70.42	73.24
Camel Q9TUM0	66.52	73.80		75.42	76.13	75.57	76.27	74.58	75.71
Cow P24627	63.84	69.44	75.42		96.33	92.23	91.53	73.31	71.33
Buffalo O77698	63.98	69.72	76.13	96.33		91.95	91.24	73.16	71.33
Goat Q29477	63.42	70.56	75.57	92.23	91.95		97.60	73.73	71.19
Sheep Q5MJE8	63.56	70.70	76.27	91.53	91.24	97.60		74.45	71.89
Pig P14632	64.27	70.42	74.58	73.31	73.16	73.73	74.45		70.34
Horse O77811	64.41	73.24	75.71	71.33	71.33	71.19	71.89	70.34	

LF is the essential iron-binding protein in milk. Crystal structures are available for the Lfs of five species, [53] human, [55, 56] cow, [57] buffalo, [58] horse, [59] and camel. [60] As shown in Figure 1, LF is a single polypeptide chain, of about 80 kDa, folded into two globular lobes, referred to N- and C-terminal lobes, [61, 62] linked by a short α-helix, which gives LF additional flexibility. [63, 64] Each lobe consist of about 345 residues and contains two domains, referred to as N1, N2 and C1, C2. The disposition of the two domains in each lobe forms an interdomain pocket (cleft) in which the iron-binding site is located. [65, 66] The iron-binding is accompanied by synergistic binding of carbonate anions. Furthermore it can bind the Cu^{2+}, Zn^{2+}, and Mn^{2+} ions [67, 68] but it can also bind negatively charged molecules, such as DNA, [69] heparin, other glycosaminoglycan and lipopolysaccharides (LPSs). [70] This variable reactivity corresponds to the several suggested functions of this protein.
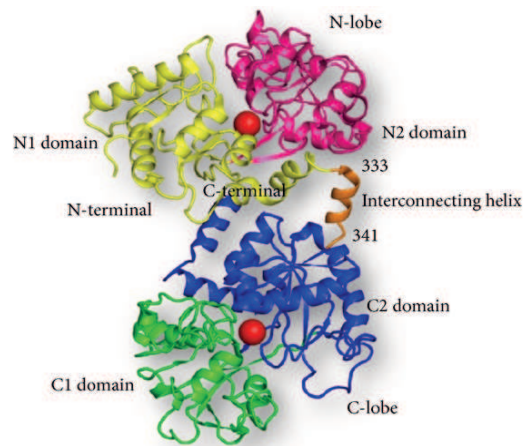
Figure 1: Schematic diagram of the bovine lactoferrin molecule (PDB code; 1BLF). The N1 and N2 domains are colored in yellow and pink, respectively, white the C1 and C2 domains are colored in green and blue, respectively. The interconnecting helix betwwen the lobes is colored in orange. The two iron atoms are shown as red spere.

Lactoferrin's functions

Lactoferrin (LF) shows an array of biological properties, such as antibacterial, [71] antiviral activity, [72, 73] antioxidant, a modulator of the immune and inflammatory responses, [74, 75] a growth regulator, [76] transcription factor, [77] regulator of the mitochondrial death signaling pathway, [78] and an iron-absorbing agent. [79]

Antimicrobial activity is mainly attributable to the ability of this protein to sequester with high affinity iron from pathogens, which are using this essential metal, and to retain the bound iron under acidic conditions. However, it was shown that some of the antimicrobial characteristics of LF are independent of its iron-binding mechanism [80, 81, 82] and are attributed to direct interference of LF with the bacterial cell surface. [83, 84, 85] LF exerts distinctive antimicrobial activities against a broad range of Gram-positive bacteria, Gram-negative bacteria, fungi, yeasts, viruses [86, 87, 88] and parasites. [89] Curiously, LF does not only have profound antimicrobial activity against pathogenic microbes, but was shown to promote the growth of beneficial bacteria, such as those belonging to the Lactobacillus and Bifidobacteria genera. [90] Furthermore, in vitro hydrolysis of human and bovine LF with pepsin elucidated that the N-terminus of LF, designated as lactoferricin H and B, which is distinct from the iron binding region, had stronger antimicrobial activity than the parental LF against various kinds of both Gram-positive and Gram–negative bacteria, with the exception of Bifidobacteria. [91, 92]

Most of the *antiviral activities* attributed to milk can be ascribed to LF alone. [93] The antiviral activity of LF is exerted by interfering with a cellular target, in the case of hepatitis B virus (HBV), HS-adapted Sindbis virus, Semliki Forest virus, cytomegalovirus (CMV), herpes simplex virus-1 (HSV-1), and HSV-2. [94, 95, 96, 97] In contrast, in the case of adenovirus, feline herpes virus (FHV-1), hepatitis C virus HCV, and HIV, the antiviral activity of LF is exerted by interfering directly with viral particles. [98, 99, 100] Furthermore, bovine and human LFs are potent inhibitors of human immunodeficiency virus (HIV) infection in vitro. [101 , 102 , 103] Harmsen et al. [104] demonstrated that among many proteins evaluated, LF was the only protein to prevent HIV-1 replication inside MT4 cells. Both Harmsen et al. and Puddu et al. [105] showed that the mechanism of LF action against HIV involves the interruption by LF of the adsorption of the virus to its target cells in an early phase of infection. LF can bind to the GPGRAF domain of the gp120 glycoprotein of HIV [106]. It is possible that the binding of LF to gp120 is responsible for its antiviral effects against HIV, because the binding of gp120 to the CD4 or chemokine receptors on the target cells plays an important role in the adsorption and entry of HIV into those cells. [107, 108,109] LF has also antiviral activity *in vitro* against rotavirus,[110] poliovirus[111] and Mayaro virus (MAYV) [112].

Lactoferrin has a unique *immunomodulatory action* on adaptive cellular functions, on both T and B lymphocytes and other immune cells, by promoting the maturation of Tcell precursors into competent helper cells and the differentiation of immature B-cells into efficient antigenpresenting cells (APCs). [113, 114, 115]

Lactoferrin peptides's functions

Bioactivities of milk are manly derived from intact proteins and peptides. However, a variety of bioactivities is exerted only after the digestion of proteins in the gastrointestinal tract. Protein digestion is initiated in the stomach by pepsin under acidic pH conditions. Then, the digesta are further hydrolyzed by pancreatic enzymes such as pepsin, trypsin and chymotrypsin and membrane peptidases resulting in peptides of various lengths. The size of bioactive sequences generally varies from 2 to around 20 amino acid residues, and some of the peptides are known to have multiple functions. [116] Bioactive peptides exert their bioactivities directly in the gastrointestinal lumen or at

peripheral organs after being absorbed at the intestinal mucosa. The mechanisms of intestinal absorption are mainly classified into 3 categories:

- ✓ *The carrier-mediated pathway*: the peptides transported into intestinal epithelial cell are generally hydrolyzed into amino acids by cytoplasmic peptidases, but certain peptides may resist digestion. Thus, some bioactive di- and tripeptides may be absorbed via this carrier-mediated pathway and exert their bioactivities at target organs; [117]
- ✓ *Transcytosi:* some oligopeptides, especially basic and/or hydrophobic polypeptides, may interact with the epithelial cell membrane surface and be transported by an intracellular vesicle present in intestinal epithelial cells. [118]
- ✓ *Paracellular diffusion*: larger oligopeptides as well as di- and tripeptides can pass through the pores between the intestinal epithelial cells by passive diffusion. [119], [120], [121] This paracellular pathway is non-digestive and is considered to play an important role in the absorption of bioactive peptides in intact forms.

The most of the LF activities are exerted before the gastrointestinal release of bioactive peptides. However, recent studies report as four bovine lactoferrin bioactive peptides, corresponding to peptides of sequences LIWKL, RPYL, LNNSRAP [122] and LRPVAA [123] exerted in vitro inhibitory effects on angiotensin I-converting enzyme (ACE) activity. Angiotensin I-converting enzyme (ACE) is one of the major regulators of blood pressure through two different reactions in the renin-angiotensin-aldosterone system (RAAS). Inhibition of ACE results in a lowering of blood pressure, since ACE acts as a catalyst in the conversion of the inactive angiotensin I into the powerful vasoconstrictor angiotensin II. Another peptidic system, the endothelin (ET) system, also has an increasingly recognized role in blood pressure regulation. In the ET system, the endothelin-converting enzyme (ECE) cleaves the biologically inactive intermediate termed big ET-1 to form ET-1 which has powerful vasoconstrictor and pressure properties. Two bovine lactoferrin bioactive peptides, corresponding to peptides of sequences GILRPY and REPYFGY, both exerted in vitro inhibitory effects on ECE activity. [124]

Lactoferrin (LF) is a glycosylated globular protein. Glycosylation renders the glycoprotein resistant to digestion and is critical for its biological functions, including pathogen decoy, prebiotic activities [125, 126] and nonspecific innate immune defense against invading pathogens in the gut by inhibition of the pathogen adhesion and infection [127, 128], respectively. The glycosilation profile of lactoferrin in human, bovine and goat has been described. [129, 130, 131, 132, 133]

The presence of glycans on human milk LF (hmLF) is long known. [134] The human lactoferrin (hLF, P02788) sequence contains three potential N-glycosilation sites, located at Asn 156, Asn 497 and Asn 624. All these three sites were found glycosilated [135] even if the third site (Asn 624) is mostly unglycosylated [136, 137]. For hLF 17 and 19 different compositions were identified in two different investigations. [129, 130] Because some of these compositions were identical, a gross total of 26 different compositions were found (Table 2). Most of the observed glycans are biantennary and triantennary complex-type structures highly fucosylated, including two neutral non-fucosylated hybrid/complex glycans, ten neutral fucosylated hybrid/complex glycans, one sialylated glycans and thirteen sialylated fucosylated glycans were found. Distribution of the N-glycans in the glycosilation sites was not determined because no glycopeptides were investigated.

In bovine lactoferrin (bLF, P24627), five of the five potential glycosilation sites at Asn 252, Asn 300, Asn 387, Asn 495 and Asn 564 were found N-glycosilated with 31 diffent compositions (Table 2) diffently distributed in the glycosilation sites (Figure 2).[131]
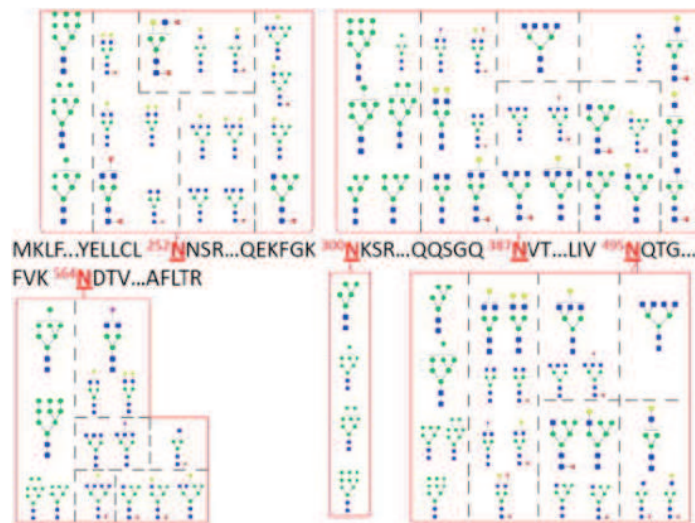

Figure 2: N-Glycans associated with each glycosite site for the bLF glycoprotein.

The bLF glycopeptides compositions found indicate that most of the glycans are high-mannose type N-glycans; however, fucosylated complex type are also observed. More detailed analysis shows that Asn 252, Asn 387 and Asn 495 are the most glycosylated sites on bLF as they yield the higher number of glycopeptides. Conversely, Asn 564 and Asn 300 were less glycosylated with significantly fewer glycopeptide products. The glycosilation compositions found include six high mannose glycans, eleven neutral non-fucosylated hybrid/complex glycans, twelve neutral fucosylated hybrid/complex glycans, two sialylated glycans and five sialylated fucosylated glycans.

32 different N-glycan compositions were found from the goat lactoferrin [133] (Table 2). To the best of our knowledge, there are no published reports about the actual occupancy of goat lactoferrin's glycosylation sites. Prediction studies reveal the presence of five potential sites (Asn 233, Asn 281, Asn 368, Asn 476 and Asn 545) just like in bovine lactoferrin. All togheter the N-glycan compositions found in goat lactoferrin reveal that most of the glycans identified are high-mannose, hybrid and complex type like the identified N-glycans of bovine milk lactoferrin. Actualy, five high mannose glycans, nine neutral non-fucosylated hybrid/complex glycans, six neutral fucosylated hybrid/complex glycans and twelve sialylated glycans, eight out of which with N-glycolyl neuramin acid (NeuGc), were found. The N-glycolyl neuramin acid (NeuGc), which is generally not found in humans, was found only in goat lactoferrin. It was not confirmed in bovine lactoferrin although in a recent study of Nwosu et al. (2012) [138] NeuGc was found as a costituent of the N-glycome in bovine milk (Table 2). Although N-acetylneuraminic acid (NeuAc) sialylation was observed in human and bovine milk samples, the NeuGc residue was only observed in bovine milk. To the best of our knowledge, this Nwous's study is the first MS based confirmation of NeuGc in milk protein bound glycans, as well as the first comprehensive N-glycan profile of bovine milk proteins.

Table 2: N-Glycan compositions of human, bovine and goat the lactoferrin. The last two columns are about the N-Glycome in human and bovine milk.

N-Glycan composition	N-Glycan Residue $(-H_2O)$	Human LF[1]	Bovine LF[2]	Goat LF[3]	N-Glycan in Human Milk[4]	N-Glycan in Bovine Milk[4]
High Mannose						
HexNAc$_2$Hex$_4$	1054.37		*	*		
HexNAc$_2$Hex$_5$	1216.42		*	*	*	*

Glycan	Mass					
HexNAc$_2$Hex$_6$	1378.48		*	*	*	*
HexNAc$_2$Hex$_7$	1540.53		*	*	*	*
HexNAc$_2$Hex$_8$	1702.58		*	*	*	*
HexNAc$_2$Hex$_9$	1864.63		*		*	*
Neutral Complex/Hybrid						
HexNAc$_2$Hex$_4$	1054.37				*	
HexNAc$_3$Hex$_3$	1095.40		*			
HexNAc$_3$Hex$_4$	1257.45		*			
HexNAc$_3$Hex$_5$	1419.50		*	*	*	*
HexNAc$_3$Hex$_6$	1581.56		*	*		*
HexNAc$_3$Hex$_7$	1743.61			*		*
HexNAc$_4$Hex$_3$	1298.48		*	*		*
HexNAc$_4$Hex$_4$	1460.53	+	*	*	*	*
HexNAc$_4$Hex$_5$	1622.58	*+	*	*	*	*
HexNAc$_4$Hex$_6$	1784.63		*	*		*
HexNAc$_5$Hex$_3$	1501.56		*			
HexNAc$_5$Hex$_4$	1663.61		*	*		*
HexNAc$_5$Hex$_6$	1987.71					*
HexNAc$_6$Hex$_3$	1704.63		*	*		*
HexNAc$_7$Hex$_8$	2717.98				*	
Neutral Complex/Hybrid Fucosylated						
HexNAc$_2$Hex$_3$Fuc$_1$	1038.38				*	
HexNAc$_2$Hex$_4$Fuc$_1$	1200.43				*	
HexNAc$_3$Hex$_3$Fuc$_1$	1241.45	+	*			
HexNAc$_3$Hex$_4$Fuc$_1$	1403.51	+	*			
HexNAc$_3$Hex$_4$Fuc$_2$	1549.57		*			
HexNAc$_3$Hex$_5$Fuc$_1$	1565.56		*	*	*	
HexNAc$_3$Hex$_5$Fuc$_2$	1711.62				*	
HexNAc$_3$Hex$_6$Fuc$_1$	1873.67		*			
HexNAc$_4$Hex$_3$Fuc$_1$	1444.53		*	*		*
HexNAc$_4$Hex$_3$Fuc$_2$	1590.59		*			
HexNAc$_4$Hex$_4$Fuc$_1$	1606.58	+	*	*	*	*
HexNAc$_4$Hex$_4$Fuc$_2$	1752.64		*			
HexNAc$_4$Hex$_5$Fuc$_1$	1768.64	*+	*	*	*	*
HexNAc$_4$Hex$_5$Fuc$_2$	1914.70	*+			*	*
HexNAc$_4$Hex$_5$Fuc$_3$	2060.76	*+			*	*
HexNAc$_4$Hex$_6$Fuc$_1$	1930.69			*	*	*
HexNAc$_4$Hex$_6$Fuc$_2$	2076.75				*	*
HexNAc$_4$Hex$_6$Fuc$_3$	2222.81				*	*
HexNAc$_5$Hex$_3$Fuc$_1$	1647.61		*			
HexNAc$_5$Hex$_3$Fuc$_2$	1793.67		*			
HexNAc$_5$Hex$_4$Fuc$_1$	1809.67			*		*
HexNAc$_5$Hex$_5$Fuc$_1$	1971.72					*
HexNAc$_5$Hex$_6$Fuc$_1$	2133.77	+				
HexNAc$_5$Hex$_6$Fuc$_2$	2279.83	*+				
HexNAc$_5$Hex$_6$Fuc$_4$	2571.95	+				
HexNAc$_6$Hex$_3$Fuc$_1$	1850.69					*
HexNAc$_6$Hex$_3$Fuc$_2$	1996.75				*	
HexNAc$_6$Hex$_3$Fuc$_3$	2142.81					*
HexNAc$_6$Hex$_4$Fuc$_1$	2012.75				*	
HexNAc$_6$Hex$_4$Fuc$_2$	2158.80				*	
HexNAc$_6$Hex$_6$Fuc$_2$	2482.91				*	
HexNAc$_6$Hex$_7$Fuc$_1$	2498.90	+				
HexNAc$_6$Hex$_7$Fuc$_3$	2791.02					*
HexNAc$_7$Hex$_3$Fuc$_2$	2199.83				*	
HexNAc$_7$Hex$_3$Fuc$_3$	2345.89				*	

		1	2	3	4
Sialylated Complex/Hybrid					
HexNAc3Hex5NeuAc1	1710.63		*	*	*
HexNAc3Hex5NeuGc1	1726.59		*		
HexNAc3Hex6NeuAc1	1872.65			*	*
HexNAc3Hex6NeuGc1	1888.65		*		
HexNAc3Hex7NeuGc1	2050.70		*		
HexNAc4Hex3NeuAc1	1589.57	*			*
HexNAc4Hex4NeuAc1	1751.62				*
HexNAc4Hex4NeuGc1	1767.62		*		
HexNAc4Hex5NeuAc1	1913.68	*+	*	*	*
HexNAc4Hex5NeuAc2	2204.77			*	*
HexNAc4Hex5NeuGc1	1929.67		*		*
HexNAc4Hex5NeuGc2	2236.76				*
HexNAc4Hex5NeuAc1NeuGc1	2220.77				*
HexNAc4Hex6NeuAc1NeuGc1	2382.82				*
HexNAc5Hex3NeuAc1	1792.65	*			
HexNAc5Hex4NeuAc1	1954.70				*
HexNAc5Hex4NeuGc1	1970.70				*
HexNAc5Hex5NeuAc1	2116.76				*
HexNAc6Hex3NeuAc1	1995.73				*
Sialylated Fucosylated Complex/Hybrid					
HexNAc3Hex4Fuc1NeuAc1	1694.60			*	
HexNAc3Hex5Fuc1NeuAc1	2018.71			*	
HexNAc3Hex6Fuc1NeuAc1	1856.66				*
HexNAc3Hex6Fuc1NeuGc1	2034.70		*		
HexNAc4Hex4Fuc1NeuAc1	1897.68		*		*
HexNAc4Hex5Fuc1NeuAc1	2059.74	*+	*	*	*
HexNAc4Hex5Fuc1NeuGc1	2075.73		*		*
HexNAc4Hex5Fuc1NeuGc2	2382.82				*
HexNAc4Hex5Fuc2NeuAc1	2205.79	*+		*	
HexNAc4Hex5Fuc3NeuAc1	2351.85	+		*	
HexNAc4Hex5Fuc1NeuAc2	2350.83	*+			
HexNAc5Hex4Fuc1NeuAc1	2100.76			*	*
HexNAc5Hex4Fuc1NeuGc1	2116.76		*		
HexNAc5Hex6Fuc1NeuAc1	2424.87	*			
HexNAc5Hex6Fuc2NeuAc1	2570.93	*+			
HexNAc5Hex6Fuc3NeuAc1	2716.98	*		*	
HexNAc5Hex6Fuc4NeuAc1	2863.04	*		*	
HexNAc6Hex4Fuc1NeuAc1	2303.84	+			
HexNAc6Hex7Fuc1NeuAc1	2790.00	*			
HexNAc6Hex7Fuc2NeuAc1	2936.06	*			
HexNAc6Hex7Fuc3NeuAc1	3082.12	*			
HexNAc6Hex7Fuc4NeuAc1	3228.17	*			
HexNAc7Hex3Fuc1NeuGc2	2667.95				*

1*T. Yu et al (2011);[129] +M. Barboza et al (2012)[130];

2*C. C. Nwosu et al (2010);[131]

3*A. Le Parc et al (2014);[133]

4*C. C. Nwous et al (2012).[138]

From the comparison among human, bovine and goat lactoferrin N-glycans (Table 3), high mannose N-glycans were confirmed only for bovine and goat lactoferrin. Moveover, human milk lactoferrin contains only 7,70% of neutral N-glycans. This value is lower than in bovine milk lactoferrin and goat milk lactoferrin (35,48% and 28,12% of all N-glycans respectively). The majority of the N-glycans in human lactoferrin shows sialylated fucosylated complex/hybrid and fucosylated complex/hybrid structures. Moreover, the total percentage of fucosylation in human milk lactoferrin (88,50%) is higher than in bovine milk lactoferrin and goat milk lactoferrin (38,72% and 34,38%) . Also the total percentage of sialylation in human milk lactoferrin (53,80%) is higher than in bovine milk lactoferrin and goat milk lactoferrin (6,45% and 37,50%). Furthermore, from the total percentage of sialylation in goat milk lactoferrin (37,50%), 25% is from sialylated complex/hybrid and sialylated fucosylated complex/hybrid with N-glycolyl neuramin acid (NeuGc), that, at the moment, was only found in goat lactoferrin while only 12.50% present N-acetylneuraminic acid (Neu5Ac) sialylation. Finally, the number of only sialylated N-glycans of human milk lactoferrin (3,80%) was slightly lower than that on bovine milk lactoferrin (6,45%) and goat lactoferrin (6.25% with N-acetylneuraminic acid NeuAc and 15.63% with N-glycolyl neuramin acid (NeuGc) sialylation).

Table 3 : Comparison among human, bovine and goat lactoferrin N-glycans type.

N-Glycan Composition	Human LF		Bovine LF		Goat LF	
	N-Glycan found	Percentage %	N-Glycan found	Percentage %	N-Glycan found	Percentage %
High Mannose	0	0	6	19,35	5	15,63
Neutral Complex/Hybrid	2	7,7%	11	35,48	9	28,12
Neutral Fucosylated Complex/Hybrid	10	38,5%	12	38,72	6	18,75
Sialylated Complex/Hybrid	1	3,8%	2	6,45	7	21,88
			2 0*	6,45 0*	2 5*	6,25 15,63*
Sialylated Fucosylated Complex/Hybrid	13	50%	0	0	5	15,63
			0 0*	0 0*	2 3*	6,25 9,38*

*Glycans containing N-glycolyl neuramin acid (NeuGc)

GLYCOSYLATION

Glycosylation is increasingly recognized as a common and biologically significant post-translational modification (PTM) of proteins. It is estimated that > 50 % of mammalian proteins are glycosylated. [139] Therefore, glycosylation can be considered the most abundant PTM of proteins.

Protein glycosylation is involved in many biological processes, such as immune function, cellular division/migration/adhesion, host–pathogen interactions and enzyme catalysis. [140, 141] Glycosylation is also recognized as an important element of disease pathophysiology (e.g., cancer, autoimmune, diabetes, Alzheimer's, hematologic disorders, and allergies) and numerous investigators and organizations like the Human Disease Glycomics/Proteome Initiative (HGPI) [142] seek to translate global glycosylation profiles into diagnostic biomarkers.

Monosaccharides are the basic structural units of glycans

A glycoprotein is a glycoconjugate in which a protein carry an array of monosaccharide or oligosaccharide units (the glycone), which are generically referred to as "glycans", covalently linked to a polypeptide backbone (the aglycone).

An oligosaccharide that is not attached to an aglycone possesses the reducing power of the aldehyde or ketone group in its terminal monosaccharide unit. This end of a sugar chain is therefore often called the reducing terminus or reducing end, terms that tend to be used even when the sugar chain is attached to an aglycone and has thus lost its reducing power. Correspondingly, the outer end of the chain tends to be called the nonreducing end.

Although several hundred distinct monosaccharides are known to occur in nature, only a small number of these are commonly found in animal glycans. They are listed below, along with their standard abbreviations

- ✓ *Pentoses*: five-carbon neutral sugars, i.e. D-xylose (Xyl);
- ✓ *Hexoses*: six-carbon neutral sugars, i.e. D-glucose (Glc), D-galactose (Gal), and D-mannose (Man);
- ✓ *Hexosamines*: hexoses with an amino group at the 2-position, which can be either free or, more commonly, N-acetylated, i.e. *N*-acetyl-D-glucosamine (GlcNAc) and *N*-acetyl-D-galactosamine (GalNAc);

✓ *Deoxyhexoses*: six-carbon neutral sugars without the hydroxyl group at the 6-position, i.e. L-fucose (Fuc);

✓ *Uronic acids*: hexoses with a negatively charged carboxylate at the 6-position, i.e. D-glucuronic acid (GlcA) and L-iduronic acid (IdoA);

✓ *Sialic acids*: family of nine-carbon acidic sugars (generic abbreviation Sia), of which the most common is *N*-acetylneuraminic acid (Neu5Ac, also sometimes called NeuAc or historically, NANA).

This limited set of monosaccharides dominates the glycobiology of more recently evolved (so-called "higher") animals, but several others have been found in "lower" animals (i.e. tyvelose), bacteria (i.e. keto-deoxyoctulosonic acid, rhamnose, L-arabinose, and muramic acid), and plants (i.e. arabinose, apiose, and galacturonic acid). A variety of modifications of glycans enhances their diversity in nature and often serve to mediate specific biological functions. Thus, the hydroxyl groups of different monosaccharides can be subject to phosphorylation, sulfation, methylation, O-acetylation, or fatty acylation. Although amino groups are commonly N-acetylated, they can be N-sulfated or remain unsubstituted. Carboxyl groups are occasionally subjected to lactonization to nearby hydroxyl groups or even lactamization to nearby amino groups.

In Table 4 it is shown a monosaccharide symbol set, modified from the first edition of *Essentials of Glycobiology,* [140] which has also been adopted by several other groups interested in presenting databases of structures (i.e. the Consortium for Functional Glycomics), names and monoisotopic mass for common monosaccharides. Each monosaccharide class (i.e. hexose) has the same shape, and isomers are differentiated by color. The same shading/color is used for different monosaccharides of the same stereochemical designation (i.e. Gal, GalNAc, and GalA). To minimize variations, sialic acids and uronic acids are in the same shape, and only the major uronic and sialic acid types are represented. When the type of sialic acid is uncertain, the abbreviation Sia can be used instead. Only common monosaccharides in vertebrate systems are assigned specific symbols. All other monosaccharides are represented by an open hexagon or defined in the figure legend. Unless otherwise indicated, all of these vertebrate monosaccharides are assumed to be in the D-configuration (except for fucose and iduronic acid, which are in the L-configuration). All glycosidically linked monosaccharides are assumed to be in the pyranose form, and all glycosidic linkages

are assumed to originate from the 1-position (except for the sialic acids, which are linked from the 2-position).

Table 4: Symbols, names and monoisotopic mass for common monosaccharides, found in eukaryotic system. The monosaccharide symbol set from Essentials of Glycobiology

Symbol	Name	Mass	Symbol	Name	Mass
⬤	Galactose (Gal)	162.05 Da	★	Xylose (Xyl)	132.04 Da
◻	N-acetylgalactosamine (GalNAc)	203.08 Da	◆	N-acetylneuraminic acid (Neu5Ac)	291.10 Da
◺	Galactosamine (GalN)	161.07 Da	◇	N-Glycolylneuraminic acid (Neu5Gc)	307.1 Da
⬤	Glucose (Glc)	162.05 Da	◈	2-Keto-3-deoxynononic acid (Kdn)	250.07 Da
◼	N-Acetylglucosamine (GlcNAc)	203.08 Da	▲	Fucose (Fuc)	146.06 Da
◺	Glucosamine (GlcN)	161.07 Da	⬡	Glucoronic acid (GlcA)	176.03 Da
⬤	Mannose (Man)	162.05 Da	⬡	Iduronic acid (IdoA)	176.03 Da
◼	N-acetylmannosamine (ManNAc)	203.08 Da	⬡	Galacturonic acid (GalA)	176.03 Da
◺	Mannosamine (ManN)	161.07 Da	⬡	Mannuronic acid (ManA)	176.03 Da

Most of eukaryotic glycoprotein are produced in the secretory pathway (in the endoplasmic reticulum (ER) and the Golgi apparatus), whose enzymes are responsible for the biosynthesis of multiple classes of glycoproteins. Golgi- and ER-resident glycosyltransferases transfer sugars from nucleotide sugar donors to glycoprotein substrates as they traffic through the secretory pathway. Glycosyltransferases are membrane proteins localized at specific subsites within the secretory pathway. The localization of these enzymes dictates the order in which glycosylation events occur. In this way, the secretory pathway serves as an assembly line for glycoprotein biosynthesis. Unlike the production of nucleic acid and peptide polymers, glycan biosynthesis is not template directed. The nontemplated nature of glycan biosynthesis results in a key characteristic of glycosylation: heterogeneity. Heterogeneity occurs both at the level of the occupancy of potential glycosylation sites as well as at the diversity of structures present at each site. The term "glycoforms" refers to different isoforms of a protein that vary with respect to the number or structure of attached glycans.

N-glycosilations

There are two major classes of protein glycosylations: N-linked and O-linked.

N-linked glycans are carbohydrate structures attached to the nitrogen atom of asparagine. These structures are usually found when the asparagine residue occurs in the trypeptidic consensus sequence N-X-S/T, where X can be any amino acid but proline.[143] The synthesis of these glycan structures take place in the Endoplasmic Reticulum (ER) where an oligosaccharide precursor is assembled, consisting of two residues of N-Acetylglucosamine (GlcNAc), nine mannose (Man) residues and three glucose (Glc) residues linked to dolichol phosphate lipid.[144] Saccharide residues are then removed and added in order to achieve a specific, mature glycan structure.

Mature N-linked glycans, which maintain in common only the core structure (Manα1-6 (Manα1-3) Manβ1-4 GlcNAcβ1-4 GlcNAcβ1-Asn), can be further divided into three subclasses (Figure 3):

- ✓ Complex tructures
- ✓ High mannose structures
- ✓ Hybrid structures

The complex structures contain one o more elongations from the core structure referred as *antennae* originating from the α-linked mannoses. These antennae are usually composed of *N*-Acetylglucosamine and galactose, and may be terminated with sialic acids (commonly *N*-Acetylneuraminic acid in mammals). The high mannose type is, instead, composed by the common core structure and a varying number of mannose residues. In mammals, structures containing 5-9 mannose residues are the most common, while other organisms as yeast have shown much larger structures. The hybrid structures have properties of both complex and high mannose types, containing both the complex antennae branching out from the core-linked Manα1-3 and oligomannoses, branching out from Manα1-6. Hybrid structures may also be terminated with sialic acids.

The core structure might undergo modifications, such as the addition of a fucose to the innermost *N*Acetylglucosamine. Such core modifications are usually seen in complex or hybrid structures. Moreover, several deviations from the common structures have been observed in many organisms.
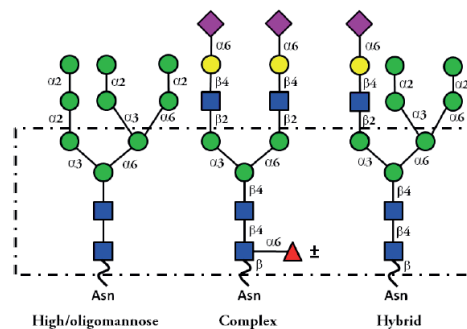
Figure 3: The three types of N-glycans. The common core structures (Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn) are framed in the dashed rectangle. See table 1 for symbols representing the different monosaccharide components. From left to right is represented: high/oligomannose; complex and hybrid. A unit of fucose, represented by the red triangle, is added to some Nglycans.

Furthermore depending on the number of antennae, the structure may be called monoantennary, biantennary, triantennary etc. The biantennary structures are referred to *G* structures, since they were first identified on Immuglobulin G. Depending on the number of galactoses in the antennae, the structures are called either G0, G1 or G2. When naming the high mannose structures, the glycan is referred to *Mannose* proceeded by the number of mannose residues in the structure, including the core. The hybrid structures are referred to as the highest number of mannose residues in a high mannose structure proceeded by the additional monosaccharides.

A residue may be observed in a glycosylated and non-glycosylated form, this is referred to as macroheterogeneity, or degree of site occupancy. A variety of different glycoforms may be found to occupy the same glycosylation site of a given protein, which is referred to as microheterogeneity. In certain diseases, the macro- and microheterogeneity are affected, producing either over-/underglycosylation, or variances in the observed glycoforms. Changes in glycosylations have been associated with a certain number of diseases, such as breast cancer, ovarian cancer, prostate cancer [145, 146, 147, 148] and rheumatoid arthritis. [149] These changes may be used as biomarkers when screening for these diseases.

O-glycosylations

The other major form of protein glycosylation, O-linked glycosylations, are attached to hydroxyl group of serine and threonine residues, but also rarely to tyrosine. Instead of one common core structure, eight different cores have been identified for the O-linked glycans, all containing a GalNAc residue as the innermost monosaccharide. [150] Since the work featured in this report is mainly focused on the N-linked glycans, the O-glycans

will not be discussed in greater detail. However, it will be mentioned that O-linked glycosylation cores are composed of two-three monosaccharides. The O-linked glycans are often found in great numbers adjacent to one another.

ANALYSIS OF GLYCOPROTEINS BY MASS SPECTROMETRY

The most commonly employed technique for the analysis of protein glycosylation is high-performance liquid chromatography (HPLC) coupled to mass spectrometry (MS). Although recent advances in mass spectrometry have made large-scale identification of proteins feasible, it is still very challenging to analyze protein glycosylation. This is due to the fact that glycopeptides often constitute a minor portion of the total peptide mixture, that signal intensity of glycopeptides is often low compared to nonglycosylated peptides and that the signal is often suppressed in the presence of other peptides. [151,152] For this reason, glycopeptide enrichment and separation is of main importance when performing glycoproteomic studies.

Sample Enrichment

LC-MS approaches to study protein glycosylation can be categorized as glycoprotein- or glycopeptide-based analysis. The former begins with enrichment of glycoproteins followed by protein digestion and LC-MS analysis, while in the latter, glycoproteins are initially digested and the resulting mixture is enriched at the glycopeptide level. Moreover, glycosylation-specific preparative strategies are less defined because of high chemical heterogeneity of glycans derived from diverse monosaccharide building blocks, anomeric configurations, branching, and elaboration by other chemical moieties (i.e. acetylation and sulfation). When sample amounts are non-limiting, combining different techniques in multidimensional formats provide the greatest overall experimental dynamic range.

Hydrophilic Interaction Liquid Chromatography (HILIC)

In 1990, Alpert [153] introduced the term HILIC for the normal-phase chromatographic mode with a polar stationary phase and a less polar mobile phase. However, in contrast to classical normal-phase chromatography, the mobile phase in HILIC comprises a

mixture of a water-miscible organic solvent (usually acetonitrile) with a certain amount of water (45% or less). The popularity of HILIC is related to the opportunity to separate polar compounds, compatibility of mobile phases based on acetonitrile, and a volatile pH adjusting reagents, with mass spectrometric detectors and its orthogonality to RP HPLC. [154],[155] The mechanism of HILIC with respect to different stationary phases, [156],[157],[158],[159] efficiency, [160] retention and selectivity [161],[162] is still under thorough investigation. [163] However, the diverse nature of available HILIC stationary phases makes the practical application of a particular stationary phase to a given compound classes difficult [164] and optimal separation conditions are usually selected based on trial and error.

HILIC provides a versatile tool for enrichment of glycopeptides before mass spectrometric analysis, particularly when used for solid phase extraction (SPE), or in combination with other chromatographic resins or ion-pairing reagents.

In this project HILIC was used as a solid phase extraction method of enrichment for glycopeptides in order to reduce the complexity of the protein digestes. Poly (2-hydroxyethyl) aspartarmide, commercially known as PolyHydroxyEthyl A (PolyLC), was used as stationary phase (Figure 4).
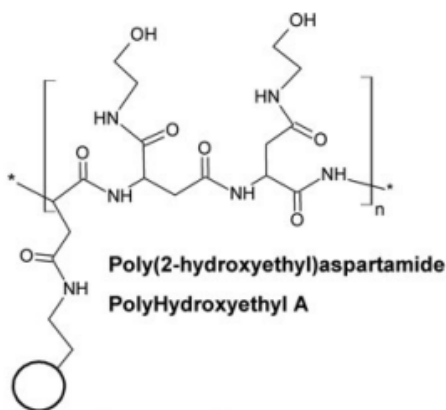

Figure 4: PolyHydroxyEthyl A (PolyLC)

Titanium dioxide (TiO$_2$)

Titanium dioxide (TiO$_2$) was originally proposed for phosphopeptide enrichment from proteolytic digests [165] can also be used for glycopeptide enrichment.

TiO$_2$-chromatography is highly selectivity for sialylated glycoproteins, as reported by Larsen *et al.*. [166] The binding-mechanism of TiO$_2$ is not precisely understood, but is

theorized to involve interaction between oxygen-containing groups such as the carboxyl groups of SAs and TiO_2 as shown in Figure 5.
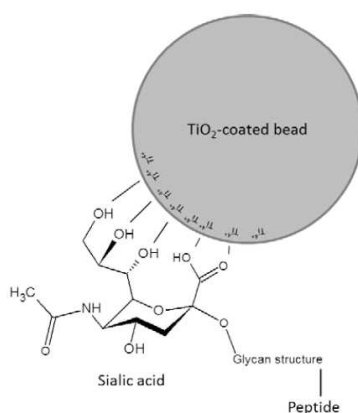

Figure 5:Possible binding mechanisms for sialylated glycopeptides to TiO₂ through multi dentate bindings between the O-atoms and Ti-atoms.

TiO_2 is highly selective for both phosphorylated peptides and sialylated glycopeptides, and if one wished to enrich solely for glycopeptides, especially if the sample has a high content of phosphoproteins, it is imperative to dephosphorylate the sample prior to enrichment. TiO_2 has successfully been used to enrich for sialylated glycopeptides from different biological samples of different levels of complexity. [167,168,169,170] Different modes of TiO_2-chromatography, including the use of an online coated column [171] to separate analytes,[172] or by coated magnetic beads, [173] are used. Throughout the present work, the coated magnetic beads are used.

HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)

The high-performance liquid chromatography, more simply known by the acronym HPLC, allows to perform separations of complex mixtures, and to determine the quantitative composition. The principle of separation of the different components present in a mixture is based on the formation of a pseudo-equilibrium between each component of the sample to be analyzed, the stationary phase, consisting of porous regular microparticles, and the liquid mobile phase flowing between the particles of the stationary phase. The particles of the stationary phase usually have small diameters, comprised between 3 and 10 μm, which thus confer a high efficiency. The eluent flows in the column thanks to the pressure that is applied by a pump to the column head. In this way not only the chromatography is faster, but the process of separation of the

different components takes place also through a larger number of theoretical plates, which results in better resolution. The most common and widespread type of liquid chromatography for the separation of protein and peptide mixtures is the reverse-phase chromatography (RP-HPLC), which adopts a non-polar stationary phase and a mobile polar phase. The apolar stationary phase consists of more or less long (C4, C12, C18) alkyl chains, linked to small spheres of silica. Regarding the mobile phase, in the RP-HPLC, in general, it is not used a single solvent but a mixture of two or more different solvents, whose flow rate is regulated by the respective pump. This makes possible to work in two different chromatographic conditions:

✓ isocratic conditions, where the same mobile phase composition is maintained from the beginning to the end of the analysis;

✓ conditions of gradient elution, in which two o more eluents are mixed in order to obtain a mobile phase, whose composition and consequently its polarity is decreasing in time.

Among the solvents used in the RP-HPLC there are water as polar solvent and methanol or acetonitrile as apolar organic solvents.

When a separation is effected with elution gradient the elution of the components is carried out starting from a composition of the mobile phase rich in the more polar solvent and, thereafter, increasing over the time the solvent with non-polar characteristics. In this way the more polar components of the mixture are eluted first, and the more apolar ones, which have a greater affinity for the organic eluent, later. The main components of a HPLC system are (Figure 6):

✓ containers for solvents, constituting the mobile phase, with attached a degassing system;

✓ pumps which generate high pressures to push the mobile phase in the system and which must give a flow as stable and reproducible as possible to avoid the background noise. Among the types of pumps, those most commonly employed are piston pumps, constituted by a small cylindrical chamber which is filled and empted by the movement of a piston. Such pumps have the ability to generate high output pressures (> 10000 psi), rapid adaptability to the change of the gradients in the course of the analysis.

- ✓ system for the introduction of the sample (sample loop) that allows the injection in the system of reproducible volumes, even very small (1-10μl), os sample. This device are equipped with interchangeable loop of variable capacity, usually from 1μL to 1 mL;

- ✓ column in which the separation of the analytes occurs; the HPLC columns can have different dimensions and characteristics depending on the type of analyses to be performed, on the system used, the type of detector and, not least, the amount of sample available. Thus, for example, the length of the columns can vary from 10 to 50 cm and the inside diameter from 4 (analytical column) to 10 mm (semipreparative columns), generally packed with particles of diameter varying from 5 to 10 µm. Lower internal diameters (0.0075-2.1 mm, microbore and narrowbore) chromatographic columns of 3-50 cm in the length packed with particles of diameter from 3 to 5 µm, are generally used in the case of small quantities of sample available. In the Ultra High Perfomance Liquid Chromatography (UPLC) the columns are packed with particles of diameter of 1.8-2.0 µm. In these types of column the efficiency of the separation increases together with the high pressure that is required for a reasonable flow rate of the mobile phase;

- ✓ detector: the components of a mixture once separated are sent to the detection system, which generates a signal when the analyte is eluted. Universal and highly sensitive detectors for the HPLC do not exist and therefore the detection system used depends on the needs dictated by the nature of the sample. The detectors most widely used for liquid chromatography are based on the measure of the absorption of ultraviolet or visible light. For example the detection of proteins is generally carried out at 220-224 nm. On the other hand a particularly sensitive and versatile detector is represented by a mass spectrometer with electrospray ionization, which today is widely used in proteomic studies.
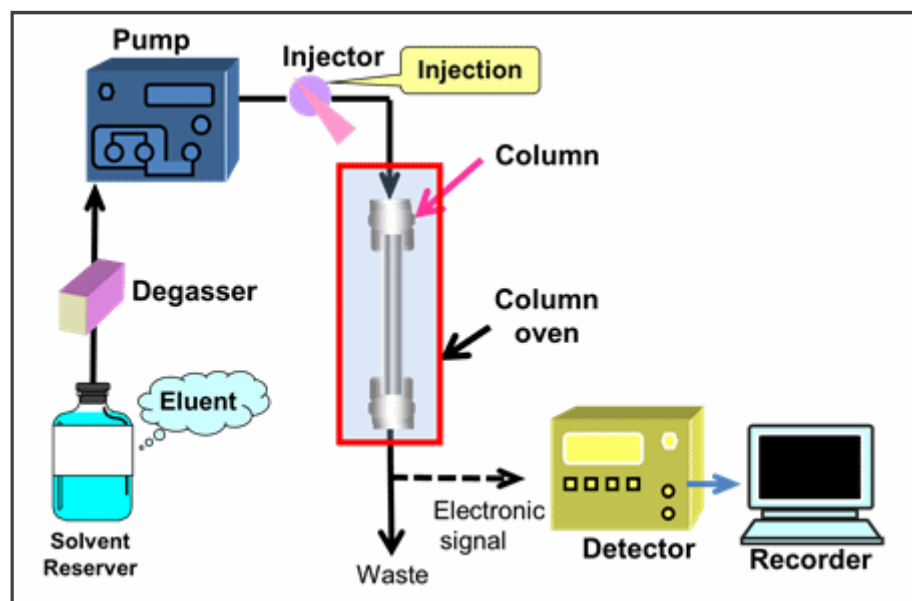
Figure 6: Main components of a system HPLC.

MASS SPECTROMETRY

Mass spectrometry is an analytical technique based on the ionization of a molecule and the subsequent separation on the basis of their different mass/charge ratio (m / z) and detection of the ions produced. At the end the technique allows to obtain the mass spectrum, a graph representing the relative abundance of the ions according to their m/z ratio.

A mass spectrometer is basically constituted by four main elements (Figure 7):

- ✓ the source, in which the ionization of the sample occurs;
- ✓ the analyzer, where there is a separation of the ions produced in the source according to their m/z ratio;
- ✓ the detector, where the separated ions are detected;.
- ✓ the vacuum system, whose task is to keep the various parts of the instrument under vacuum, the presence of which (the pressure is around 10^{-6}-10^{-8}Torr) is needed primarily to avoid the collision of the gas phase ions with the atmospheric gases.
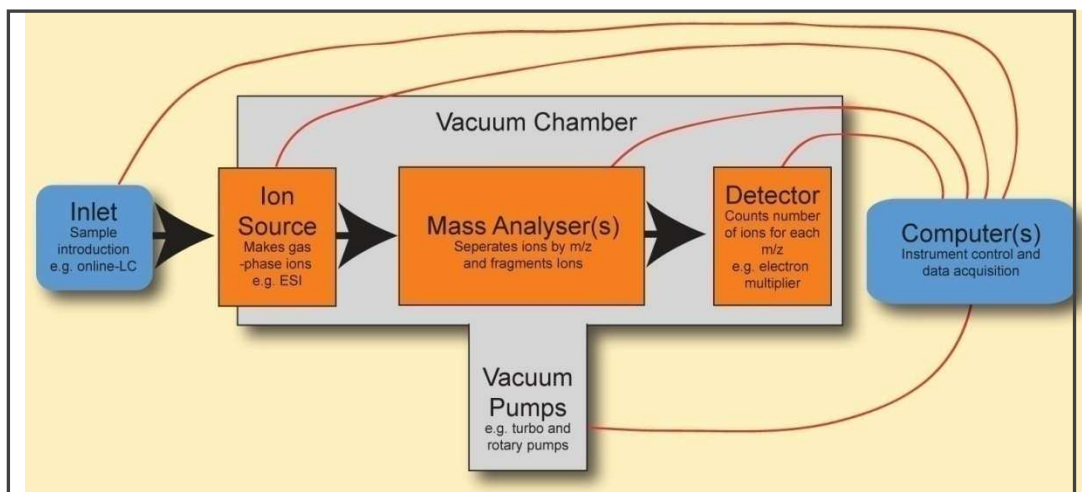
Figure 7: Block diagram of a mass spectrometer.

The ability to perform analysis to exceptionally high levels of sensitivity combined with a low error in the mass measurement has made mass spectrometry particularly suited for the study of protein structure.

Ion sources

From a historical point of view the first sources for ionization of the sample were the electron impact (EI) and the chemical ionization (CI), which were suitable only for molecules with low molecular weight and low polarity, which could be transferred easily in the gas phase. Such sources, in addition to ionize the sample, provoked a more or less extended fragmentation. Over the years, in order to extend the applicability of this technique to samples with higher molecular weights and/or higher polarity, new types of sources (FAB, thermospray etc), have been introduced. However, only in the late '80, thanks to the introduction of two new methods for desorption ionization, known as MALDI (Matrix-Assisted Laser Desorption/ Ionization) and ESI (Electrospray Ionization), mass spectrometry has assumed an important role in the study and characterization of biomolecules. In particular, the MALDI mass spectrometry is used for the characterization of relatively simple mixtures of peptides or proteins, while the Electrospray ionization, which can be coupled with liquid chromatography, can be used for the study of even very complex protein or peptide mixtures.

Matrix Assisted Laser Desorption/Ionization (MALDI)

The concept of Matrix Assisted Laser Desorption/Ionization Mass Spectrometry, or MALDI-MS was first introduced in 1985 by Hillenkamp and coworkers.[174] Before the introduction of MALDI, methods such as fast atom bombardment and plasma ionization were the methods of choice for the analysis of large molecules such as intact proteins. Tanaka *et al*. in 1988 [175] showed that MALDI-MS proved efficient of analyzing such molecules such as intact proteins. In MALDI, the analyte is co-crystallized in a matrix, constituted by a solid organic compound. The matrix/analyte crystal layer, deposited on a metal plate, is irradiated by a laser pulse, thereby protecting the analyte from degradation. Upon irradiation, the matrix absorbs energy, which causes both the matrix molecules and the analytes to disperse into the gas phase. The method of ionization is not fully understood, however several theories exist.[176] The choice of the matrix is highly dependent on the nature of the analyte, i.e. some matrices are best suited for molecules such as peptides, whereas others matrices are best suited for molecules such as proteins, modified peptides.

Electrospray ionization (ESI)

The electrospray ionization (ESI), is a soft ionization technique, since it does not produce fragmentation of the sample. This ionization technique has assumed an important role in the field of mass spectrometry not only for the ability to bring into phase gas ionized macromolecules of biological origin but also because it is the ideal interface for the on line coupling with a liquid chromatographic system (RP-HPLC/ESI-MS). In such cases the components eluted from the chromatographic column are directly sent to the ESI source and analyzed by mass spectrometer. Electrospray mass spectrometry (ESI-MS) allows to obtain single- and multiple charged ions which are thus sent to the analyzer and to the detection system.

The protein solution is introduced into the source through a capillary tube of silica and it reaches the end of a metal needle. Inside the ionization chamber, between the metallic tip of the needle and a counter electrode, it is present a strong electric field (3-5 kV) that disperses the solution emerging from the needle into an aerosol of droplets with a high charge concentration: the spray. Sometimes, a stream of nitrogen suitably heated, promotes the desolvation of the droplets of the spray. The generally used solvent is water mixed with an organic solvent (acetonitrile, methanol, or propanol) and small

amounts of or an acid (trifluoroacetic acid, acetic acid or formic acid) or a base (ammonia solution) to facilitate the ionization of the sample and the formation, respectively, of positive or negative ions.

The mechanism through which the ions are desolvated starting from the charged droplets of solution has not yet been completely clarified; in this regard several models have been proposed, including a qualitative model compatible with the mechanisms proposed by Smith, Fenn and Röllgen.[177, 178] In this model, the first step consists in the formation of micro-droplets whose dimensions are related to their surface tension; the hot gas stream causes the desolvation of these micro-droplets, tending to bring together the charged molecules. When the Coulomb repulsion equals the surface tension of the droplet (Rayleigh limit), this explodes producing smaller droplets (nano-droplets), [179] that are subject to further desolvation (Figure 8).
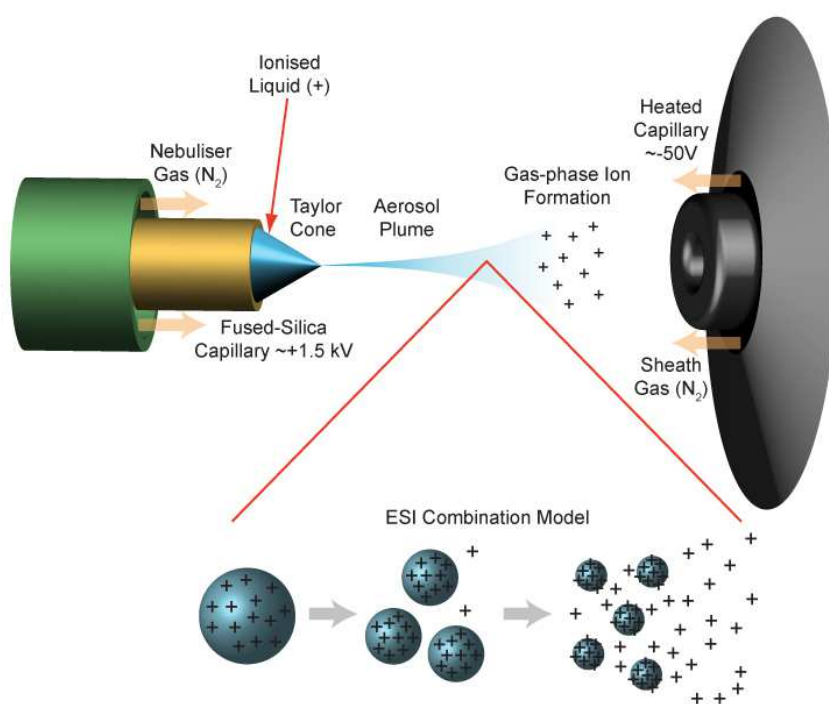
Figure 8: ESI source and model of formation of ions.

Only a part of the formed ions, passing through a small orifice, arrives in a pre-chamber which is located at a pressure of 10^{-1}-10^{-2} Torr. Subsequently a series of electrostatic lenses (skimmers) focus the ion beam that reaches the analyzer (10^{-6} - 10^{-7} Torr) where separation takes place based on the value of the m/z ratio. The formation of multiple charged ions allows to determine the molecular weight of molecules with high mass

also working with analyzers that have limited mass range and, therefore, it makes this ionization method an excellent tool for the analysis of peptides and protein. A typical ESI spectrum of positive ions consists of a set of peaks, each of which it is generated from the analyte that has linked a specific number of protons. The mass spectrometer measures mass/charge ratios (m/z) of each peak of the series obtained by:

$$m = \frac{M + nH}{n}$$

where **M** is the molecular mass of the analyte (eg. a protein), **n** is the number of protons, that is, the positive charges, and **H** is the mass of the proton (1.008 Da). Consequently, if the value of n for all ions of the series is known, from the mass/charge ratios measured it is possible to calculate the molecular weight of the sample:

$$M = n(m\text{-}H)$$

To determine n we consider two consecutive peaks of the series and then we have:

$$m_2 = \frac{M + nH}{n} \qquad m_1 = \frac{M + (n+1)H}{n+1}$$

where **m₁** is the peak to the lower mass and **m₂** that to greater mass. By solving this system of two equations one finds that the n charge of **m₂** is given by:

$$n = \frac{m_1 - H}{m_2 - m_1}$$

Since, as already said, each peak of a series differs from the next one for a proton, the "data system" identifies the charge associated with each peak, calculates the molecular mass from each peak and, finally, the molecular weight of the analyte. The series of multi-charged protein ions generated in the source is mainly related to the protonation of basic sites of molecules. In general, in a protein, the number of basic amino acid residues determines the maximum number of protons that the molecule can take. For small molecules there is a precise correlation between the number of basic sites present in the structure and multi-charged ions that are observed in spectrum. As the size increases, this correlation is not so rigorous, because, depending on the particular conformations that the molecule can assume, some of the basic sites will be located inside the protein itself, and accordingly will be protonable with difficulty. This explains why for protein of high molecular weight, the attack on the basic sites is

closely related to the conformation that the protein assumes in solution under the experimental conditions (pH, temperature, presence of denaturing agents).

ESI mass spectrometry constitutes a particularly powerful and versatile detector for the classic techniques of chromatographic separation such as liquid chromatography (HPLC). The set of data obtainable using the HPLC/ESI-MS technique consists of a series of mass spectra that are acquired one after another. The abundance of the various ions in each spectrum can be added together to obtain the total ion current (TIC), that is, a chromatogram whose appearance is quite similar to that obtained by a conventional chromatographic detector. Each chromatographic peak present in TIC represents one or more eluted compounds, which can be identified by its relative mass spectrum.

Mass analyzers

The ion source can be interfaced with different mass analyzers, the most commonly used are quadrupole (Q), ion trap (IT), time-of-flight (TOF) and Orbitrap. These mass analyzers show differences both in principles of operation and performance and can be used alone or coupled with other analyzers to improve performance (Table 5).

<u>Quadrupole (Q)</u>

A quadrupole mass filter consists of four precisely parallel rods with a hyperbolic or circular profile, connected in pairs in an angle of 180° thereby able to form a "tube" of electric fields. The ions fly between the rods and their motions are controlled by direct current (DC) voltage and radiofrequency (RF) voltage applied on the opposite pairs of rods. Each opposite pair of rods has same direct current (DC) voltage and superimposed radio-frequency (RF) potential that generate the electric field. The combination of DC and RF determines which ions with certain m/z value can pass through the quadrupole field. Other ions, which do not satisfy the chosen DC and RF values, collapse into the quadrupole. Changing the DC (U) and RF voltage (V), maintaining constant the 2U/V ratio, allows the whole MS range to be scanned. Another application of a quadrupole is ion-focusing, which can be achieved by setting the DC value to 0. In this mode, quadrupole can be used as collision cells in tandem instruments.

Quadrupole ion trap(QIT)

A quadrupole ion trap consists of three electrodes, one ring electrode and two electrodes hyperbolic in shape (end caps). Due to their similarities, an ion trap can be considered as a three-dimensional variant of a quadrupole mass analyzer. Ions are trapped by DC and RF voltages applied to the electrodes, and ions with certain m/z values can be permanently trapped or ejected from the ion trap to the detector. This trap can be used for multiple step tandem mass spectrometry.

Compared to conventional quadrupole mass analyzer, an ion trap has advantages due to the possibility of performing MS/MS measurements with a single analyzer.

Linear quadrupole ion trap(LTQ)

The production of quadrupole ion traps requires a high level of quality control due to the complicated shape of the electrodes. The introduction of a linear ion trap was mainly aimed at simplifying the production of ion traps. Apart from this, linear ion traps have a greater ion capacity and faster scan rate. A linear ion trap consists of a quadupole mass filter and two additional electrodes as end caps. Ions are trapped radially by the same principle as in a quadrupole, and static electrical potential is applied to the end caps to trap the ions axially. The frequency of oscillation depends on the m/z value, and certain RF values bring certain ions into resonance. This excitation can be used for selective ion ejection, scanning a m/z range or to perform fragmentation by collision with a neutral gas within the trap. Due to its multipurpose characteristic, linear ion traps are often used in hybrid tandem mass spectrometers. As mass analyzer, linear ion traps are relatively fast and sensitive but have low accuracy because the low resolution that can be achieved.

Time-of-flight (TOF)

Time-of-flight (TOF) analyzer is the mass-analyzer often coupled to MALDI. The TOF analyzer separates ions after their initial acceleration by an electric field, according to their velocities when they drift in a field-free region called "a flight tube" between the source and the detector. Thus, m/z ratios are determined by measuring the time that ions take to move through the flight tube. The linear TOF analyzer has poor mass resolution,

because ions of the same m/z ratio but with different kinetic energy reach the detector at slightly different times. This can be improved by introducing an electrostatic reflector behind the field-free region opposed to the ion source, which is termed as reflectron. The reflectron corrects the initial kinetic energy dispersion of the ions leaving the source with the same m/z ratio. Ions with higher kinetic energy and hence with higher velocity will penetrate the reflectron more deeply, i.e., will spend more time in the reflectron, with respect to ions of the same m/z but with lower kinetic energy. Thus, the slower ions will catch up and reach the detector at the same time as the faster ions of the same m/z. However, the reflectron increase the mass resolution at the expense of sensitivity and limited mass range.

Two successive TOF analyzers can be applied, called tandem TOF mass spectrometry (TOF/TOF MS). In tandem mass spectrometry, the first analyzer selects, isolates a precursor ion of interest, which is fragmented by collision. The fragment ions produced are reaccelerated and their mass and intensities determined by the second analyzer. For this project, a 4800 Plus MALDI TOF/TOF (AB SCIEX, Foster City, CA, USA) mass spectrometer was used.

Table 5: Comparison of performance characteristic of commonly used mass spectrometers for proteomics.

Instrument	Mass resolution	Mass accuracy (ppm)	Sensitivity	m/z range	Scan rate	Dynamic range	MS/MS capability	Ion source	Main applications
QIT	1000[a]	100–1000	Picomole	50–2000; 200–4000	Moderate	1E3	MS[nd]	ESI	Protein identification of low complex samples; PTM identification
LTQ	2000[a]	100–500	Femtomole	50–2000; 200–4000	Fast	1E4	MS[nd]	ESI	High throughput large scale protein identification from complex peptide mixtures by on-line LC–MS[n]; PTM identification
Q-q-Q	1000	100–1000	Attomole to femtomole	10–4000	Moderate	6E6	MS/MS	ESI	Quantification in selective reaction monitoring (SRM) mode; PTM detection in precursor ion and neutral loss scanning modes
Q-q-LIT	2000[a]	100–500	Femtomole	5–2800	Fast	4E6	MS[nd]	ESI	Quantification in SRM mode; PTM detection in precursor ion and neutral loss scanning modes
TOF	10 000–20 000	10–20[b]; <5[c]	Femtomole	No upper limit	Fast	1E4	n/a[e]	MALDI	Protein identification from in-gel digestion of gel separated protein band by peptide mass fingerprinting
TOF–TOF	10 000–20 000	10–20[b]; <5[c]	Femtomole	No upper limit	Fast	1E4	MS/MS	MALDI	Protein identification from in-gel digestion of gel separated protein band by peptide mass fingerprinting or sequence tagging via CID MS/MS
Q-q-TOF	10 000–20 000	10–20[b]; <5[c]	Femtomole	No upper limit	Moderate to fast	1E4	MS/MS	MALDI; ESI	Protein identification from complex peptide mixtures; intact protein analysis; PTM identification
FTICR	50 000–750 000	<2	Femtomole	50–2000; 200–4000	Slow	1E3	MS[nd]	ESI; MALDI	Top-down proteomics; high mass accuracy PTM characterization
LTQ-Orbitrap	30 000–100 000	<5	Femtomole	50–2000; 200–4000	Moderate to fast	4E3	MS[nd]	ESI; MALDI	Top-down proteomics; high mass accuracy PTM characterization; protein identification from complex peptide mixtures; quantification

[a] Mass resolution achieved at normal scan rate; higher resolution achievable at slower scan rate.
[b] With external calibration.
[c] With internal calibration.
[d] n > 2, up to 13.
[e] Fragmentation achievable by post-source-decay.

Orbitrap

In 1999 it was introduced to the market the first mass analyzer based on a new physical principle of ion separation: the separation in an electric field. This analyzer, introduced by the mathematician Alexander Makarov and called Orbitrap, (Figure 9) is constituted by an inner electrode (central) and an outer electrode, axially symmetrical, which create a combined square logarithmic electrostatic potential.

The ions rotate around a center electrode and oscillate with harmonic motion along its axis (z direction) with a frequency characteristic of their values of m/z (Figure 9).
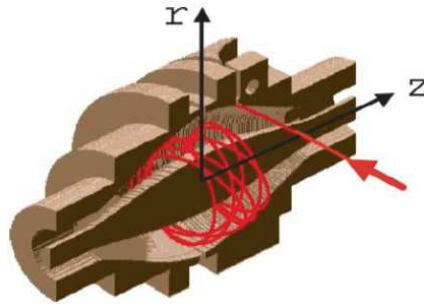

Figure 9: Ion motion within an Orbitrap analyzer.

As mentioned, within this analyzer, the axial symmetric electrodes create a square logarithmic **U** electrostatic potential, which can be calculated through the equation:

$$U(r,z) = \frac{k}{2}\left(z^2 - \frac{r^2}{2}\right) + \frac{k}{2}(R_m)^2 \ln\left[\frac{r}{R_m}\right] + C$$

where **r** and **z** are the cyclic coordinates, **C** is a constant, **k** is the field curvature and **R$_m$** is the characteristic radius. In this U field, the stable trajectories of the ions are the combination of a rotational motion around the electrode and of an oscillatory motion along the axes, which results in a complex spiral. The equations of motion for this mass analyzer are very complex. From these equations it follows that the mass and the charge are correlated with the frequency of axial oscillations, expressed in radiant/second:

$$\omega = \sqrt{(q/m)\,k}$$

ω is completely independent of the energy and position of the ions, and thus can be used for analysis of mass (in fact in the expression appears just the ratio q/m). All ions have then a harmonic oscillatory motion of the same amplitude but of different ω frequency. These frequencies are measured in a non-destructive way by a differential amplifier which acquires the signals of the current image in the time domain; through the successive application of Fourier transform the mass spectrum is obtained. For each ion is produced a wave function; therefore a mixture of ions gives rise to overlapped signals that can be converted to a mass spectrum thanks to Fourier transform.

The first Orbitrap instrument, LTQ Orbitrap Classic was released in 2005 as a tandem instrument that also included a linear ion trap for performing collision-induced dissociation (CAD or CID) fragmentation. Two years later, in 2007, a new instrument, LTQ Orbitrap XL, was introduced. In this instrument, a modified C-trap allows high energy collision dissociation (HCD) to be studied. This instrument can also be modified to perform electron transfer dissociation (ETD). Orbitrap Velos was introduced in 2009 and was the first serious upgrade of Orbitrap. A new S-lens was added that led to a 10-fold increase in ion transfer efficiency. The quadrupole linear ion trap was replaced with a dual ion trap to improve ion isolation and increase the speed of MS/MS. Furthermore, an independent HCD collision cell was added after C-trap to allow rapid ion extraction, sensitivity and trapping capabilities. In 2011 a combination of a stand-alone Orbitrap mass spectrometer ("Exactive) and a quadrupole mass analyzer was introduced as a new smaller Orbitap ("Q-exactive). In the same year, a new compact high field Orbitrap analyzer was realized as a core of Orbitrap Elite. The resolving power of this instrument together with an enhanced Fourier transform (eFT) algorithm was 4 times better compared to its predecessors. In 2013, a new instrument, Orbitrap Fusion Tribrid Mass Spectrometer, was introduced. This instrument combines the best of quadrupole, linear ion trap and Orbitrap mass analysis in a revolutionary Tribrid architecture. This instrument presents ultrahigh resolution up to 450,000 FWHM and removes interferences. Moreover, the precursor selection using a quadrupole mass filter allows the ion trap and Orbitrap mass analyzers to operate in parallel for excellent sensitivity and selectivity. Also, multiple dissociation techniques, (CID, HCD, and optional ETD) with ion trap or Orbitrap detection at any level of MSn maximize flexibility for research applications and a novel Orbitrap analyzer design and dual-pressure ion trap configuration provide fast MS/MS scan rates up to 15 Hz with unmatched spectral quality.
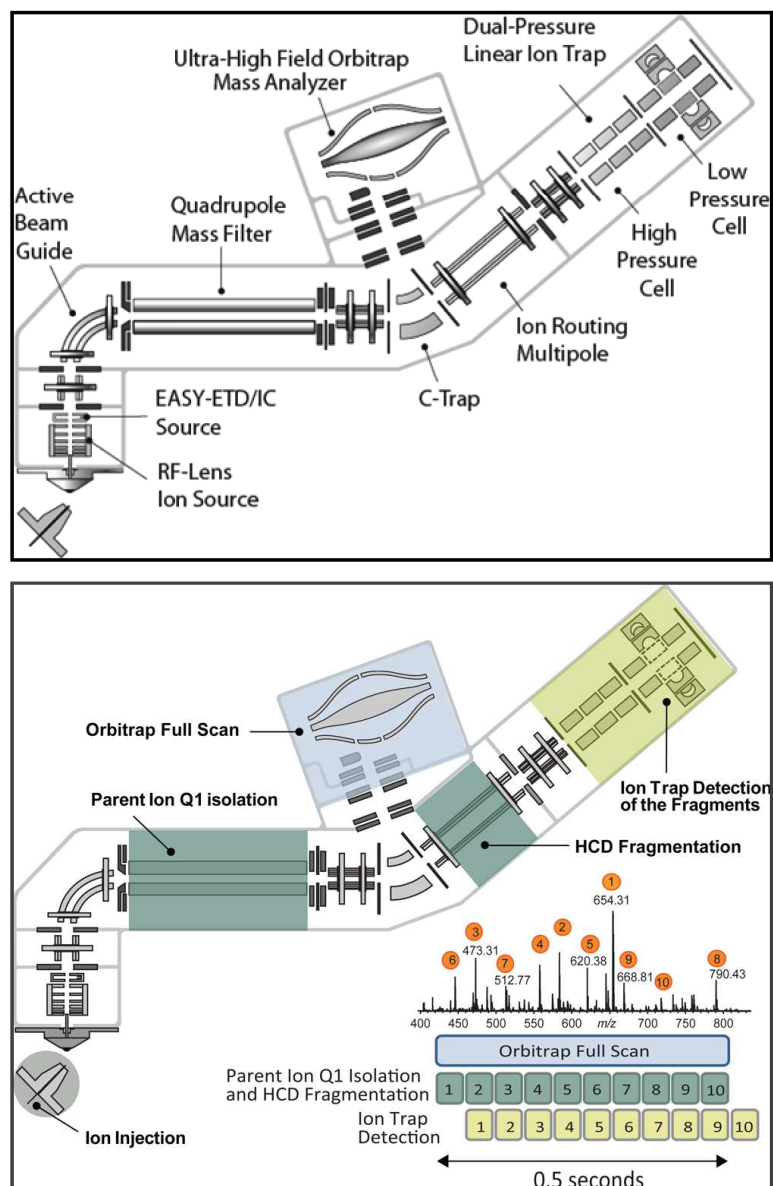
Figure 10: **Connection between the phases that will produce spectra MS and MS/MS and instrumental components (From Senko et al. 2013).**

The data analyzed in thiwork were reordered using a Q-Excative and a Obitrap Velos.

Fragmentation methods

Three principal fragmentation methods are used today in proteomic analysis:

*CID***:** Collision Induced Dissociation (CID) is the most commonly used method of fragmentation in modern proteomics, producing good fragment coverage when fragmenting peptides. [180] In CID, selected precursor ions are collided with an inert gas; upon anelastic collision. The initial kinetic energy of the accelerated ion is converted to

vibrational energy which, when dispersed across the ion, producing a series of fragment ions. For the unmodified peptides, CID fragmentation usually occurs at the peptide bond between the carboxyl group and amino group. The produced fragments are referred to as y- and b-ions (Figure11) according to the nomenclature first proposed by Roepstorff and Fohlman.[181] Glycopeptides, do not behave as non-glycosylated peptides, with regards to fragmentation. A glycopeptide is composed of a relatively constant peptide part, and a glycan moiety which may differ rather much, depending on the given microheterogeneity. When glycopeptides are fragmented using CID or HCD, peptide fragments are rarely obtained. Since the glycosidic linkages between the monosaccharide residues are easier broken than peptide bonds, partial degradation of the glycan moiety is often observed [182] while the peptide is left intact, functioning as a constant mass tag throughout the spectrum. Two main types of glycopeptide fragment ions are produced; Y-ions, carrying the reducing end of the glycan, conjugated to a peptide, and B-ions, which are free single charged glycan fragments, originating from the non-reducing end, according to the nomenclature first proposed by Domon and Costello(Figure11). [183] The Y-ions provide greater glycostructural details, but B-ions (henceforth referred to as oxonium ions) can be used to identify glycopeptides from a complex dataset, and provide an overview of the structural elements. However, the CID process generates only very limited ions diagnostic of the amino acid sequence of glycopetides, so that an alternative fragmentation approach is often needed.
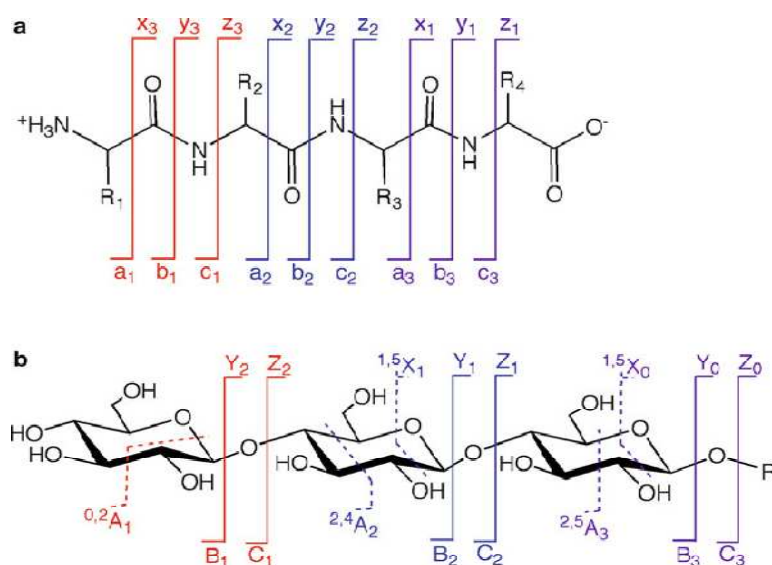

Figure 11: (a) Nomenclature for peptide fragmentation, as proposed in 1984 by Roepstorff and Fohlmann; (b) Nomenclature for glycopeptide fragmenation, as proposed in 1988 by Domon and Costello.

HCD**:** High-energy Collison Dissociation (HCD) is a fragmentation method which produces the same y-/b- ions and Y-/B-ions as CID. It can be performed in instruments equipped with a HCD fragmentation cell and utilizes higher energy than CID. In contrast to ion trap CID, "one-third" m/z cut-off restriction is absent in HCD, thus the smaller fragments corresponding to monosaccharide or disaccaride fragment can be detected. Thus, the oxonium ions m/z 163.06 $[Hex+H]^+$, m/z 204.09 $[HexNAc+H]^+$, m/z 325.11 $[2Hex+H]^+$, m/z 366.14 $[HexNAc+Hex+H]^+$, m/z 292.10 $[NeuAc+ H]^+$, m/z 274.10 $[NeuAc–H_2O+H]^+$, m/z 657.23 $[HexNAc + Hex + Sia + H]^+$, m/z 407.17 $[2HexNAc + H]^+$, 308.10 $[NeuGc + H]^+$ and 290.10 $[NeuGc–H_2O+H]^+$ can proven to be highly sensitive markers for the presence of glycopeptides

In addition, a series of Y-type ions resulting from the sequential neutral losses of monosaccharides from the glycan moieties of glycopeptides are also displayed in the HCD MS/MS spectra.

ETD: Electron-transfer dissociation (ETD) is based on the same principle as electron capture dissociation, ECD, introduced in 1998 by Zubarev. An electron is transferred to the ionized analyte and this causes the fragmention of the molecole in a non-ergodic way. The exact mechanism of the electron transfer is still debated, but it is known that ECD produces mostly backbone fragmentation of the c/z-type, and is a good compliment to gain more information about a peptide sequence; furthermore, labile modifications such as glycosylation and phosphorylation are maintained on the peptide. ECD can be performed in a Fourier transform ion cyclotron resonance (FTICR) mass spectrometers, but FTICR-instruments are very expensive to run and maintain. Ion trap instruments are therefore much more common, and ETD has been developed as an alternative to ECD for these instruments, because the strong electric field used to manipulate the ions in a trap affects the electrons and they are expelled from the trap. In ETD, ions in the gas-phase react with radical anion fluoranthene instead of thermal electrons and the resulting fragmentation patterns are analogous to those produced by ECD.

For this project, were used: Collision Induced Dissociation (CID) and Higher Energy Collision Dissociation (HCD).

BIOINFORMATIC ANALYSIS OF DATA

The fundamental problem that biological research is facing today is to have a considerable and growing amount of data on gene and protein sequences accumulated from studies of genomics and proteomics. It is therefore necessary to have adequate support of programs that can not only store data but also to analyze them in a reasonably short time, to identify proteins, to compare sequences, to predict secondary and tertiary structures etc., and construct molecular models.

In this thesis, LC–MS/MS data were processed using Proteome Discoverer, GPMAW and MassAI softwares.

Proteome Discoverer v1.4.1.14 (Thermo Scientific).

In order to identify proteins, LC –MS/MS data were processed using **Proteome Discoverer v1.4.1.14** (Thermo Scientific). Data were searched in **SwissProt** database using MASCOT algorithm (Matrix Science, London, UK, version 2.5.1).

The software Proteome Discoverer evaluates possible homologies in the databases based on the information of the mass spectrum generated from peptide maps eliminating unthinkable manual calculations needed to match precisely the masses of the fragments of the peptide digestions to those predicted by the digestion conditions. In addition to the choice of the database there are other parameters that must be set before launching the search. These parameters make the search more or less restrictive, affecting the evaluation of the results:

- ✓ **taxonomy**: if known, this parameter will force the program to compare only to proteins belonging to indicated taxonomy;
- ✓ **enzyme digestion**: the program will compare theoretical peptides born from cutting mode selected enzyme;
- ✓ **missed cleavage**: it can be expected that the enzyme goes to encounter errors in the recognition of cleavage sites; in the program you may indicate the maximum number of missed cleavage sites;
- ✓ **mass of protein**: if you can predict what will be the mass of the protein (for example based on the area of origin of the spot in a two-dimensional gel) can limit the search to the value of the mass of the intact protein. This will eliminate any false positives at the risk of losing information that might be true (protein fragment);

- ✓ **charging the peptide**: if you are introducing a peak list from a spectrum, must indicate if it is a positive or negative mass spectrum;
- ✓ **monoisotopic or average mass**: it can be given the mass of a peptide, or the average of all peaks belonging to the isotopic peptide or the first peak (mass monoisotopic); this depends on the set during the peak-list allocation;
- ✓ **fixed modifications:** they are the modifications to which the peptide has been subjected during the treatment, such as the reduction of disulfide bonds and subsequent alkylation. The program then will compare the theoretical peptide to whose mass this modification is applied;
- ✓ **variables modifications**: you can indicate the changes that you believe the protein may be subjected. In this category are the post-translational modifications, such as phosphorylations, oxidation of cysteine and methionine and the formation of pyroglutamic acid after the cyclization of glutamine or glutamic acid, in the N-terminal. To indicate any variables changes can lead to the identification of peptides that otherwise might not be recognized;
- ✓ **tolerance of mass**: if the spectrum is well calibrated and is high esolution a value of 10 ppm should lead to a good identification. Raise tolerance means increasing the risk of false matches. The results are processed by the algorithm Mowse71. This algorithm is not limited only to the counting of peptides resulting from the match but uses empirically derived factors to assign a statistical weight to each pairing. From this it is obtained the total score.

The score is the absolute probability that the observed event is a random event. The result is not however reported directly, both because it includes a very wide range of sizes, and a "high" rating corresponds to a "low" probability, and this can create problems of ambiguity. The rating is then reported as $-10 * LOG_{10} (P)$, where P is the absolute probability. A minimum threshold for defining a significant result is that the event is significant if it is expected to occur randomly at a frequency less than the 5% (p <0.05).

Factors to consider when evaluating a result are:
- ✓ **the score**: must be greater than the significance threshold. At least 2 or 3 peptides must be paired;
- ✓ **the coverage sequence**: the program allows you to visualize and quantify the percentage coverage of the protein. Good coverage exceeds 30%;

✓ **the molecular weight**: if the mass of the result does not match the estimated (by position spot for example) this may be due to a false positive or you can make several considerations. If the resulting mass is higher than the estimated one, the protein may have been subjected to post-translational modifications; in that case you have to evaluate whether the increase can be attributed to value of some modifications, and restart the search to explore possible modification (but not always detectable in MS).

GPMAW

GPMAW, developed by Lighthouse data, is a bioinformatical tool, used for interpretation of mass spectrometric data. In this project, the program was used to create *in-silico* digests of the target proteins, and thereby determining the properties of the expected peptides and glycopeptides, both before and after MS analysis.

MassAI

MassAI is a general search engine, developed in-house, by MassAI Bioinformatics (downloadable from www.massai.dk), for analysing LC-MS/MS dataset, which identifies glycosilation on intact glycopeptides in several steps. By providing the program with a database of proteins along with common search parameters, such as proteases and modifications, the program is able to identify proteins and peptides from a mascot generic file (mgf). The identification of glycopeptides is based on a flexible glycan database, and losses of common glycan components on the MS/MS level.

First, MS/MS scans containing peaks from glycan oxonium ions are singled out and forwarded to next step. Second, selected scans are searched for "anchor" peaks (i.e., the peptide peaks where the glycan moities are attached), which can determine the peptide mass. Generally, the glycan fragments dominate the CID and HCD scans. Third, the peptide mass is calculated from the above anchor peaks, and the molecular weight of glycan is searched against a database of known glycan mass. Once a pair has been made, it is fragmented in silico to generate B/Y –type ions, which are then matched against the observed peaks. In the present work all identified results were manually validated.

AIM OF THE WORK

Proteins represent one of the most important milk components from nutritional and physiological viewpoints. On the other hand, the wide use of cow's milk in human diet has shown that a considerable percentage of subjects are allergic to its protein components. This disorder is commonly known as cow milk protein allergy (CMPA). Particularly, cow's milk, used as substitute of breast feeding when mother's milk is not available or advisable, represents the main source of allergens in infants. Approximately 2-3% of infants younger than one years of age are allergic to cow's milk proteins. This allergy is normally outgrown in the first year of life but 15% of allergic children remain allergic. The proteins most frequently and most intensively recognized by IgE are the caseins and β-LG, even if lower abundant (i.e. lactoferrin, IgG and bovine serum albumin) and trace components appear to be potential allergens.

Taking into account the high incidence of CMPA in infants and considering that breast feeding is not always possible, indicated or sufficient, alternative supply becomes indispensable. Therefore, one of the major objectives of the pharmaceutical industries is the production of milk and milk-based foods (i.e. infant formulae) close to breast milk. For infants affected of CMPA infant formulas, which are products based on bovine milk modified by enzymatic and/or thermal treatments, represent the preferred choices in the treatment. However, it should be considered that allergy in these products is reduced, but never completely suppressed and adverse reactions have been experienced also with these preparations.

Many investigations have reported real benefits from using non-bovine milks as alternative in cases of CMPA. However, allergies to non-bovine milk proteins have also been documented. On this respect, many clinical trials have shown that equine milk, and in particular, donkey's milk (DM) has special nutritional and therapeutic properties, and may represent a safe and alternative food in CMPA, providing dietary adequacy and good palatability.

Notwithstanding the increasing knowledge of DM protein fraction, up to the present, no informations are available for the primary structure and glycan composition of donkey lactoferrin, one of the most important glycoproteins, that confers high hygienic qualities

to DM and presents an array of biochemical properties, including immune-modulation, iron-binding, antioxidant, antibacterial and antiviral activities.

In the frame of our research line oriented to the characterization of the donkey milk proteins and its genetic polymorphism, in this project it is reported for the first time the direct characterization of the primary structure and a comprehensive site specific glycan profile of donkey lactoferrin by coupling RP-HPLC, ion exchange chromatography (IEC), enzymatic digestions and mass spectrometric analysis.

MATERIALS AND METHODS

Materials

Acetic acid (AA), formic acid (FA), trifluoracetic acid (TFA), Glycolic acid (GA), α-cyano-4-hydroxycinnamic acid (CHCA), endoproteinase GluC (*Staphylococcus aureus strain V8*), endoproteinase AspN (*Pseudomonas fragi*), α-chymotrypsin from bovine pancreas, albumin from chicken egg white, fetuin from fetal bovine serum were, alpha 1 acid glycoprotein were obtained from Sigma-Aldrich (Milan, Italy). Modified porcine trypsin was purchased from Promega (Madison, WI, USA). Ammonium bicarbonate, sodium acetate, HPLC grade H_2O and acetonitrile (ACN) were provided by Carlo Erba (Milan, Italy). Dithiothreitol (DTT), iodoacetamide (IA), ammonium acetate, sodium acetate and sodium chloride were purchased from Fischer-Chemical. N-glicosidase F (PNGaseF) was purchased from Roche. Micro-Spin regenerated cellulose filters (0.45 μm pore size) and nylon membranes (0.22 μm pore size) used for filtration of samples and IEC-HPLC solvents, respectively were from Alltech (Milan, Italy). IEC-HPLC solvents were filtrated using Nylon filters (0.22 μm pore size, Alltech, Milan, Italy). Spectra-Por Float-A-Lyzer dialysis tubes (cut off 3.5-5 kDa) were obtained from Sigma-Aldrich (Milan, Italy). TiO2 beads were obtained from GL Sciences (Japan). PolyHYDROXYETHYL A^{TM} was provided by PolyLCinc (Columbia, MD, USA). POROS Oligo R3 was purchased from Applied Biosystems (Framingham, MA, USA). GELoader tip were obtained from Eppendorf (Hamburg, Germany). All the materials used had a high degree of purity and were used without further purification.

Milk sample preparation and lactoferrin purification

The individual milk sample was collected in Eastern Sicily (ASILAT srl farm, Milo, Catania) from a Ragusano breed donkey. After milking, the sample was immediately frozen and stored at –20 °C until used. The milk was defatted by centrifugation at 3500 rpm, 4°C for 20 min and the casein fraction precipitated from the skimmed milk at pH 4.3 with sodium acetate/acid acetic buffer and separated from the supernatant (i.e. whey protein fraction) by centrifugation at 8000 rpm, 4°C for 20 min. The whey protein fraction was dialyzed against water, freeze-dried and resuspended in CH_3COONH_4 50mM, pH 5.5 at the concentration of about 10 mg/mL.

Isolation of lactoferrin from the whey fraction was performed by an IEC-HPLC analysis carried out on a Jasco Pu-980 HPLC system equipped with a detector Uv – Jasco Uvidec-1000-III. Data were acquired with a PC using Borwin Chromatography software.

In detail, the whey fraction solution (in CH_3COONH_4 50mM, pH 5.5) was filtered on Micro-spin filters (Alltech, Milan, Italy) and 500 µL loaded onto a sulfonic column (GE Healthcare Mini S 4.6/50 PE; 4.6 x 50 mm). Whey proteins were separated at a flow rate of 0.83 mL/min and room temperature by holding 100% of solvent A (CH_3COONH_4 50mM, pH 5.5) for 3 minutes and then with a linear gradient of solvent B (CH_3COONH_4 50mM + NaCl 1M, pH 5.5) in A from 0% to 100% in 30 min. Peaks were detected by their absorption at 280 nm, collected manually, dialyzed against water and freeze-dried.

Reduction, alkylation, enzymatic digestions and deglycosylation reactions to study the primary structure

In order to characterized the IEC fraction containing donkey lactoferrin, the corresponding chromatographic peak eluted at 14 min was dialyzed and then subjected to reduction, alkylation and enzymatic digestion using different enzymes. In detail, an aliquot corresponding to 5 µg was dissolved in 100 mM NH_4HCO_3, pH 8.3, at a concentration approximately of 1 µg/µL. DTT (50-fold molar excess over the disulfide bonds) was added, and the reduction was carried out at 57°C for 15 min in nitrogen atmosphere. At the end of the reduction reaction, IA was added to the mixture at a molar IA/DTT ratio of 2:1, and the alkylation was performed at room temperature for 15 min in the dark. The reaction was stopped by cooling in liquid nitrogen. Then, with the aim to maximize the sequence coverage, reduced and alkylated chromatographic fraction was separately subjected to enzymatic digestions using modified porcine trypsin, α-chymotrypsin, endoproteinase GluC and AspN. Each 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 reactions were stopped by cooling in liquid nitrogen.

Digested mixtures were deglycosylated using PNGase F. Briefly, an aliquot corresponding to1 µg of reduced and alkylated peptides in 50 mM NH_4HCO_3 (pH 8.3) were added to 1 µL of PNGase F solution and incubated at 37°C overnight. Finally,

deglycosylated peptide mixtures were dried and resuspended in 5 µL of 0.1% FA prior to LC/MS analyses.

Enzymatic digestions to study N-glycan composition

With the aim to obtain glycopeptides with a low molecular mass, chromatographic fraction was reduced, alkylated and enzymatic 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.

4-sulfophenyl isothiocyanate (SPITC) derivatization

The reagent solution was prepared by dissolving SPITC in 50 mM $NaHCO_3$ (pH 8.3) to a concentration of 10 µg/µL. The reaction was carried out in a 1.5 mL microcentrifuge tube by mixing 8 µL of reagent solution with about 40 pmol of tryptic digest. After incubation at 55 °C for 1h, the reaction was stopped by adding 1 µL of 5% TFA. [184] The mixtures of derivatized tryptic peptides were subjected to a desalting/concentration step prior to analysis by MALDI MS. In detail, an aliquot of about 15 pmol of derivatized tryptic mixture was loaded onto a homemade 5 mm nanocolumn packed with POROS R2 in a constricted GELoader tip. [185] A syringe was used to force liquid through the column by applying gentle air pressure. The column was washed with 10 µL of 0.1 % TFA and the bound peptides subsequently eluted directly onto the MALDI target with 0.6 µL of matrix solution (5 µg/µL CHCA in 70% [v/v] ACN and 0.1% [v/v] TFA).

Glycopeptides enrichments

TiO$_2$ enrichment[166]

The digested mixture containing non-glycosylated peptides and glycopeptides was dried down and diluited in 100 µL of loading buffer (80% Acetonitrile, 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, were precipitated by centrifigutation, helping the selective precipitation of sialylated glycopeptides. The supernatant, containing only neutral glycopeptides and non-glycosylated peptides 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% Acetonitrile, 1% TFA), third with 50 µL of washing buffer 2 (10% Acetonitrile, 0.5% TFA). In each step the supernatant was stored. Then the beads with sialylated glycopeptides were dried and the sialylated glycopeptides eluted from beads with 50 µL of Elution buffer (60 µL Ammonia solution (28%) in 940 µL H2O, pH 11,3.

HILIC enrichment[186]

A home-made column was done with PolyHydroxyEthyl (PolyLC) 3µm as stationary phase. 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. In detail, the previous supernatant stored from TiO_2 enrichment, containing only neutral glycopeptides and non-glycosylated peptides, was dried down, resuspended in 20 µl of 80% ACN and 2% formic acid FA, loaded in the home-made column, and eluted with 20 µl 2% FA. The flow-through contained only non glycosylated peptides whereas the neutral glycopeptides were eluted from the column with 2% FA.

Mass spectrometry analysis to study the primary structure

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). 5 µL of the reconstituted samples (in 0.1% FA) were loaded onto a home-made µ-pre-column (100 µm x 2 cm, 5 µm ReproSil Pur C18). After washing the trapping column with solvent A (0.1% FA) at a flow rate of 5 µL/min for 5 min, the solution was switched from the trapping column onto a home-made reversed phase C18 column (75µm x 17 cm, 3µm ReproSil Pur C18). Peptides were separated by elution at a flow rate of 250 nL/min and 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). Mass spectrometer operated in data dependent mode acquisition as follows: i) survey scans of peptide precursors from 400 to 1400 *m/z*, performed at 70K resolution (@ 200 *m/z*); ii) MS/MS analysis, in the *m/z* range of 200-2000, of the twelve 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 msec. Dynamic exclusion was set to 15 s. Doubly-, triply- and quadruply- charged ions were fragmented into the HCD collision cell using a normalized collision energy (NCE) of 30. Subsequent MS/MS were acquired using an AGC target value of 2×10^4, a maximum injection time of 200 ms and a resolution of 17,5 K.

Mass spectrometry data were also acquired on a Thermo Scientific™ LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) combined with a proxeon Easy-nLC system (Thermo Fisher Scientific, Odense, Denmark). 5 µL of the reconstituted samples in 0.1% FA were loaded and separated using the same µ-pre-column, reversed phase C18 column, flow rate and gradient described above.

Eluting peptide cations were converted to gas-phase ions by the Thermo Scientific Nanospray Ion Source, using a source voltage of 2.3 kV and introduced into the mass spectrometer through a heated ion transfer tube (270 °C). Mass spectrometer operated in data dependent mode acquisition as follows: i) survey scans of peptide precursors from 350 to 1800 *m/z*, performed at 60K resolution (@ 400 *m/z*); ii) MS/MS analysis, in the *m/z* range of 100-2000, 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 msec. Dynamic exclusion was enabled with repeat count 1 and set to 20s. Only doubly- triply- and quadruply charged peptide ions were fragmented into the CID collision cell using normalized collision energy (NCE) of 35. Subsequent MS/MS were acquired using an AGC target value of 5×10^4, a maximum injection time of 200 ms and a resolution of 15 K. Mass spectrometers calibration was performed using the Pierce® LTQ Velos ESI Positive Ion Calibration Solution (Thermo Fisher Scientific). MS data acquisition was performed using the *Xcalibur* v. 3.0 software (Thermo Fisher Scientific).

MALDI MS analyses were acquired using a 4800 Plus MALDI TOF/TOF (AB SCIEX, Foster City, CA, USA) mass spectrometer operating in reflecton, positive ion mode, and using an acceleration voltage of 20 kV. Laser intensity and number of laser exposures were varied to optimize the spectral quality. The mass range was set to 700-5000 Da. For all MALDI TOF/TOF MS/MS fragmentations, air was used as collision gas. β-lactoglobulin tryptic peptides (*m/z* 837.48 and 2313.26) were used for external mass calibration. MS data acquisition and data handling were performed using Data Explorer (AB SCIEX version 4.6). MALDI MS/MS spectra of derivatized tryptic peptides were manually interpreted.

Mass spectrometry analysis to study the N-glycan composition

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). 5 µL of the reconstituted samples (in 0.1% FA) were loaded onto a home-made µ-pre-column (100 µm x 2 cm, 5 µm ReproSil Pur C18). After washing the trapping column with solvent A (0.1% FA) at a flow rate of 5 µL/min for 5 min, the solution was switched from the trapping column onto a home-made reversed phase C18 column (75µm x 17 cm, 3µm ReproSil Pur C18). Peptides were separated by elution at a flow rate of 250 nL/min and 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). Mass spectrometer operated in data dependent mode acquisition as follows: i) survey scans of peptide precursors from 700 to 2000 m/z, performed at 70K resolution (@ 200 m/z); ii) 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 msec. Micro scans 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 into 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).

Primary structure characterization by database searching

LC/MS/MS data were processed by Proteome Discoverer v. 1.4.1.14 (Thermo Scientific). Data were searched against the "Mammalia" UniProt database (SwissProt release July 2015, containing 66979 entries) using the Mascot algorithm (Matrix Science, London, UK, version 2.5.1) and SEQUEST algorithm. Full enzymatic peptides with a maximum of 3 missed cleavage sites were subjected to bioinformatic search. Cysteine carbamidomethylation was set as fixed modification, whereas oxidation of methionine, transformation of N-terminal glutamine and N-terminal glutamic acid

residue in the pyroglutamic acid form, and deamidation of glutamine and asparagine residues were included as variable modifications. The precursor mass tolerance threshold was 10 ppm and the max fragment mass error was set to 0.1 Da (for data acquired by Q Exactive) and 0.6 Da (for data acquired by Velos Pro). Peptide spectral matches (PSM) were validated using Target Decoy PSM Validator node based on q-values at a 1% FDR.

In addition, the General Protein/Mass Analysis for Windows (GPMAW) software (http://welcome.to/gpmaw) was used for the sequence handling and storage.

N-glycan composotion characterization by database searching

Glycopeptides were identified with a software tool, developed in-house, by MassAI Bioinformayics,[187] The identification was performed by a Mascot Generic File (mgf), providing the program with a database of protein and using a flexible glycan database. The N-glycans from the library were composed of hexose (Hex), HexNAc, fucose, N-acetylneuraminic acid (NeuAc) and N-glycolylneuraminic acid (NeuGc). Full chymotryptic peptides with a maximum of 3 missed cleavage sites were subjected to bioinformatic search. Cysteine carbamidomethylation was set as fixed modification, whereas oxidation of methionine and deamidation of glutamine residues were included as variable modifications. The precursor mass tolerance threshold was 10 ppm and the max fragment mass error was set to 0.1. All identified results were manually validated.

RESULTS

PRIMARY STRUCTURE

An aliquot of 500 µL of the dialyzed whey fraction was loaded on an analytical sulfonic column, in order to isolate donkey lactoferrin. The resulting chromatogram (Figure 12) shows the presence of four peaks, named F1, F2, F3 and F4, which were manually collected. The collected fractions were dialyzed, reduced, alkylated, digested with modified porcine trypsin, and the resulting peptide mixture analyzed by MALDI MS. MS data revealed that in the chromatographic fraction F1, F2 and F3 eluted β-lactoglobulin, α-lactalbumin and lysozyme, respectively. Instead, most of the signals present in the MALDI-TOF mass spectrum of the tryptic digest of peak F4 were identified as arising from the theoretical cleavages of the mare lactoferrin genomic deduced sequence (UniProtKB Acc. Nr. O77811), indicating that in this fraction is eluted the homologous donkey lactoferrin, whose sequence is not present in the database.
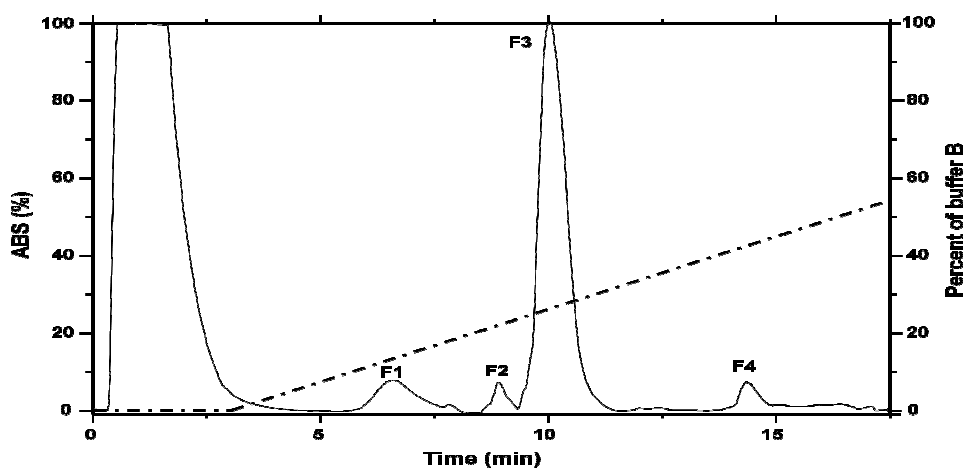
Figure 12: UV profile (λ=280 nm) of the IEC run of the whey protein fraction. The lactoferrin was identified in the chromatographic fraction labelled as F4. Other proteins identified were: β-lactoglobulin (F1); α-lactalbumin (F2), lysozyme (F3).

In order to characterize the primary structure of donkey lactoferrin, the tryptic peptides of fraction 4 were deglycosylated using PNGase F to remove N-glycans from glycosylated asparagines and thus facilitate MS data interpretation. The deglycosylated

tryptic mixture was investigated by RP-HPLC/nESI-MS/MS on a Q Exactive mass spectrometer using HCD for peptide molecular ion fragmentation and MS/MS data were used for database searching. Determination of donkey lactoferrin sequence was achieved using as reference the sequence of the homologous mare lactoferrin.
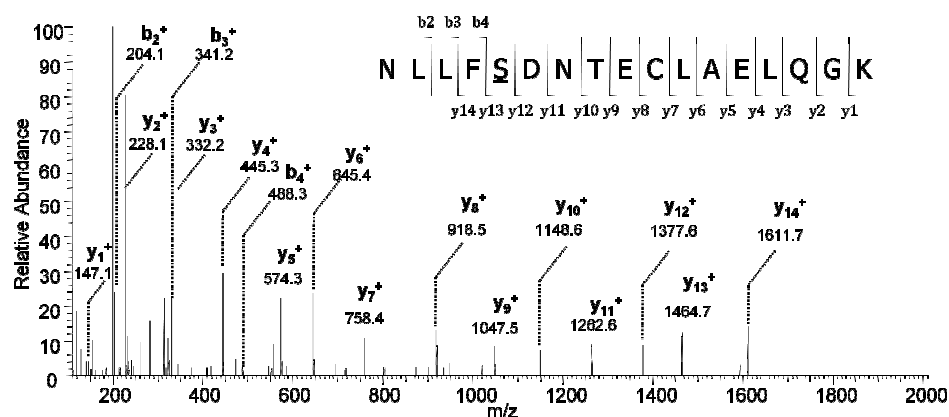
The MS/MS spectra of the doubly- or triply-charged molecular ions of the peptides showed that the sequences of lactoferrin in these two close phylogenetic related species present a high level of similarity. In fact, most of the tryptic peptides can be interpreted as arising from the theoretical cleavages of mare lactoferrin sequence, as reported in Table 6.

In addition to the above described peptides, LC/MS/MS analysis of the lactoferrin tryptic digest revealed the presence of several signals, which remained unassigned. In order to characterize the amino acid sequence of the unassigned signals, the corresponding MS/MS of their doubly- or triply- charged ions were manually interpreted. This analysis showed that these signals are due to specific donkey tryptic peptides presenting amino acid point substitutions with respect to the sequence of the reference mare lactoferrin. In detail, interpretation of the MS/MS spectrum of the doubly- charged molecular ion at m/z 780.93 (MH$^+$ 1560.85, Figure. 13a) allowed to deduce the sequence GKPQTHYYAVAVVK, which corresponds to the tryptic fragment $T_{(15+16)}$ of mare's lactoferrin, carrying the amino acid substitution Arg→His at position 91 (Table 6). The MS/MS spectrum of doubly-charged ion at m/z 976.47 (MH$^+$ 1951.93, Figure 13b) is due to a peptide having the sequence NLLFSDNTECLAELQGK, which corresponds to the tryptic fragment T_{76} of mare lactoferrin with the amino acid substitution Asn→Ser at position 642 (Table 6). Analogously, de novo interpretation of the MS/MS spectra of the doubly charged ion at m/z 1111.55 (MH$^+$ 2222.09) and of the triply charged ion at m/z 840.79 (MH$^+$ 2520.35) shown in Figure 13c and 13d, respectively, permitted to deduce the sequences TTYEQYLGSEYVTAITNLR, and IPSQIDSGLYLGANYLTAL/IQNRL. The two sequences match with the mare lactoferrin tryptic fragments T_{77} and T_{41} carrying the amino acid substitutions Ser→Ala at position 668, and Ile/Leu at position 328, respectively (Table 6). Finally, Figure 13e shows the MS/MS of peptide TAGWNIPMGLLFNQTGSCK of the doubly charged ion at m/z 1056.50 (MH$^+$ 2111.99) matching with the deamidated form of T_{58} of mare lactoferrin and carrying the amino acid Ala→Gly at position 466.
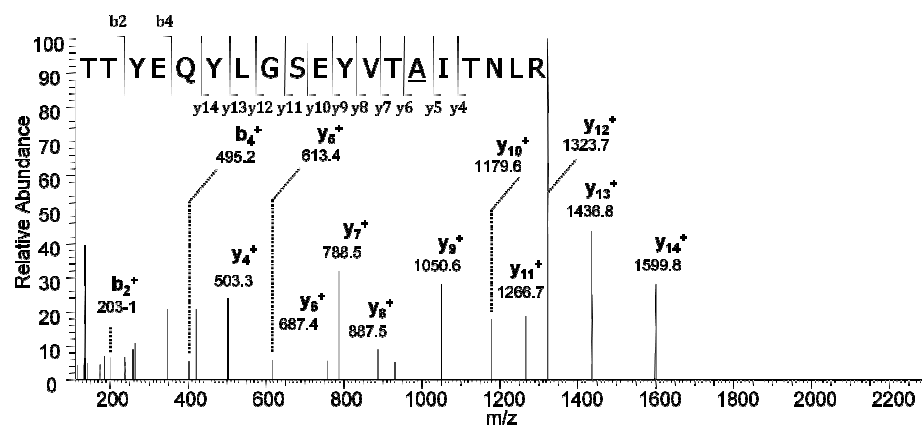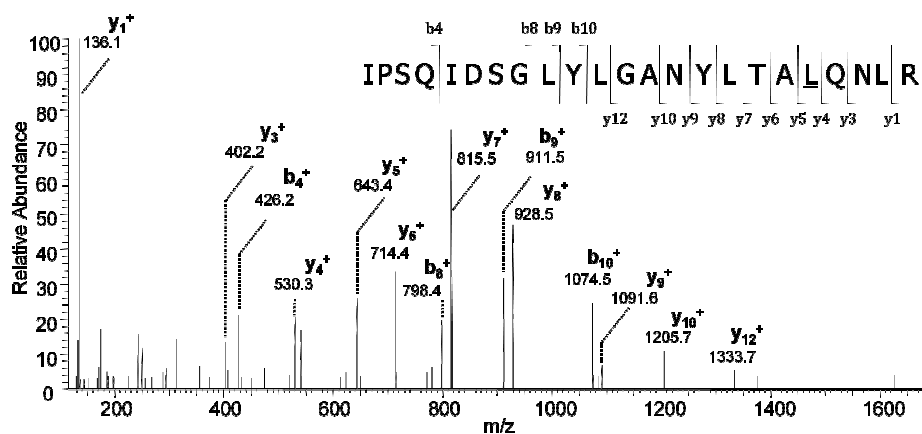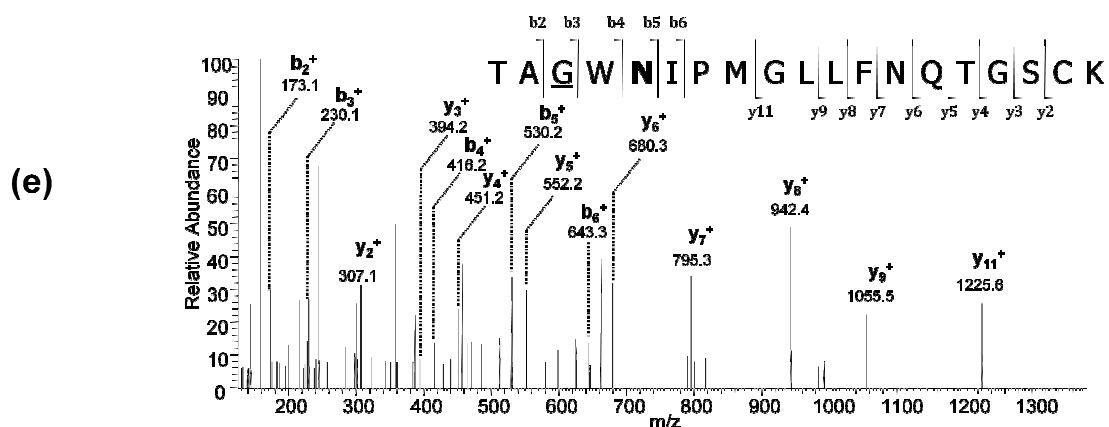
(a)

Relative Abundance vs m/z

Sequence: G K P Q T H(underlined) Y Y A V A V V K

b2 b6 b7 b8 b9 b10 b11 b12 b13
y12 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1

y_1^+ 147.1
y_2^+ 248.2
y_5^+ 515.4
y_6^+ 586.4
b_2^+ 186.1
y_3^+ 345.2
y_4^+ 416.3
b_9^{2+} 523.7
b_{10}^{2+} 573.3
b_{11}^{2+} 608.8
b_6^+ 649.3
y_{12}^{2+} 688.4
y_7^+ 749.5
b_7^+ 812.4
y_8^+ 912.5
b_9^+ 975.5
y_9^+ 1049.6
b_{10}^+ 1145.6
y_{10}^+ 1150.6
b_{11}^+ 1216.6
b_{12}^+ 1315.7
y_{12}^+ 1375.7

(b)

Relative Abundance vs m/z

Sequence: N L L F S(underlined) D N T E C L A E L Q G K

b2 b3 b4
y14 y13 y12 y11 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1

b_2^+ 204.1
b_3^+ 341.2
y_2^+ 228.1
y_3^+ 332.2
y_4^+ 445.3
b_4^+ 488.3
y_6^+ 645.4
y_1^+ 147.1
y_5^+ 574.3
y_7^+ 758.4
y_8^+ 918.5
y_{10}^+ 1148.6
y_{12}^+ 1377.6
y_{14}^+ 1611.7
y_9^+ 1047.5
y_{11}^+ 1262.6
y_{13}^+ 1464.7

(c)

Relative Abundance vs m/z

Sequence: T T Y E Q Y L G S E Y V T A(underlined) I T N L R

b2 b4
y14 y13 y12 y11 y10 y9 y8 y7 y6 y5 y4

b_4^+ 495.2
y_6^+ 613.4
y_{10}^+ 1179.6
y_{12}^+ 1323.7
y_{13}^+ 1436.8
b_2^+ 203.1
y_4^+ 503.3
y_7^+ 788.5
y_9^+ 1050.6
y_{11}^+ 1266.7
y_{14}^+ 1599.8
y_6^+ 687.4
y_8^+ 887.5

(d)

Relative Abundance vs m/z

Sequence: I P S Q I D S G L Y L G A N Y L T A L(underlined) Q N L R

b4 b8 b9 b10
y12 y10 y9 y8 y7 y6 y5 y4 y3 y1

y_1^+ 136.1
y_3^+ 402.2
b_4^+ 426.2
y_5^+ 643.4
y_7^+ 815.5
b_9^+ 911.5
y_4^+ 530.3
y_6^+ 714.4
b_8^+ 798.4
y_8^+ 928.5
b_{10}^+ 1074.5
y_9^+ 1091.6
y_{10}^+ 1205.7
y_{12}^+ 1333.7

(e)

Figure 13: MS/MS spectra of selected molecular ions in the RP-HPLC/ESI-MS analysis of the tryptic digest of donkey lattoferrin, demonstrating point substitution with respect to the mare lactoferrin sequence: (a) doubly charged molecular ion at m/z 780.93 (MH+ 1560.85, Arg→His); (b) doubly charged molecular ion at m/z 976.47 (MH+ 1951.93, Asn→Ser); (c) doubly charged ion at m/z 1111,55 (MH+ 2222.09, Ser → Ala); (d) triply charged ion at m/z 840,79 (MH+ 2520.35, Thr → Ile/Leu); (e) doubly charged ion at m/z 1056.50 (MH+ 2111.99, Ala→Gly). The sequences of the peptides are displayed with the fragment ions observed in the spectrum

These results are summarized in Table 6, where the tryptic fragments carrying the substitutions are marked with an asterisk, whereas the amino acid substitutions with respect to the mare counterpart are underlined, and in Figure 15 where identified peptides are underlined by a solid line. In summary, MS/MS analysis of the tryptic digest allowed the characterization of about 91.1% of the sequence of donkey lactoferrin, giving also clear evidence of five amino acid substitutions with respect to its mare counterpart. Inspection of Table 6 also evidences that the uncharacterized peptides were mostly low molecular weight or very hydrophilic peptides, which very likely were not detected because their molecular masses were below the scanned mass range or were not recovered because eluting in the void volume of the HLPC column, or both.

Table 6: Fragment nomenclature, position, calculated monoisotopic and experimentally measured MH⁺ of tryptic fragments of donkey lactoferrin eluting in the IEC chromatographic fraction F4 (Figure 12) characterized by HPLC/ESI-MS/MS. Amino acid sequence of mare lactoferrin (UniProtKB Acc. Nr. O77811) was used as reference. The modified tryptic fragments of donkey lactoferrin are marked with an asterisk, whereas the amino acid substitutions with respect to the mare's counterpart are underlined. The hypothetical glycosylation sites are reported in bold.

Fragment	Position	Calculated MH⁺	Measured MH⁺ ESI-MS/MS (z)	Reference Sequence (mare's lactoferrin Acc. Nr. O77811)
T_1	1-3	343.21	-	APR
T_2	4-4	147.11	-	K
T_3	5-7	361.22	-	SVR
T_4	8-18	1233.59	1233.59 (2)	WCTISPAEAAK
T_5	19-21	378.18	-	CAK
T_6	22-24	450.25	-	FQR
T_7	25-27	392.20	-	NMK
T_8	28-28	147.11	-	K
T_9	29-30	274.19	-	VR
T_{10}	31-38	875.44	875.44 (2)	GPSVSCIR
$T_{(9+10)}$	29-38	1130.61	1130.61 (2)	VRGPSVSCIR
T_{11}	39-39	147.11	-	K
T_{12}	40-53	1539.75	1539.75 (2)	TSSFECIQAIAANK
$T_{(11+12)}$	39-53	1667.84	1667.84 (2)	KTSSFECIQAIAANK
T_{13}	54-73	2089.06	2089.06 (2)	ADAVTLDGGLVYEAGLHPYK
T_{14}	74-85	1402.78	1402.78 (2)	LRPVAAEVYQTR
T_{15}	86-91	686.39	-	GKPQTR
T_{16}	92-99	912.52	-	YYAVAVVK
$T_{(15+16)}$ *	86-99	1560.85	1560.85 (2)	GKPQT<u>H</u>YYAVAVVK [b]
T_{17}	100-100	147.11	-	K
T_{18}	101-113	1375.73	1375.73 (2)	GSGFQLNQLQGVK [a]
$T_{(17+18)}$	100-113	1503.83	1503.83 (2)	KGSGFQLNQLQGVK [a]
T_{19}	114-121	887.42	887.41 (2)	SCHTGLGR
T_{20}	122-147	2906.52	2906.52 (3)	SAGWNIPGTLRPYL**N**WTGPPEPLQK [g]
T_{21}	148-164	1800.80	1800.80 (2)	AVANFFSASCVPCADGK [a]
T_{22}	165-171	950.45	950.45 (2)	QYPNLCR [a]
T_{23}	172-180	964.44	964.44 (2)	LCAGTEADK
T_{24}	181-197	1958.80	1958.80 (2)	CACSSQEPYFGYSGAFK
T_{25}	198-210	1379.66	1379.67 (2)	CLENGAGDVAFVK [a]
T_{26}	211-224	1607.72	1607.72 (2)	DSTVFENLPDEADR
T_{27}	225-226	262.14	-	DK
$T_{(26+27)}$	211-226	1850.84	1850.84 (2)	DSTVFENLPDEADRDK
T_{28}	227-236	1280.59	1280.59 (2)	YELLCPDNTR
T_{29}	237-243	804.46	804.46 (2)	KPVDAFK
T_{30}	244-249	785.37	-	ECHLAR

T_{31}	250-258	935.54	935.54 (2)	VPSHAVVAR
T_{32}	259-263	533.27	-	SVDGR
T_{33}	264-269	831.44	-	EDLIWR
$T_{(32+33)}$	259-269	1354.69	1345.69 (2)	SVDGREDLIWR
T_{34}	270-273	538.35	-	LLHR
T_{35}	274-280	836.39	836.39 (2)	AQEEFGR
T_{36}	281-282	262.16	-	NK
$T_{(35+36)}$	274-282	1079.51	1079.51 (2)	AQEEFGR**N**K [g]
T_{37}	283-290	927.49	927.49 (2)	SSAFQLFK
T_{38}	291-296	675.33	-	STPENK
T_{39}	297-301	635.38	-	DLLFK
$T_{(38+39)}$	291-301	1291.69	1291.69 (2)	STPENKDLLFK
T_{40}	302-309	864.46	864.46 (2)	DSALGFVR
T_{41} [*]	310-332	2520.35	2520.34 (3)	IPSQIDSGLYLGANYLTA(<u>I/L</u>)QNLR [c]
T_{42}	333-341	917.47	917.47 (2)	ETAAEVAAR
T_{43}	342-342	175.12	-	R
T_{44}	343-344	304.16	-	ER
T_{45}	345-356	1430.67	1430.67 (2)	VVWCAVGPEEER
T_{46}	357-357	147.11	-	K
$T_{(45+46)}$	345-357	1558.77	1558.77 (2)	VVWCAVGPEEERK
T_{47}	358-359	307.14	-	CK
T_{48}	360-367	991.46	991.46 (2)	QWSDVSNR
T_{49}	368-368	147.11	-	K
$T_{(48+49)}$	360-368	1119.55	1119.56 (2)	QWSDVSNRK
T_{50}	369-386	1922.96	1922.96 (2)	VACASASTTEECIALVLK
T_{51}	387-404	1809.90	1809.90 (2)	GEADALNLDGGFIYVAGK
T_{52}	405-416	1327.70	1327.70 (2)	CGLVPVLAENQK [a]
T_{53}	417-440	2936.30	2936.29 (4)	SQNSNAPDCVHRPPEGYLAVAVVR [a]
T_{54}	441-441	147.11	-	K
T_{55}	441-454	1393.66	1393.66 (2)	SDADLTWNSLSGK [a]
$T_{(54+55)}$	441-454	1521.75	1521.76 (2)	KSDADLTWNSLSGK [a]
T_{56}	455-455	147.11	-	K
T_{57}	456-463	873.40	873.40 (2)	SCHTGVGR
T_{58} [*]	464-482	2111.99	2111.99 (2)	TA<u>G</u>WNIPMGLLF**N**QTGSCK [d, g, h]
T_{58} [*]	464-482	2095.99	2096.00 (3)	TA<u>G</u>WNIPMGLLF**N**QTGSCK [d, g]
T_{59}	483-485	409.21	-	FDK
T_{60}	486-514	3201.37	3201.37 (3)	FFSQSCAPGADPQSSLCALCVGNNENENK [a]
$T_{(59+60)}$	483-514	3591.56	3591.56 (3)	FDKFFSQSCAPGADPQSSLCALCVGNNENENK
T_{61}	515-522	1022.40	-	CMPNSEER
T_{62}	523-531	1097.51	1097.51 (2)	YYGYTGAFR
T_{63}	532-536	620.31	-	CLAEK
T_{64}	537-544	806.44	806.44 (2)	AGDVAFVK
T_{65}	545-555	1189.61	1189.61 (2)	DVTVLQNTDGK [a]
T_{66}	556-562	831.40	831.40 (2)	NSEPWAK [a]
T_{67}	563-565	375.22	-	DLK
T_{68}	566-578	1595.74	1595.74 (2)	QEDFELLCLDGTR
$T_{(67+68)}$	563-578	1951.94	1951.94 (3)	DLKQEDFELLCLDGTR
T_{69}	579-591	1467.74	1467.74 (3)	KPVAEAESCHLAR
T_{70}	592-603	1280.63	1280.63 (2)	APNHAVVSQSDR [a]

T_{71}	604-608	596.35	-	AQHLK
T_{72}	609-609	147.11	-	K
T_{73}	610-628	2076.98	2076.98 (2)	VLFLQQDQFGGNGPDCPGK [a]
$T_{(72+73)}$	609-628	2205.08	2205.08 (2)	KVLFLQQDQFGGNGPDCPGK [a]
T_{74}	629-633	714.36	-	FCLFK
T_{75}	634-637	464.24	-	SETK
T_{76} [*]	638-654	1951.94	1951.94 (2)	NLLFSDNTECLAELQGK [e]
T_{77} [*]	655-673	2222.10	2222.10 (2)	TTYEQYLGSEYVTAITNLR [f]
T_{78}	674-674	175.12	-	R
T_{79}	675-688	1610.77	1610.76 (2)	CSSPLLEACAFLR
T_{80}	689-689	90.05	-	A

[a] fragments also found with deamidated asparagine;

[b] fragment carrying the substitution $Arg^{91} \rightarrow$ His compared to the lactoferrin from mare;

[c] fragment carrying the substitution $Thr^{328} \rightarrow$ Ile/Leu compared to the lactoferrin from mare;

[d] fragment carrying the substitution $Ala^{466} \rightarrow$ Gly compared to the lactoferrin from mare;

[e] fragment carrying the substitution $Asn^{642} \rightarrow$ Ser compared to the lactoferrin from mare;

[f] fragment carrying the substitution $Ser^{668} \rightarrow$ Ala compared to the lactoferrin from mare;

[g] asparagine residues found exclusively in deamidated form. These residues, reported in bold, were tentatively assumed to be the glycosylation sites;

[h] Met^{471} oxidized to methionine sulfoxide .

In order to obtain additional information, a chemical N-terminal derivatization of the tryptic peptides with SPITC, prior to MALDI-TOF/TOF MS, was performed. The presence of a positive charge at the C-terminus of arginine-terminated peptides, simplify the interpretation of the MS/MS spectra, because y-series ions are predominantly obtained. [188, 189] The MALDI-TOF MS and MS/MS analyses of derivatized tryptic mixture confirmed the results already obtained by the LC/MS/MS analysis and, in addition, allowed to increase the sequence coverage by characterizing the amino sequence of the tryptic peptide T_{34} (LLHR), not identified by the LC/MS/MS approach. The MS/MS spectrum of the fragment ions of the SPITC derivatized T_{34} protonated molecular ion (Figure 14) shows exclusively y- type ions. Additionally, immonium ions of Leu and His residues are found in the low mass range at m/z 86.1 and 110.1, respectively. The increase in molecular mass due the SPITC derivatization also contributed to the detection of the T_{34} peptide.
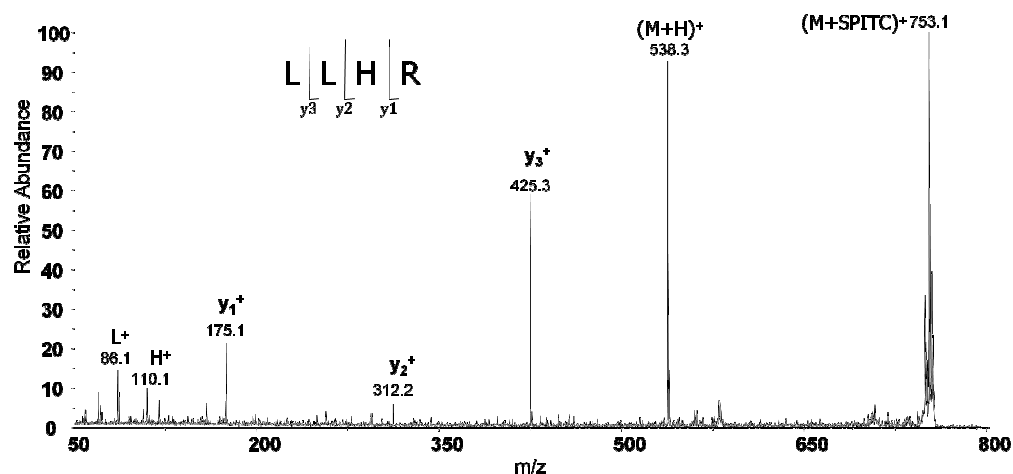
Figure 14: MS/MS spectrum of the SPITC derivatized tryptic peptide T$_{34}$ (LLHR) in the tryptic digest of donkey lactoferrin.

Complementary data were also obtained using different enzymes, such as α-chymotrypsin, AspN and GluC. Characterization of the resulting enzymatic peptide mixtures by LC/MS and the MALDI MS allowed increasing the sequence coverage by the identification of amino acid traits previously uncovered. Selected fragments identified in the different digestion mixtures and used to improve the sequence coverage are summarized in Table 7.

Table 7: Position, calculated and measured monoisotopic MH$^+$ of selected fragments, used to increase the coverage of the amino acid sequence, detected in the trypsin, α-chymotrypsin, GluC and AspN digestion mixtures of donkey lactoferrin eluted in the IEC chromatographic fraction F4 (Figure 12) and characterized by HPLC/ESI-MS/MS or MALDI-MS analysis. Amino acids not previously covered by HPLC/ESI-MS/MS of tryptic digestion are underlined. The sequence of lactoferrin (UniProtKB Acc. Nr. O77811) from mare was taken as reference

Position	Calculated MH$^+$	Measured MH$^+$ ESI-MS/MS (z)	Measured MH$^+$ MALDI-MS	Reference Sequence (UniProtKB Acc. Nr. O77811)
Trypsin				
19-28	1310.68		1310.66	CAKFQRNMKK
270-273	538.35		538.33	LLHR[a]
629-633	714.36		714.34	FCLFK
α-chymotrypsin				
1-8	999.58		999.65	APRKSVRW
9-22	1553.75	1553.74 (2)		CTISPAEAAKCAKF[c]
248-268	2333.25	2333.25 (4)		ARVPSHAVVARSVDGREDLIW [b]
272-278	916.43	916.43 (2)		HRAQEEF [c]

63

332-347	1899.03	1899.03 (4)		RETAAEVAARRERVVW [b]
347-361	1776.82		1776.95	CAVCPEEERKCKQW
531-542	1336.67	1336.66 (2)		RCLAEKAGDVAF [c]
590-607	1957.01	1957.01 (4)		ARAPNHAVVSQSDRAQHL[c]
619-629	1105.47	1105.47 (2)		GGNGPDCPGKF[b, c, d]
633-639	819.46	819.46 (2)		KSETKNL [c]
633-641	1079.61	1079.61 (2)		KSETKNLLF[b]
673-686	1653.78	1653.78 (3)		RRCSSSPLLEACAF [b, c]
GluC				
522-535	1726.80	1726.80 (2)		RYYGYTGAFRCLAE [c]
683-689	804.41	808.41 (2)		ACAFLRA[c]
AspN				
240-264	2806.42	2806.42 (5)		DAFKECHLARVPSHAVVARSVDGRE [b]
602-615	1724.00	1724.00 (3)		DRAQHLKKVLFLQQ [b]

[a] derivatized with SPITC
[b] peptide characterized by CID
[c] peptide characterized by HCD
[d] peptide also found with deamidated asparagine

In summary, the MS/MS analyses allowed to characterize almost completely the primary structure of donkey lactoferrin. Actually, only a sequence of seven amino acid of homologous mare lactoferrin remained uncovered (Figure 15).

Some recent studies report six bovine lactoferrin derived peptides, corresponding to the sequences LIWKL, RPYL, LNNSRAP, [122] LRPVAA[123] GILRPY and REPYFGY[124] as antihypertensive peptides. The first four were found to exert in vitro inhibitory effects on angiotensin I-converting enzyme (ACE) activity; the others two were found to exert in vitro inhibitory effects on endothelin-converting enzyme (ECE) activity. Inhibition of ACE and inhibition of ECE result in a lower blood pressure. Comparing the bovine lactoferrin bioactive peptide sequences to donkey lactoferrin sequence, we found that the sequences RPYL and LRPVAA are also present in donkey lactoferrin at the position 133-136 and 74-79, respectively. In addition, in donkey lactoferrin are present the sequences LIWRL, GTLRPY and QEPYFGY, at the position 266-270, 130-136, 186-192, respectively. These sequences are different from bovine lactoferrin bioactive peptides LIWKL GILRPY and REPYFGY for the substitution of Lys→Arg, Leu→Thr and Arg→Gln. The best alignment for the bovine lactoferrine peptide LNNSRAP is with the donkey lactoferrin sequence LDGTRKP at the position 574-580. This observation allows to consider donkey lactoferrin as a potential source of antihypertensive peptides as well as bovine lactoferrin. The possible antihypertensive

activity of donkey lactoferrin peptides LIWKL, GILRPY and REPYFGY, differing for a point substitution from the corresponding bioactive bovine lactoferrin peptides, needs to be investigated.

Among the identified peptides, eighteen were detected carrying deamidated Asn. Indeed, ezymatic PNGase deglycosylation converts glycosylated asparagine residues into aspartic acid. On the other hand, it is well known that chemical (i.e. non-enzymatic) deamidation of Asn and Gln residues can occur in vivo or during sample preparation prior to the PNGase treatment.[190] Chemical deamidation rate can be influenced by several factors including the protein sequence immediately surrounding the Asn residue. In particular, it has been demonstrated that the deamidation is more favoured if Asn residues are preceded by polar residues or followed by small and hydrophilic amino acids (e.g. glycine and serine), whereas bulky, hydrophobic residues

preceding Asn result in low deamidation rates.[191] N-linked glycosylation in mammalians occurs via the amide group of asparagine in the consensus tripeptide sequence Asn-X-Ser/Thr, or much less frequently Asn-X-Cys, where X can be any amino acid except proline.[192] The primary structure of donkey lactoferrin deduced from our results (Figure 15) 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, asparagines 137, 281 and 476 were found exclusively in deamidated form, tryptic peptide T_{22} containing Asn 168 was 304 detected both with unmodified and deamidated asparagine, whereas Asn 513 was never observed in deamidated form. The remaining fifteen asparagines were also detected with unmodified and deamidated asparagine (Table 6). All together these findings strongly suggest that donkey lactoferrin presents three N-glycosylation sites located at the positions 137, 281 and 476, like in the mare's homologous protein, although unambiguous assignment of glycosylation sites is not possible at this stage and further control experiment providing direct evidence are required.

```
  1 APRKSVRWCTISPAEAAKCAKFQRNMKKVRGPSVSCIRKTSSFECIQAIAANKADAVTLD
 61 GGLVYEAGLHPYKLRPVAAEVYQTRGKPQTHYYAVAVVKKGSGFQLNQLQGVKSCHTGLG
121 RSAGWNIPIGTLRPYLṄWTGPPEPLQKAVANFFSASCVPCADGKQYPNLCRLCAGTEADK
181 CACSSQEPYFGYSGAFKCLENGAGDVAFVKDSTVFENLPDEADRDKYELLCPDNTRKPVD
241 AFKECHLARVPSHAVVARSVDGREDLIWRLLHRAQEEFGRṄKSSAFQLFKSTPENKDLLF
301 KDSALGFVRIPSQIDSGLYLGANYLTALQNLRETAAEVAARRERVVWCAVGPEEERKCKQ
361 WSDVSNRKVACASASTTEECIALVLKGEADALNLDGGFIYVAGKCGLVPVLAENQKSQNS
421 NAPDCVHRPPEGYLAVAVVRKSDADLTWNSLSGKKSCHTGVGRTAGWNIPMGLLFṄQTGS
481 CKFDKFFSQSCAPGADPQSSLCALCVGNNENENKCMPNSEERYYGYTGAFRCLAEKAGDV
541 AFVKDVTVLQNTDGKNSEPWAKDLKQEDFELLCLDGTRKPVAEAESCHLARAPNHAVVSQ
601 SDRAQHLKKVLFLQQDQFGGNGPDCPGKFCLFKSETKNLLFSDNTECLAELQGKTTYEQY
661 LGSEYVTAITNLRRCSSSPLLEACAFLRA
```

Figure 15: Primary structure of the donkey lactoferrin. The characterized by HPLC/ESI-MSMS of the tryptic digest are underlined by a solid line. Additional traits identified by complementary enzymatic digestion and MALDI-MS or HPLC/ESI-MSMS analysis are underlined by a dotted line. The five point substitution with respect to the mare's counterpart (UniProtKB Acc. Nr. O77811), taken as reference, are reported in bold. The hypothesized N-glycosylation sitesare marked with an asterisk.

GLYCAN COMPOSITIONS

An aliquot of 500 µL of the dialyzed whey fraction was loaded on an analytical sulfonic column and donkey lactoferrin, previously identified (Figure 12) in fraction F4, was collected. In order to obtain a comprehensive site specific glycosylation profile of donkey lactoferrin fraction F4 was dialyzed, reduced, alkylated and digested by α-chymotrypsin. α-chymotrypsin was selected for digestion because it produces smaller glycopeptides in comparison to trypsin. RP-HPLC/nESI-MS/MS analysis was selected for glycopeptides analysis. Actually, HPLC coupled to mass spectrometry is the most commonly employed technique for the analysis of protein glycosylation. Although recent advances in mass spectrometry have made large-scale identification of proteins feasible, analysis of protein glycosylation is still very challenging, because the signals of glycopeptides are often suppressed in the presence of other peptides.[151,152] For this reason, glycopeptide enrichments were performed prior RP-HPLC/nESI-MS/MS analysis. First of all, the α-chymotriptic mixture, containing non-glycosylated peptides and glycopeptides, was enriched by TiO_2. TiO_2 has been successfully used to enrich sialylated glycopeptides from different biological samples of different levels of complexity.[167, 168, 169, 170] TiO_2-beads interact selectivity with sialylated glycopeptides removing them from the solution. The supernatant, containing only neutral glycopeptides and non-glycopeptides, was removed and stored, whereas the beads containing sialylated glycopeptides were dried and the sialylated glycopeptides eluted from beads. To enrich neutral glycopeptides, the supernatant, stored from TiO_2 enrichment, was subjected to HILIC enrichment. The flow-through contained only non-glycosylated peptides, whereas the neutral glycopeptides were eluted from the column with 2% FA.

The enriched mixtures of glycopeptides were investigated by RP-HPLC/nESI-MS/MS coupled on-line with a hybrid ESI-MS LTQ/Orbitrap mass spectrometer (Q-Exactive Plus) using HCD fragmentation and MS/MS data were used for database searching. Several studies employed only MALDI MS spectrometry to characterize glycans composition. However, an additional complication in the comprehensive characterization of the N-glycans of animal milk proteins is the presence of N-glycolineuraminic acid (NeuGc), which is generally not found in human. NeuGc prevents determination of the composition based exclusively on accurate mass

because combinations of fucose (Fuc) and NeuGc yeld masses isobaric with oligosaccharides containing N-acetylneuraminc acid (NeuAc) and hexose (Hex). For example, the neutral mass 1931.69 Da may correspond to $GlcNAc_4{:}Hex_5{:}NeuAc_1$ but can also correspond to $GlcNAc_4{:}Hex_4{:}Fuc_1{:}NeuGc_1$. To discriminate between the two compositions tandem MS is required, because MS/MS spectra of glycans or glycopeptides are charaterized by the presence of carbohydrate-specific oxonium fragment ions. Diagnostic ions at m/z 308.10 $[NeuGc+H]^+$ and 290.10 $[NeuGc{-}H_2O+H]^+$ indicate the presence of NeuGc while ions at m/z 292.10 $[NeuAc+H]^+$, m/z 274.10 $[NeuAc{-}H_2O+H]^+$ indicate the presence of NeuAc. 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 $[HexNAc+Hex+Sia+H]^+$ and 407.17 $[2HexNAc+H]^+$, are indicative for the presence of glycopeptides. Figure 16b shows the ESI-MS/MS spectrum of the glycopeptide $HexNAc_4Hex_5NeuAc_1$+GRNKSSAF. The experimental 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). Based on the determined molecular mass two glycan composition can be candicated for the GRNKSSAF peptide: $GlcNAc_4{:}Hex_5{:}NeuAc_1$ or $GlcNAc_4{:}Hex_4{:}Fuc_1{:}NeuGc_1$. However, the MS/MS spectrum confirms the $GlcNAc_4{:}Hex_5{:}NeuAc_1$ composition thanks to the presence of the m/z 292.10 $[NeuAc+H]^+$ and m/z 274.10 $[NeuAc{-}H2O+H]^+$ ions. On the other hand, in the MS/MS spectrum of the glycopeptide $HexNAc_4Hex_5NeuGc_1$+GRNKSSAF shown in Figure 16c, the carbohydrate-specific oxonium fragment ions at m/z 308.10 $[NeuGc+H]^+$ and 290.10 $[NeuGc{-}H_2O+H]^+$ indicate unequivocally the presence of $NeuGc_1$.

The glycopeptide compositions were determined by accurate mass measurements of the parent ions and from the MS/MS spectra, using an in-house developed software tool, (MassAI Bioinformayics, MassAI.at http://massai.dk/index.html). This software helps to identify glycopeptides (Figure 16a) from a Mascot Generic File (mgf), using a flexible glycan database. All results were manually checked in order to validate the accurate mass of the potential glycopeptides and the compatibility of the fragment ions observed in the MS/MS spectra.
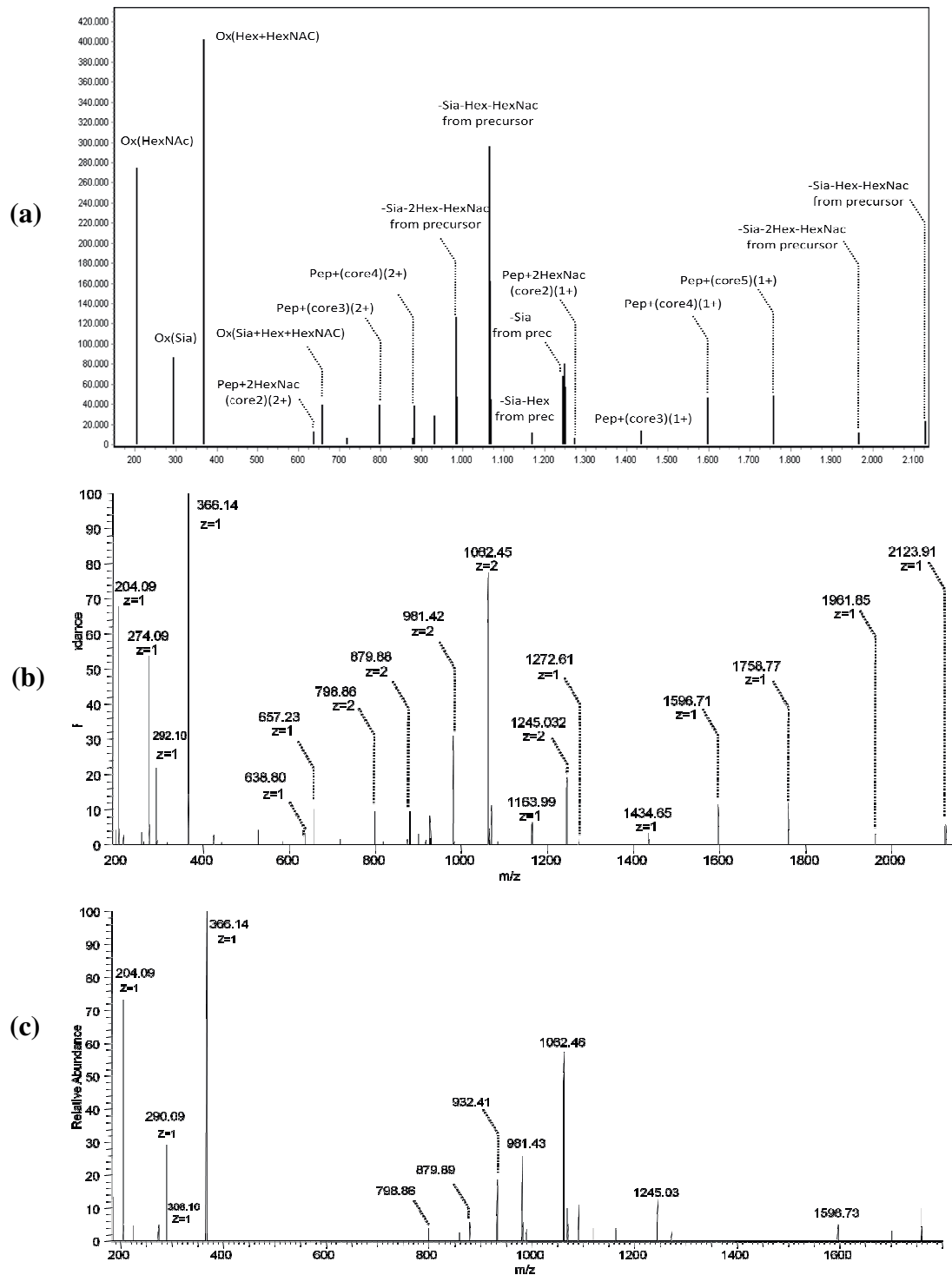
Figure 16: (a) MassAI identification of the glycopeptide HexNAc₄Hex₅NeuAc₁+GRNKSSAF; (b) MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₄Hex₅NeuAc₁+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lattoferrin. The experimental determined molecular 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) MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₄Hex₅NeuGc₁+GRNKSSAF. The experimental 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).

N-linked glycosylation in mammalian occurs via the amide group of asparagine in the consensus tripeptide sequence Asn-X-Ser/Thr, or, much less frequently, Asn-X-Cys, where X can be any amino acid except proline. [194] The primary structure of donkey lactoferrin 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 were found in glycosilated form. In this study 44 glycopeptides were identified, as reported in Table 8.

Table 8: N-Glycopeptides identified in donkey lactoferrin.

N-Glycan Composition	N-Glycan residue (-H$_2$O)	Peptide Sequence	Molecular Mass		Mass Error	
			Theor.	Exp.	ppm	Da
			Site 137			
HexNAc$_4$Hex$_5$	1622.58	LNWTGPPEPLQKAVANF	3503.561	3503.578	5	0.017
HexNAc$_4$Hex$_5$NeuAc$_1$	1913.68	LNW	2344.894	2344.889	2	0.005
HexNAc$_5$Hex$_4$	1663.61	LNWTGPPEPLQKAVANFF	3691.655	3691.646	2	0.009
HexNAc$_5$Hex$_4$Fuc$_1$NeuAc$_1$	2100.76	LNW	2531.978	2532.015	15	0.037
HexNAc$_5$Hex$_5$Fuc$_1$	1971.72	LNWTGPPEPLQKAVANFF	3999.763	3999.741	6	0.022
HexNAc$_6$Hex$_5$	2028.74	LNWTGPPEPLQKAVANFF	4056.787	4056.765	5	0.022
			Site 281			
HexNAc$_2$Hex$_5$	1216.42	GRNKSSAF	2081.864	2081.866	1	0.002
HexNAc$_3$Hex$_3$	1095.40	GRNKSSAF	1960.838	1960.805	16	0.032
HexNAc$_3$Hex$_4$	1257.40	GRNKSSAF	2122.891	2122.891	0	0.000
HexNAc$_3$Hex$_4$NeuAc$_1$	1548.545	GRNKSSAF	2413.986	2413.985	0	0.001
HexNAc$_4$Hex$_4$	1460.53	GRNKSSAF	2325.970	2325.969	0	0.001
HexNAc$_4$Hex$_4$NeuAc$_1$	1751.62	GRNKSSAF	2617.065	2617.053	5	0.012
HexNAc$_4$Hex$_5$	1622.58	GRNKSSAF	2488.023	2488.023	0	0.000
HexNAc$_4$Hex$_5$NeuAc$_1$	1913.68	GRNKSSAF	2779.118	2779.116	1	0.002
HexNAc$_4$Hex$_5$NeuGc$_1$	1929.67	GRNKSSAF	2795.113	2795.111	1	0.002
HexNAc$_4$Hex$_5$Fuc$_1$	1768.64	GRNKSSAF	2634.081	2634.088	3	0.007
HexNAc$_4$Hex$_5$Fuc$_1$NeuAc$_1$	2059.735	GRNKSSAFQLF	3313.387	3313.341	4	0.046
HexNAc$_5$Hex$_3$	1501.555	GRNKSSAF	2366.996	2366.995	0	0.001
HexNAc$_5$Hex$_4$	1663.61	GRNKSSAF	2529.049	2529.049	0	0.000
HexNAc$_5$Hex$_4$NeuAc$_1$	1954.70	GRNKSSAF	2820.145	2820.145	0	0.000
HexNAc$_5$Hex$_5$Fuc$_1$	1971.72	GRNKSSAF	2837.160	2837.126	13	0.036
HexNAc$_5$Hex$_6$	1987.71	GRNKSSAF	2853.155	2853.161	2	0.006
HexNAc$_5$Hex$_6$Fuc$_1$NeuAc$_1$	2424.87	GRNKSSAFQLF	3678.519	3678.529	3	0.010
HexNAc$_6$Hex$_5$	2028.74	GRNKSSAF	2894.181	2894.143	13	0.038
HexNAc$_6$Hex$_5$Fuc$_1$	2174.80	GRNKSSAFQLF	3428.450	3428.465	4	0.015

HexNAc$_6$Hex$_6$	2190.79	GRNKSSAFQLF	3444.445	3444.442	1	0.003
HexNAc$_6$Hex$_7$Fuc$_1$NeuAc$_1$	2790,00	GRNKSSAFQLF	3655.440	3655.520	23	0.080

N-Glycan Composition	N-Glycan residue (-H$_2$O)	Peptide Sequence	Site 476			
			Molecular Mass		Mass Error	
			Theor.	Exp.	ppm	Da
HexNAc$_2$Hex$_5$	1216.42	NQTGSCKF	2156.829	2156.827	0	0.002
HexNAc$_2$Hex$_6$	1378.48	NQTGSCKF	2318.882	2318.887	2	0.005
HexNAc$_3$Hex$_4$	1257.40	NQTGSCKF	2197.856	2197.856	0	0.000
HexNAc$_3$Hex$_4$NeuAc$_1$	1548.545	NQTGSCKF	2488.951	2488.957	2	0.006
HexNAc$_4$Hex$_4$	1460.53	NQTGSCKFDKFF	2938.193	2938.181	5	0.012
HexNAc$_4$Hex$_4$NeuAc$_1$	1751.62	NQTGSCKFDKFF	3229.288	3229.285	1	0.003
HexNAc$_4$Hex$_5$	1622.58	NQTGSCKF	2562.988	2562.987	0	0.001
HexNAc$_4$Hex$_5$NeuAc$_1$	1913.68	NQTGSCKF	2854.083	2854.082	0	0.001
HexNAc$_4$Hex$_5$Fuc$_1$NeuAc$_1$	2059.735	NQTGSCKF(Q->E)	3001.125	3001.151	9	0.026
HexNAc$_4$Hex$_5$NeuAc$_2$	2204.77	NQTGSCKF	3145.178	3145.176	1	0.002
HexNAc$_5$Hex$_4$	1663.608	NQTGSCKF	2604.014	2604.034	8	0.020
HexNAc$_5$Hex$_4$NeuAc$_1$	1954.70	NQTGSCKF	2895.110	2895.110	0	0.000
HexNAc$_5$Hex$_5$Fuc$_1$	1971.72	NQTGSCKF	2912.125	2912.093	11	0.032
HexNAc$_5$Hex$_6$Fuc$_1$	2204.77	NQTGSCKF(Q->E)	3116.188	3116.187	0	0.001
HexNAc$_6$Hex$_3$	1704.635	NQTGSCKF	2645.041	2645.062	8	0.021
HexNAc$_6$Hex$_5$	2028.74	NQTGSCKF	2969.146	2969.168	7	0.022
HexNAc$_6$Hex$_5$Fuc$_1$	2178.80	NQTGSCKF	3145.178	3145.176	1	0.002

Detailed analysis of the results summarized in Table 8 show that Asn 281 and Asn 476 are the sites bearing the major number of different glycans compositions in donkey LF. Actually 21 and 17 glycosylated peptides were found, respectively. Conversely, Asn 137 presents a minor number of glycans compositions (6 glycopeptides). Moreover, five of the six glycan compositions identified, linked at the Asn in position 137, were found also linked at Asn in position 281 and 476. 13 glycan compositions were found in common between Asn in positions 281 and 476, while 12 glycan compositions were found differently distributed between Asn in position 281 and 476, as reported in Table 9.

Table 9: N-Glycan compositins identified at Asparagine in the positions 137, 281 and 476 in donkey's lactoferrin.

N-Glycan Composition	N-Glycan Residue (-H$_2$O)	Site 137	Site281	Site 476
HexNAc$_2$Hex$_5$	1216.42		*	*
HexNAc$_2$Hex$_6$	1378.48			*
HexNAc$_3$Hex$_3$	1095.40		*	
HexNAc$_3$Hex$_4$	1257.40		*	*

N-Glycan Composition	N-Glycan Residue (-H2O)	Site 137	Site281	Site 476
HexNAc$_3$Hex$_4$NeuAc$_1$	1548.545		*	*
HexNAc$_4$Hex$_4$	1460.53		*	*
HexNAc$_4$Hex$_4$NeuAc$_1$	1751.62		*	*
HexNAc$_4$Hex$_5$	1622.58	*	*	*
HexNAc$_4$Hex$_5$Fuc$_1$	1768.64		*	
HexNAc$_4$Hex$_5$NeuAc$_1$	1913.68	*	*	*
HexNAc$_4$Hex$_5$NeuGc$_1$	1929.67		*	
HexNAc$_4$Hex$_5$Fuc$_1$NeuAc$_1$	2059.735		*	*
HexNAc$_4$Hex$_5$NeuAc$_2$	2204.77			*
HexNAc$_5$Hex$_3$	1501.555		*	
HexNAc$_5$Hex$_4$	1663.61	*	*	*
HexNAc$_5$Hex$_4$NeuAc$_1$	1954.70		*	*
HexNAc$_5$Hex$_4$Fuc$_1$NeuAc$_1$	2100.76	*		
HexNAc$_5$Hex$_5$Fuc$_1$	1971.72	*	*	*
HexNAc$_5$Hex$_6$	1987.71		*	
HexNAc$_5$Hex$_6$Fuc$_1$	2133.77			*
HexNAc$_5$Hex$_6$Fuc$_1$NeuAc$_1$	2424.87		*	
HexNAc$_6$Hex$_3$	1704.635			*
HexNAc$_6$Hex$_5$	2028.74	*	*	*
HexNAc$_6$Hex$_5$Fuc$_1$	2174.79		*	*
HexNAc$_6$Hex$_6$	2190.79		*	
HexNAc$_6$Hex$_7$Fuc$_1$NeuAc$_1$	2790.00		*	

All togheter the N-glycan compositions determined in donkey milk lactoferrin revealed that most of the glycans identified are neutral complex/hybrid. Actually, 10 neutral non-fucosylated hybrid/complex glycans and 4 neutral fucosylated hybrid/complex glycans were found. In addition, 2 high mannose glycans, 4 sialylated fucosylated hybrid/complex glycans and 6 sialylated non-fucosylated glycans, one of which with N-glycolyl neuramin acid (NeuGc), are present (Table 10).

The results above described allowed for the first time the determination of the most comprensive glycosylation profile of donkey lactoferrin with 44 glycopeptides identified involving the asparagine residues located at the positions 137, 281 and 476.

The glycosilation profile of lactoferrin in human, bovine and goat has been described.[131, 132, 133] From the comparison of the N-glycans compositions (Table 10 and 11), high mannose N-glycans were found only in bovine, goat and donkey lactoferrin, whereas this composition is not present in human lactoferrin. The majority of the N-glycans in donkey lactoferrin are neutral N-glycans (38.45%) (Table 11). This value is similar to that of bovine lactoferrin and goat lactoferrin (35.48% and 28.12% of all N-glycans, respectively) while in human lactoferrin is only 7.70%. Also, the total

percentage of fucosylation in donkey lactoferrin (30.78%) is similar to that of bovine lactoferrin and goat lactoferrin (38.72% and 34.38%), while is lower than in human lactoferrin (88.50%). The total percentage of sialylation in donkey lactoferrin (38.46%) is lower than in human milk lactoferrin (53.80%), similar than goat milk lactoferrin (37.50%) but higher than in bovine milk lactoferrin (6.45%). Furthermore, the total percentage of sialylation in goat lactoferrin (37.50%), is due for 25% to sialylated complex/hybrid and sialylated fucosylated complex/hybrid with N-glycolylneuramin acid (NeuGc), which is generally not present in humans, while only the 12.50% presents N-acetylneuraminic acid (NeuAc) sialylation. In contrast, the total percentage of sialylation in donkey milk lactoferrin (38.46%), is due for 34.62% to sialylated complex/hybrid and sialylated fucosylated complex/hybrid with N-acetylneuraminic acid (NeuAc) while only 3.84% contains the N-glycolylneuramin acid (NeuGc). Prior this study, N-glycolylneuramin acid (NeuGc) was only found in goat lactoferrin and in bovine milk (table 11).

Table 10: N-Glycan compositions present in human, bovine, goat and donkey lactoferrin. In the last two columns, N-Glycan compositions present in human and bovine milk are also reported.

N-Glycan composition	N-Glycan Residue ($-H_2O$)	Human LF[1]	Bovine LF[2]	Goat LF[3]	Donkey LF[3]	N-Glycome in Human Milk[4]	N-Glycome in Bovine Milk[4]
High Mannose							
$HexNAc_2Hex_4$	1054.37		*	*			
$HexNAc_2Hex_5$	1216.42		*	*	*	*	*
$HexNAc_2Hex_6$	1378.48		*	*	*	*	*
$HexNAc_2Hex_7$	1540.53		*	*		*	*
$HexNAc_2Hex_8$	1702.58		*	*		*	*
$HexNAc_2Hex_9$	1864.63		*			*	*
Neutral Complex/Hybrid							
$HexNAc_2Hex_4$	1054.37					*	
$HexNAc_3Hex_3$	1095.40		*		*		
$HexNAc_3Hex_4$	1257.45		*		*		
$HexNAc_3Hex_5$	1419.50		*	*		*	*
$HexNAc_3Hex_6$	1581.56		*	*			*
$HexNAc_3Hex_7$	1743.61			*			*
$HexNAc_4Hex_3$	1298.48		*	*			*
$HexNAc_4Hex_4$	1460.53	+	*	*	*	*	*
$HexNAc_4Hex_5$	1622.58	*+	*	*	*	*	*
$HexNAc_4Hex_6$	1784.63		*	*			*
$HexNAc_5Hex_3$	1501.56		*		*		
$HexNAc_5Hex_4$	1663.61		*	*	*		*
$HexNAc_5Hex_6$	1987.71				*		*

N-Glycan composition	N-Glycan Residue (-H₂O)	Human LF[1]	Bovine LF[2]	Goat LF[3]	Donkey LF[3]	N-Glycome in Human Milk[4]	N-Glycome in Bovine Milk[4]
Neutral Complex/Hybrid							
HexNAc₆Hex₃	1704.63		*	*	*		*
HexNAc₆Hex₅	2028.74				*		
HexNAc₆Hex₆	2190.79				*		
HexNAc₇Hex₈	2717.98					*	
Neutral Fucosylated Complex/Hybrid							
HexNAc₂Hex₃Fuc₁	1038.38					*	
HexNAc₂Hex₄Fuc₁	1200.43					*	
HexNAc₃Hex₃Fuc₁	1241.45	+	*				
HexNAc₃Hex₄Fuc₁	1403.51	+	*				
HexNAc₃Hex₄Fuc₂	1549.57		*				
HexNAc₃Hex₅Fuc₁	1565.56		*	*		*	
HexNAc₃Hex₅Fuc₂	1711.62					*	
HexNAc₃Hex₆Fuc₁	1873.67		*				
HexNAc₄Hex₃Fuc₁	1444.53		*	*			*
HexNAc₄Hex₃Fuc₂	1590.59		*				
HexNAc₄Hex₄Fuc₁	1606.58	+	*	*		*	*
HexNAc₄Hex₄Fuc₂	1752.64		*				
HexNAc₄Hex₅Fuc₁	1768.64	*+	*	*	*	*	*
HexNAc₄Hex₅Fuc₂	1914.70	*+				*	*
HexNAc₄Hex₅Fuc₃	2060.76	*+				*	*
HexNAc₄Hex₆Fuc₁	1930.69			*		*	*
HexNAc₄Hex₆Fuc₂	2076.75					*	*
HexNAc₄Hex₆Fuc₃	2222.81					*	*
HexNAc₅Hex₃Fuc₁	1647.61		*				
HexNAc₅Hex₃Fuc₂	1793.67		*				
HexNAc₅Hex₄Fuc₁	1809.67			*			*
HexNAc₅Hex₅Fuc₁	1971.72				*		*
HexNAc₅Hex₆Fuc₁	2133.77	+			*		*
HexNAc₅Hex₆Fuc₂	2279.83	*+					
HexNAc₅Hex₆Fuc₄	2571.95	+					
HexNAc₆Hex₃Fuc₁	1850.69						*
HexNAc₆Hex₃Fuc₂	1996.75					*	
HexNAc₆Hex₃Fuc₃	2142.81						*
HexNAc₆Hex₄Fuc₁	2012.75					*	
HexNAc₆Hex₄Fuc₂	2158.80					*	
HexNAc₆Hex₅Fuc₁	2174.79				*		
HexNAc₆Hex₆Fuc₂	2482.91					*	
HexNAc₆Hex₇Fuc₁	2498.90	+					
HexNAc₆Hex₇Fuc₃	2791.02						*
HexNAc₇Hex₃Fuc₂	2199.83					*	
HexNAc₇Hex₃Fuc₃	2345.89					*	
Sialylated Complex/Hybrid							
HexNAc₃Hex₄NeuAc₁	1548.55				*		
HexNAc₃Hex₅NeuAc₁	1710.63			*		*	*
HexNAc₃Hex₅NeuGc₁	1726.59			*			
HexNAc₃Hex₆NeuAc₁	1872.65					*	*
HexNAc₃Hex₆NeuGc₁	1888.65			*			
HexNAc₃Hex₇NeuGc ₁	2050.70			*			
HexNAc₄Hex₃NeuAc₁	1589.57		*				*

N-Glycan composition	N-Glycan Residue (-H$_2$O)	Human LF[1]	Bovine LF[2]	Goat LF[3]	Donkey LF	N-Glycome in Human Milk[4]	N-Glycome in Bovine Milk[4]
HexNAc$_4$Hex$_4$NeuAc$_1$	1751.62				*		*
HexNAc$_4$Hex$_4$NeuGc$_1$	1767.62			*			
HexNAc$_4$Hex$_5$NeuAc$_1$	1913.68	*+		*	*	*	*
HexNAc$_4$Hex$_5$NeuAc$_2$	2204.77				*	*	*
HexNAc$_4$Hex$_5$NeuGc$_1$	1929.67			*	*		*
HexNAc$_4$Hex$_5$NeuGc$_2$	2236.76						*
HexNAc$_4$Hex$_5$NeuAc$_1$NeuGc$_1$	2220.77						*
HexNAc$_4$Hex$_6$NeuAc$_1$NeuGc$_1$	2382.82						*
HexNAc$_5$Hex$_3$NeuAc$_1$	1792.65		*				
HexNAc$_5$Hex$_4$NeuAc$_1$	1954.70				*		*
HexNAc$_5$Hex$_4$NeuGc$_1$	1970.70						*
HexNAc$_5$Hex$_5$NeuAc$_1$	2116.76						*
HexNAc$_6$Hex$_3$NeuAc$_1$	1995.73						*
Sialylated Fucosylated Complex/Hybrid							
HexNAc$_3$Hex$_4$Fuc$_1$NeuAc$_1$	1694.60					*	
HexNAc$_3$Hex$_5$Fuc$_1$NeuAc$_1$	2018.71					*	
HexNAc$_3$Hex$_6$Fuc$_1$NeuAc$_1$	1856.66						*
HexNAc$_3$Hex$_6$Fuc$_1$NeuGc$_1$	2034.70			*			
HexNAc$_4$Hex$_4$Fuc$_1$NeuAc$_1$	1897.68			*			*
HexNAc$_4$Hex$_5$Fuc$_1$NeuAc$_1$	2059.74	*+		*	*	*	*
HexNAc$_4$Hex$_5$Fuc$_1$NeuGc$_1$	2075.73			*			*
HexNAc$_4$Hex$_5$Fuc$_1$NeuGc$_2$	2382.82						*
HexNAc$_4$Hex$_5$Fuc$_2$NeuAc$_1$	2205.79	*+				*	
HexNAc$_4$Hex$_5$Fuc$_3$NeuAc$_1$	2351.85	+				*	
HexNAc$_4$Hex$_5$Fuc$_1$NeuAc$_2$	2350.83	*+					
HexNAc$_5$Hex$_4$Fuc$_1$NeuAc$_1$	2100.76				*	*	*
HexNAc$_5$Hex$_4$Fuc$_1$NeuGc$_1$	2116.76			*			
HexNAc$_5$Hex$_6$Fuc$_1$NeuAc$_1$	2424.87	*			*		
HexNAc$_5$Hex$_6$Fuc$_2$NeuAc$_1$	2570.93	*+					
HexNAc$_5$Hex$_6$Fuc$_3$NeuAc$_1$	2716.98	*				*	
HexNAc$_5$Hex$_6$Fuc$_4$NeuAc$_1$	2863.04	*				*	
HexNAc$_6$Hex$_4$Fuc$_1$NeuAc$_1$	2303.84	+					
HexNAc$_6$Hex$_7$Fuc$_1$NeuAc$_1$	2790.00	*			*		
HexNAc$_6$Hex$_7$Fuc$_2$NeuAc$_1$	2936.06	*					
HexNAc$_6$Hex$_7$Fuc$_3$NeuAc$_1$	3082.12	*					
HexNAc$_6$Hex$_7$Fuc$_4$NeuAc$_1$	3228.17	*					
HexNAc$_7$Hex$_3$Fuc$_1$NeuGc$_2$	2667.95						*

1*T. Yu et al (2011);[129] +M. Barboza et al (2012)[130];

2*C. C. Nwosu et al (2010);[131]

3*A. Le Parc et al (2014);[133]

4*C. C. Nwous et al (2012).[138]

Table 11 : Comparision of N-Glycan Type linked at human, bovine, goat and donkey lactoferrin.

N-Glycan Composition	Human LF		Bovine LF		Goat LF		Donkey LF	
	N-Glycan found	Percentage %	N-Glycan found	Percentage %	N-Glycan found	Percentage %	N-Glycan found	Percentage %
High Mannose	0	0	6	19.35	5	15.63	2	7.70
Neutral Complex/Hybrid	2	7.70%	11	35.48	9	28.12	10	38.45
Neutral Fucosylated Complex/Hybrid	10	38.50%	12	38.72	6	18.75	4	15.39
Sialylated Complex/Hybrid	1	3.80%	2 (2 | 0*)	6.45 (6.45 | 0*)	7 (2 | 5*)	21,88 (6.25 | 15.63*)	6 (5 | 1)	23.07 (19.23 | 3.84)
Sialylated Fucosylated Complex/Hybrid	13	50.00%	0 (0 | 0*)	0 (0 | 0*)	5 (2 | 3*)	15.63 (6.25 | 9.38*)	4 (4 | 0)	15.39 (15.39 | 0)

*Glycans containing N-glycolyl neuramin acid (NeuGc)

CONCLUSIONS

Summarizing our result, the primary structure of donkey lactoferrin was almost completely characterized by coupling cationic exchange purification, enzymatic digestions, derivatization of tryptic digest, and different mass spectrometric platforms. Our data showed that donkey lactoferrin presents five amino acid point substitutions at positions 91 (Arg→ His), 328 (Thr→ Ile/Leu), 466 (Ala→ Gly), 642 (Asn→ Ser), and 668 (Ser→ Ala) with respect to the homologue mare lactoferrin. A sequence of seven amino acid of the homologue mare lactoferrin remained uncovered. In addition, our result allowed to know that donkey lactoferrin sequence contains the sequences RPYL and LRPVAA at the position (133-136) and (74-79), respectively. These sequences, also present in bovine lactoferrin, are known to be antihypertensive peptides. [122, 123]In addition, donkey lactoferrin contains the sequences LIWRL, GTLRPY and QEPYFGY, at the position (266-270), (130-136), (186-192), respectively. These sequences differ from bovine lactoferrin antihypertensive peptides LIWKL, GILRPY and REPYFGY[124] for a point substitution of one amino acid. This observation allows to consider donkey lactoferrin as a potential source of antihypertensive peptides as well as bovine lactoferrin. The possible antihypertensive activity of donkey lactoferrin peptides LIWKL, GILRPY and REPYFGY, differing for a point substitution from the corresponding bioactive bovine lactoferrin peptides, needs to be investigated.

Finally, the data obtained allow the identification of 26 different N-glycans compositions linked at asparagine residues located at the positions 137, 281 and 476. All togheter the N-glycan compositions determined revealed that in donkey milk lactoferrin most of the N-glycans identified are neutral complex/hybrid. Actually, 10 neutral non-fucosylated hybrid/complex glycans and 4 neutral fucosylated hybrid/complex glycans were found. In addition, 2 high mannose glycans, 4 sialylated fucosylated hybrid/complex glycans and 6 sialylated non-fucosylated glycans, one of which with N-glycolyl neuramin acid (NeuGc), are present. The amount of the neutral N-glycans in donkey lactoferrin (38.45%) is similar to that of bovine lactoferrin and goat lactoferrin (35.48% and 28,12% of all N-glycans, respectively) while in human lactoferrin is only 7.70%. Also, the total percentage of fucosylation in donkey lactoferrin (30.78%) is similar to that of bovine lactoferrin and goat lactoferrin (38.72%

and 34.38%), while is lower than in human lactoferrin (88.50%). Instead, the total percentage of sialylation in donkey lactoferrin (38.46%) is lower than in human milk lactoferrin (53.80%), similar than goat milk lactoferrin (37.50%) but higher than in bovine milk lactoferrin (6.45%). In addition, the total percentage of sialylation in goat lactoferrin (37.50%), is due for 25% to sialylated complex/hybrid and sialylated fucosylated complex/hybrid with N-glycolyl neuramin acid (NeuGc), which is generally not found in humans, while only the 12.50% presents N-acetylneuraminic acid (NeuAc) sialylation. In contrast, the total percentage of sialylation in donkey milk lactoferrin (38.46%), is due for 34.62% to sialylated complex/hybrid and sialylated fucosylated complex/hybrid with N-acetylneuraminic acid (NeuAc) while only 3.84% contains the N-glycolyl neuramin acid (NeuGc). The site specific N-glycan compositions elucidated in this study could enable future investigations on the relationship between glycosylation pattern and protein's function.
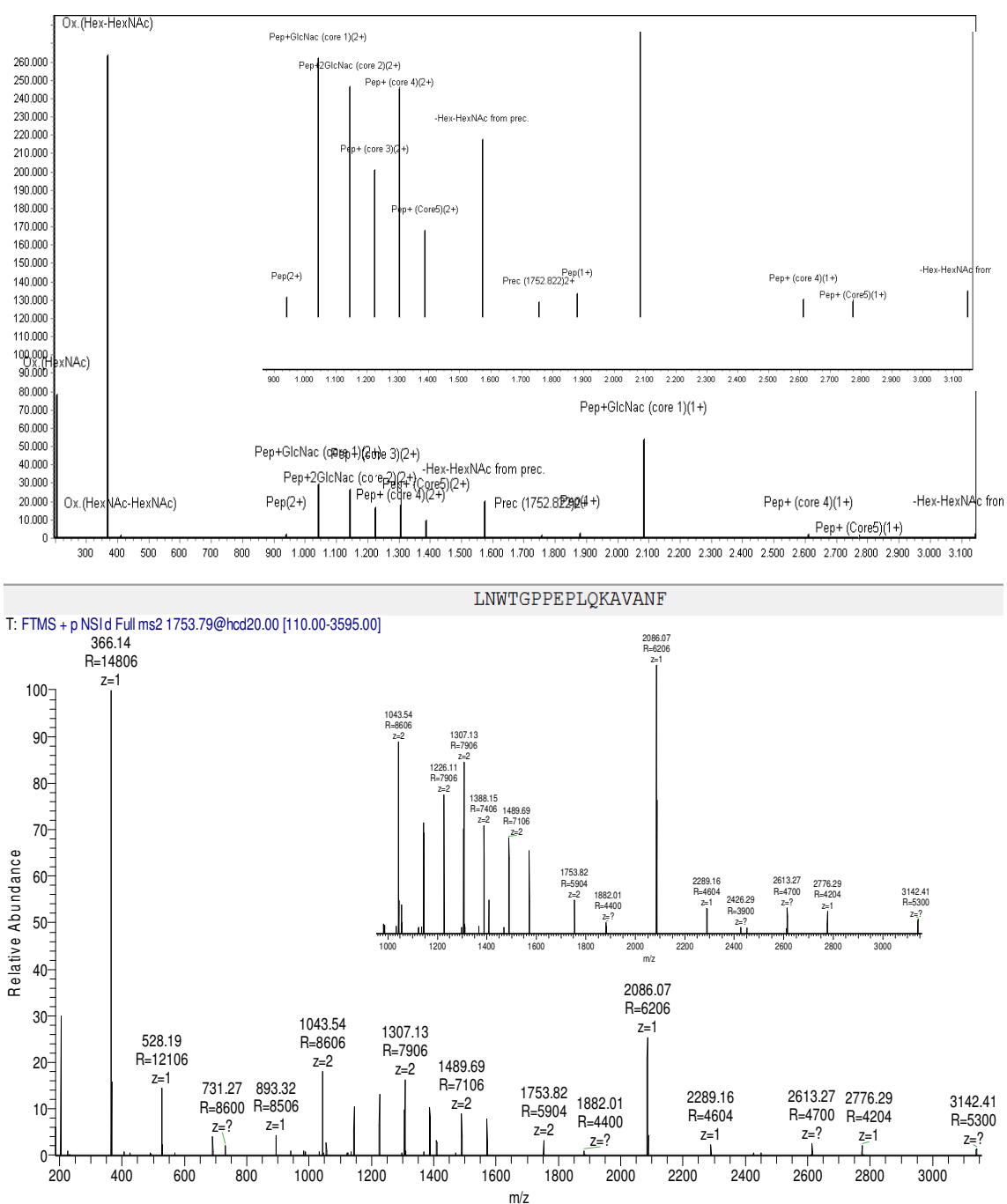
FIG.1: Bottom: ESI-MS/MS spectrum of the doubly charged molecular ion of the glycopeptide HexNAc$_4$Hex$_5$+LNWTGPPEPLQKAVANF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lattoferrin. The experimental determined molecular mass of the glycopeptide is 3503,578 Da, which corresponds to the theoretical one 3503,561, with an error of 0.017 Da (5 ppm). Top: MassAI identification.
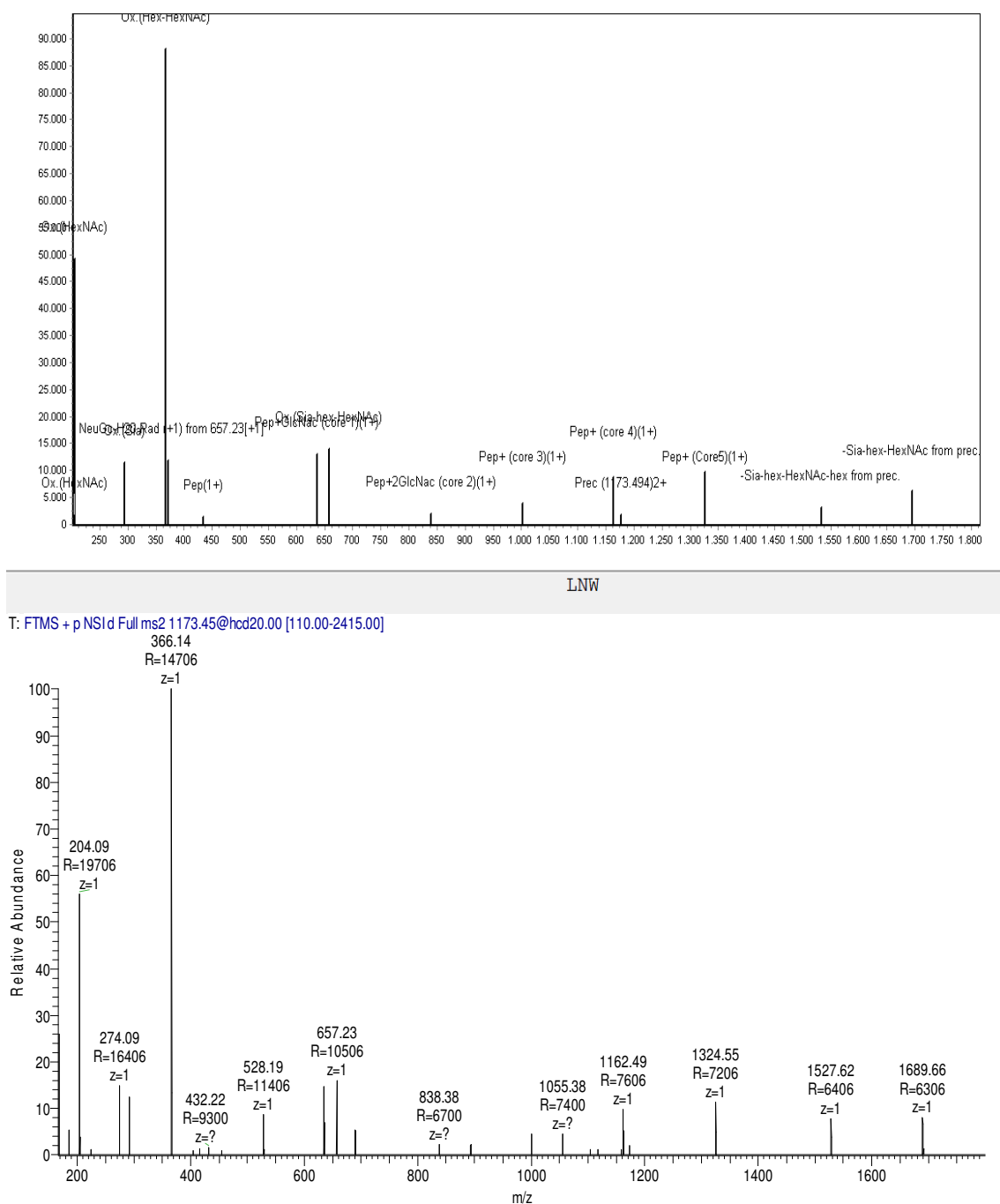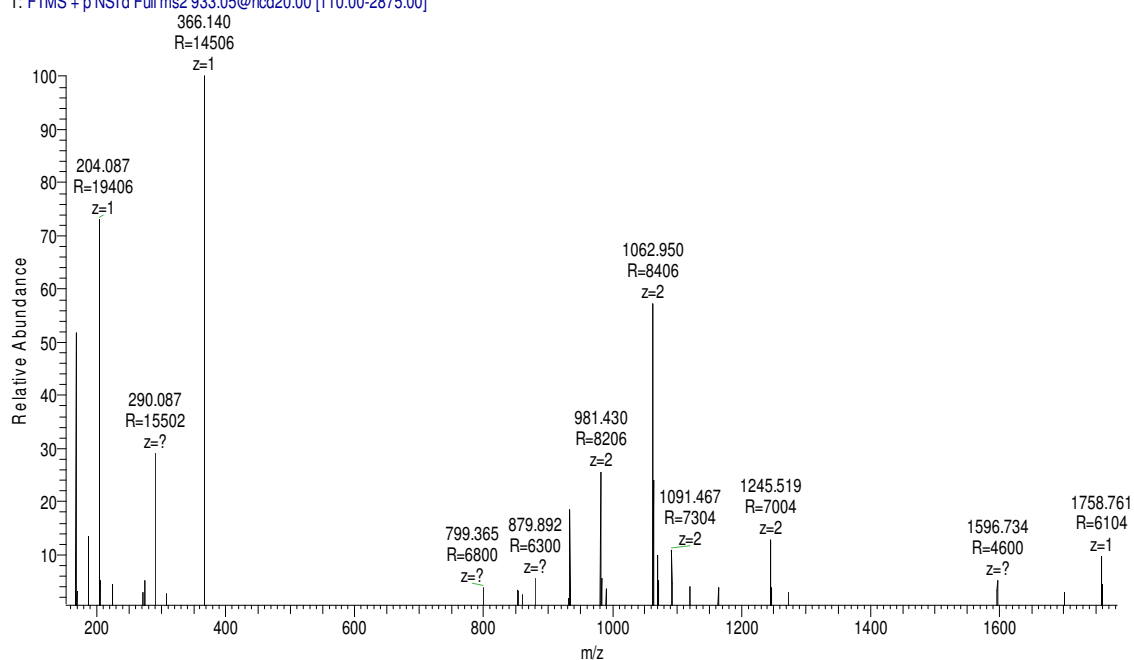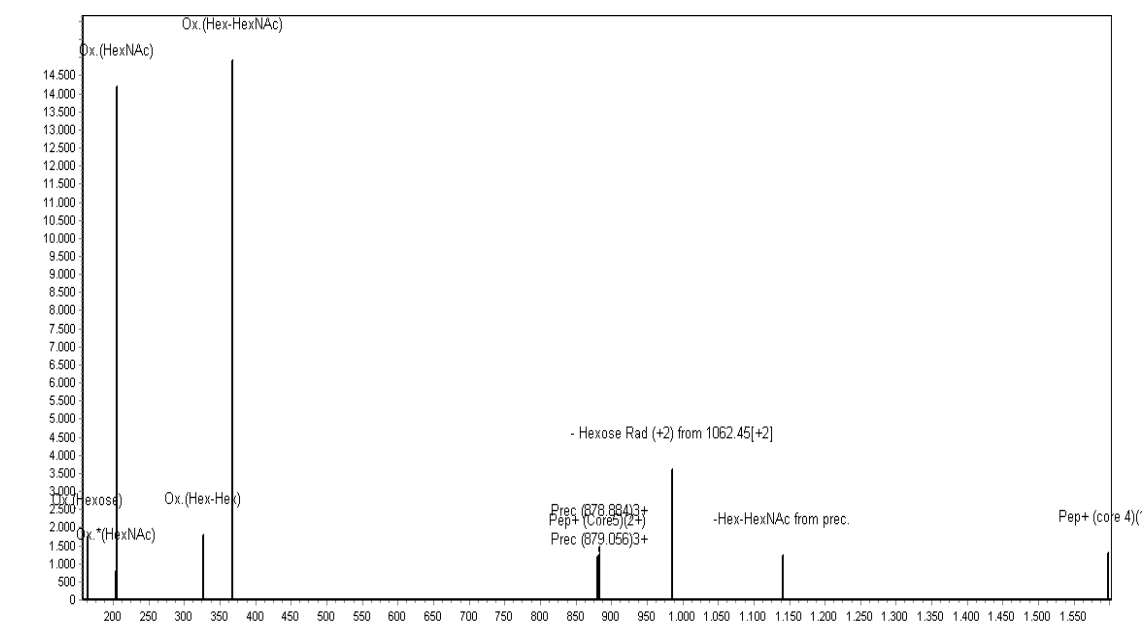
FIG. 2: Bottom: MS/MS spectrum of the doubly charged molecular ions of the glycopeptide HexNAc$_4$Hex$_5$NeuAc$_1$+LNW in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lattoferrin The experimental determined molecular mass of the glycopeptide is 2344,889 Da, which corresponds to the theoretical one 2344,894, with an error of 0.005 Da (2 ppm). Top: MassAI identification.
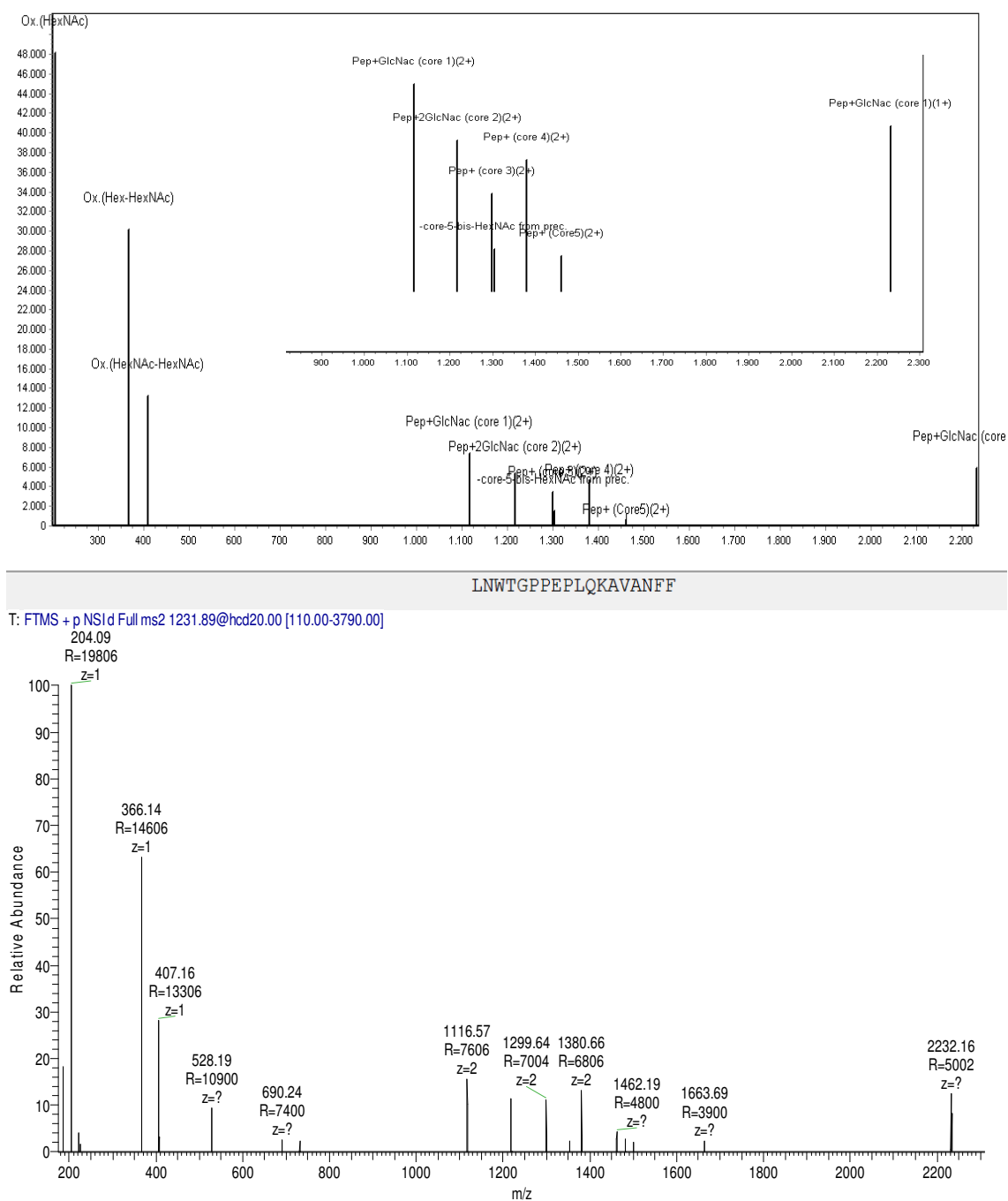
FIG. 3: Bottom: MS/MS spectrum of the triply charged molecular ions of the glycopeptide HexNAc$_5$Hex$_4$+LNWTGPPEPLQKAVANFF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lattoferrin. The experimental determined molecular mass of the glycopeptide is 3691,646 Da, which corresponds to the theoretical one 3691,655, with an error of 0.009 Da (2 ppm). Top: MassAI identification.
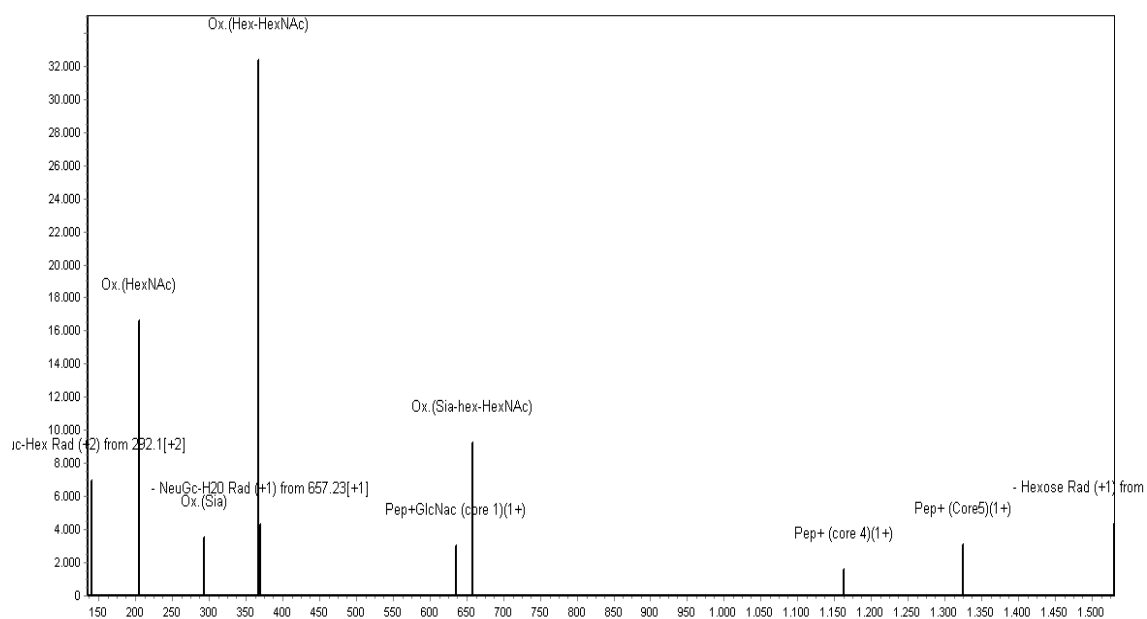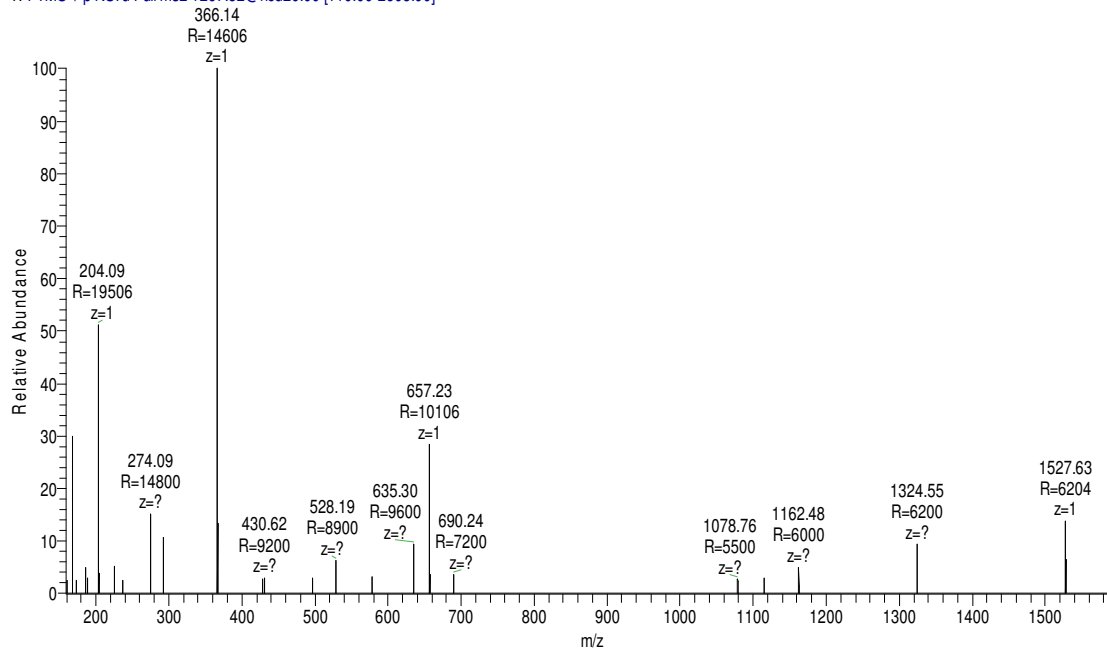
FIG. 4: Bottom: MS/MS spectrum of the doubly charged molecular ions of the glycopeptide HexNAc$_5$Hex$_4$Fuc$_1$NeuAc$_1$+LNW identified in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lattoferrin. The experimental determined molecular mass of the glycopeptide is 2532,015 Da, which corresponds to the theoretical one 2531,978, with an error of 0.037 Da (15 ppm). Top: MassAI identification.
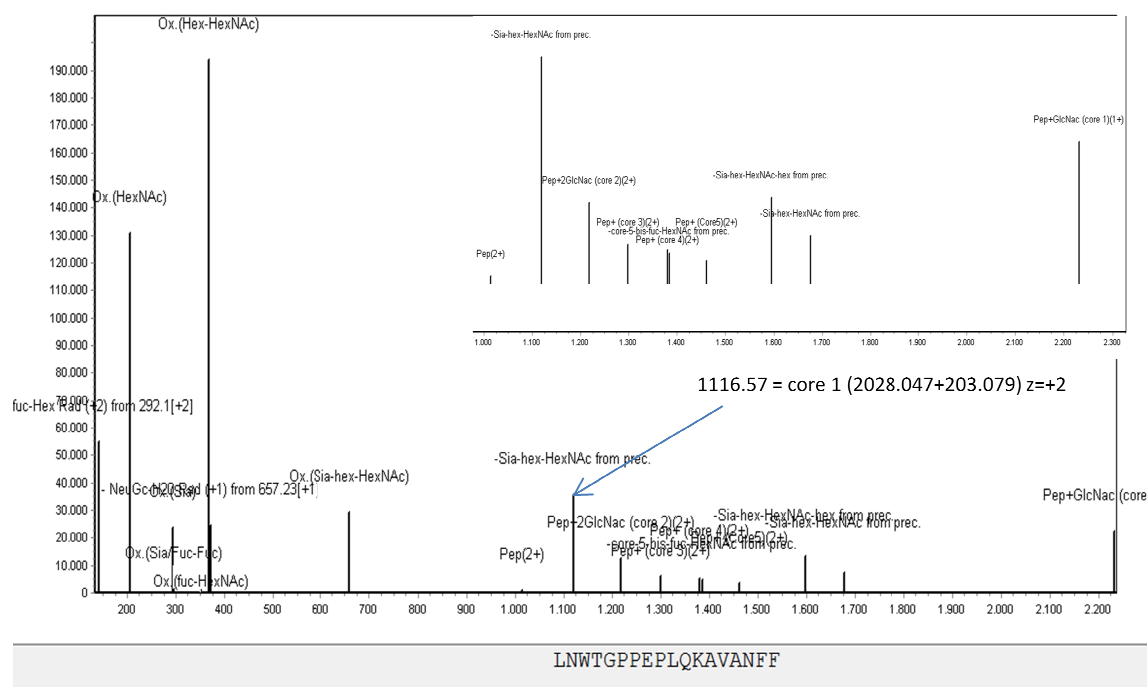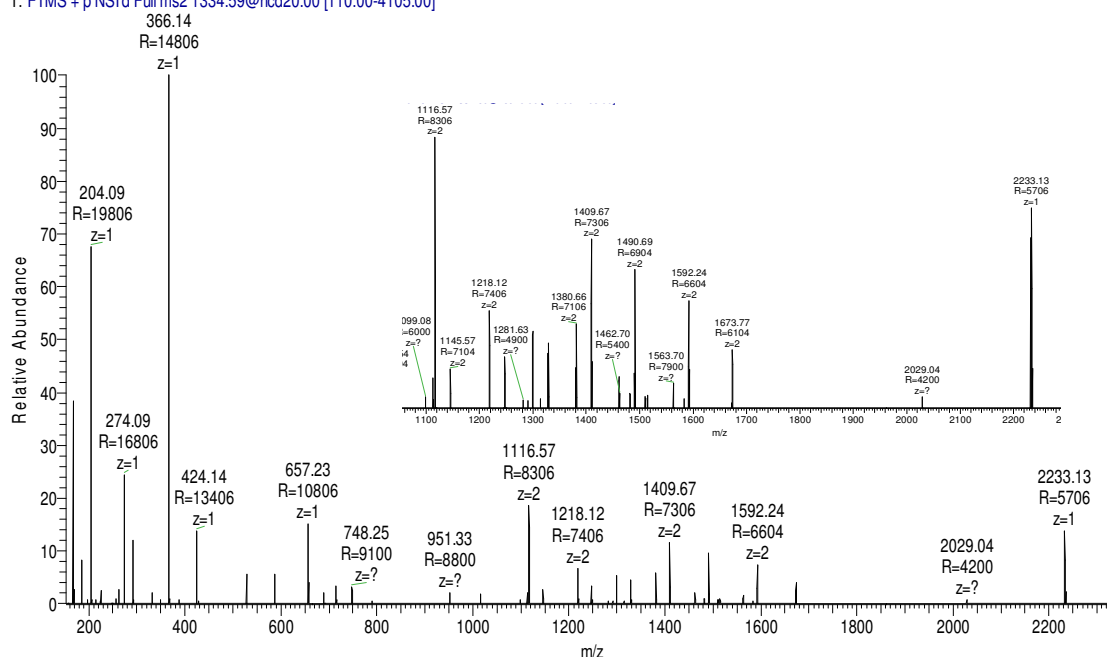
FIG. 5: Bottom: MS/MS spectrum of the triply charged molecular ions of the glycopeptides HexNAc$_5$Hex$_5$Fuc$_1$+ LNWTGPPEPLQKAVANFF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin . The experimental determined molecular mass of the glycopeptide is 3999.741 Da, which corresponds to the theoretical one 3999.763, with an error of 0.022 Da (6 ppm). Top: MassAI identification.
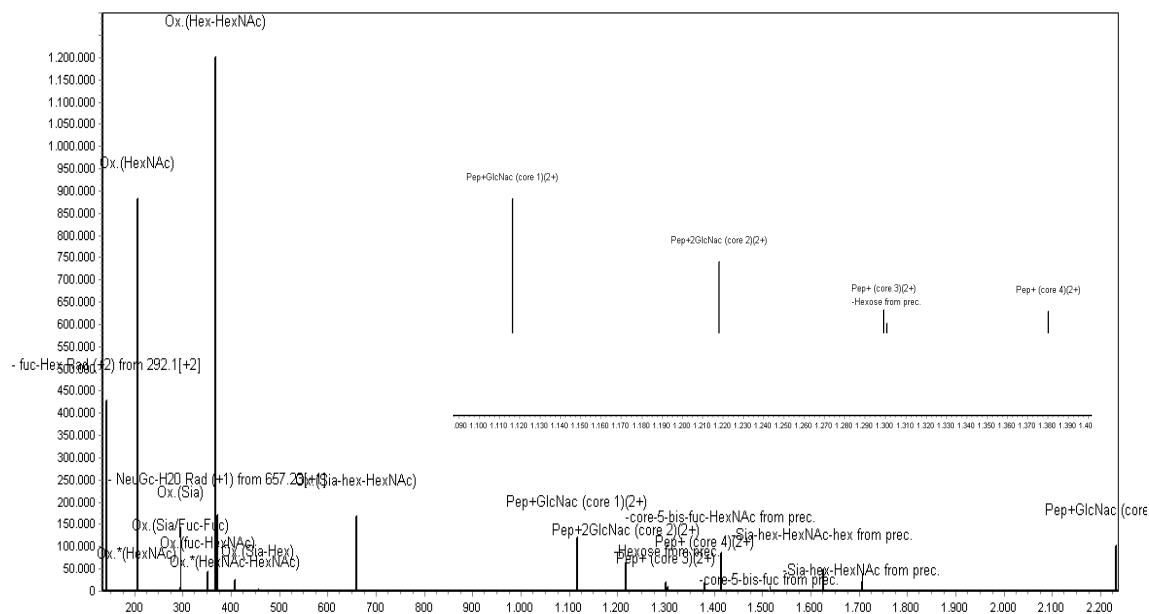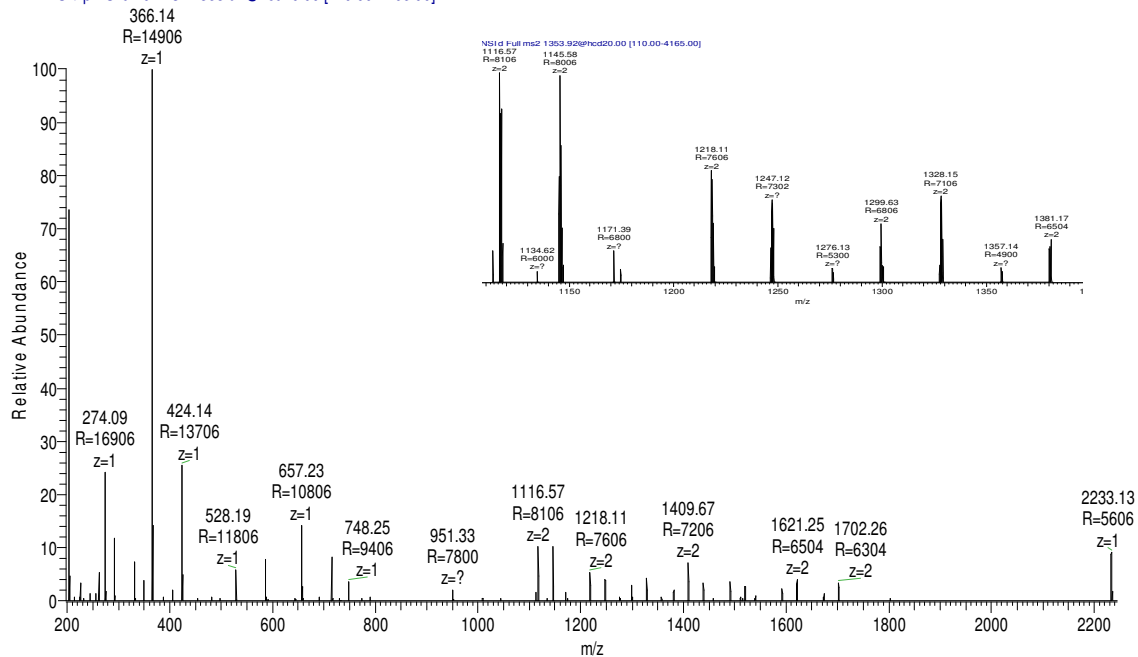
FIG. 6: Bottom: MS/MS spectrum of the triply molecular ions of the glycopeptide HexNAc$_6$Hex$_5$+LNW in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 4056.765 Da, which corresponds to the theoretical one 4056.787, with an error of 0.022 Da (5 ppm). Top: MassAI identification.

Prec (1041.938)2+

Ox.(HexNAc)

Pep+ (Core5)(2+)

Pep+GlcNac (core

Ox.(Hexose)

Ox.(Hex-Hex)

Ox.(Hex-HexNAc)

Pep+2GlcNac (core 2)(2+)

Pep+ (core 4)(2+)
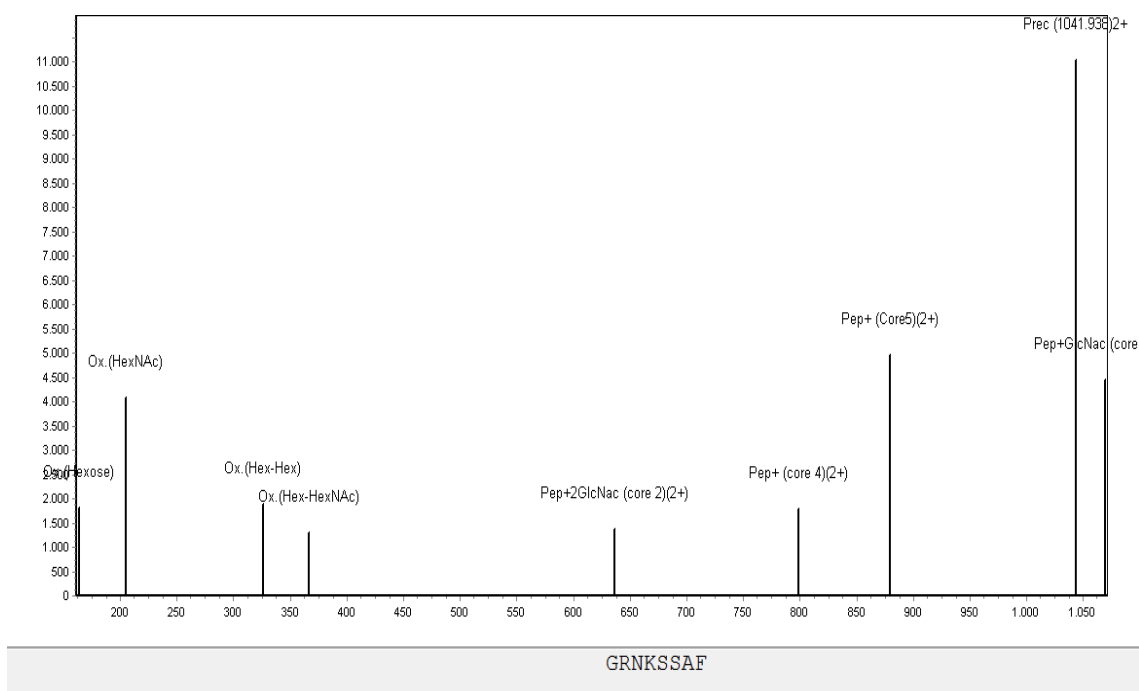
GRNKSSAF

T: FTMS + p NSI d Full ms2 1041.94@hcd20.00 [110.00-2145.00]

1041.938
R=8406
z=2

960.911
R=8306
z=2

879.888
R=8304
z=2

1069.524
R=7404
z=1

848.772
R=5900
z=?

798.864
R=6800
z=?

636.809
R=7600
z=?

204.087
R=16900
z=?

325.114
R=10900
z=?

224.530
R=12900
z=?

387.482
R=9500
z=?

Relative Abundance

m/z
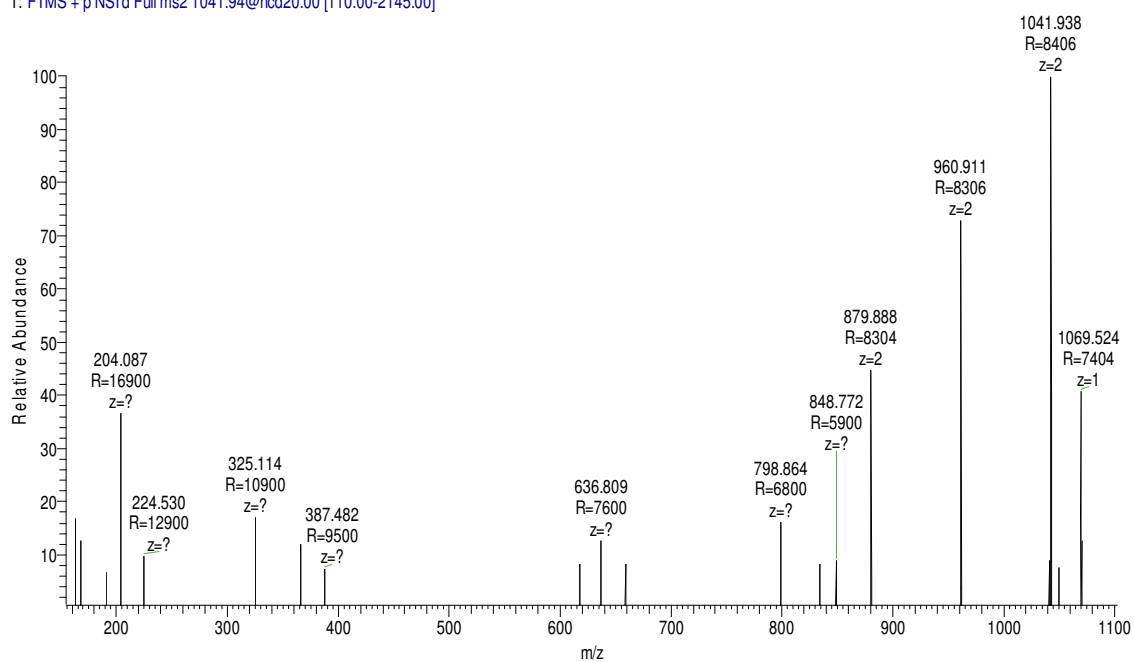
FIG. 7: Bottom: MS/MS spectrum of the doubly charged molecular ions of the glycopeptide HexNAc$_2$Hex$_5$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2081.866 Da, which corresponds to the theoretical one: 2081.864, with an error of 0.002 Da (1 ppm). Top: MassAI identification.
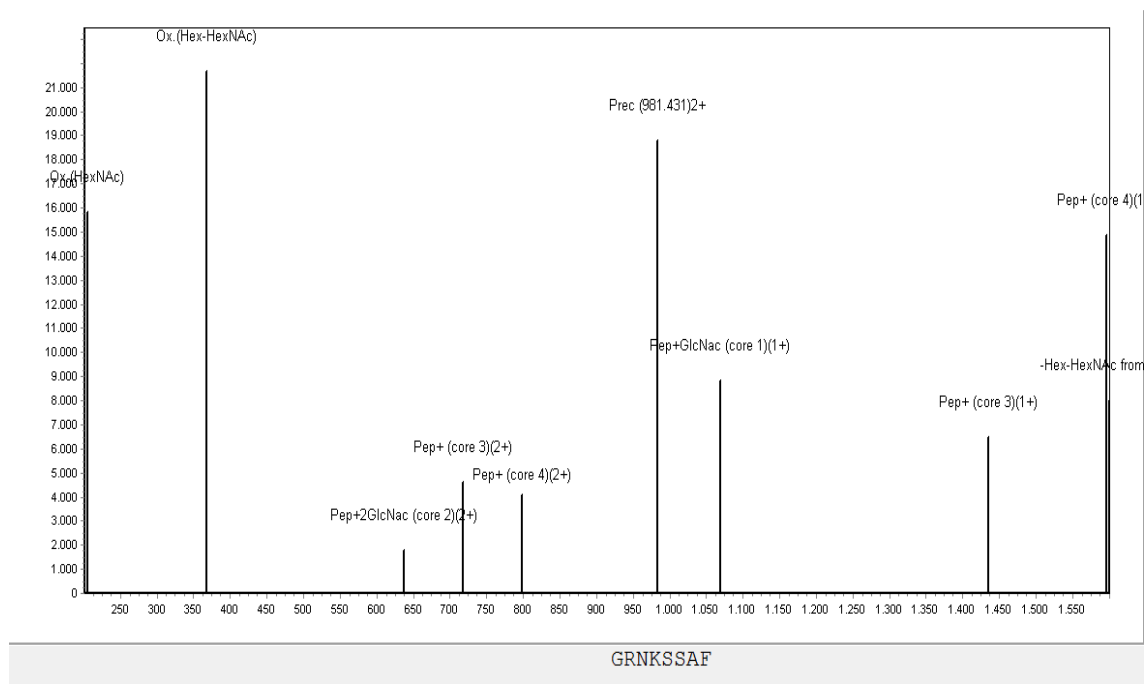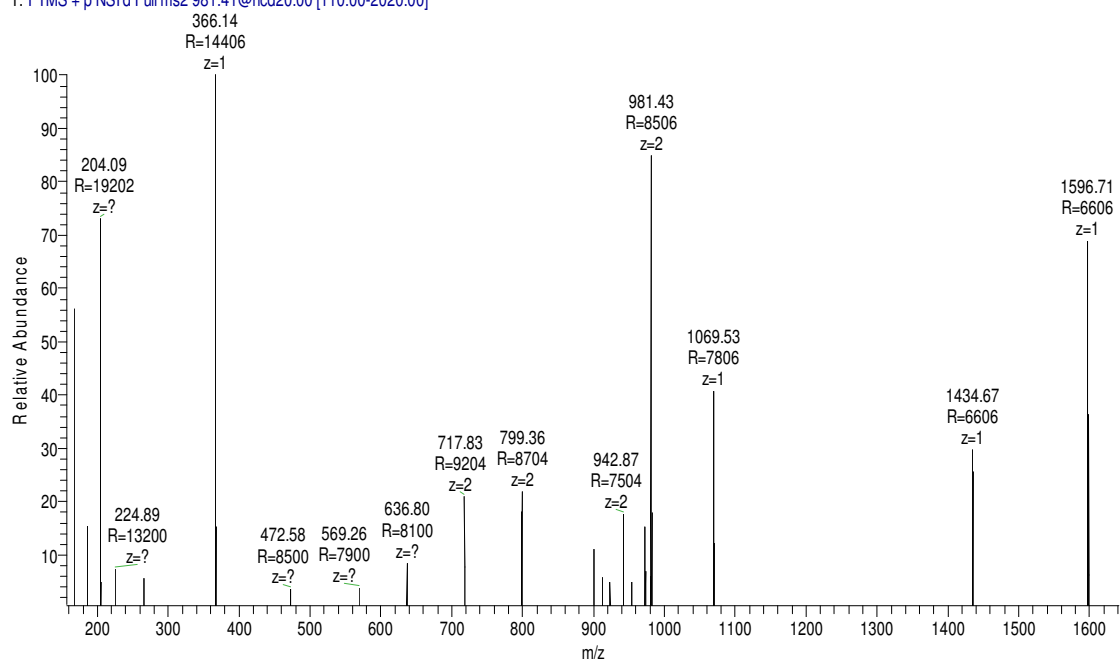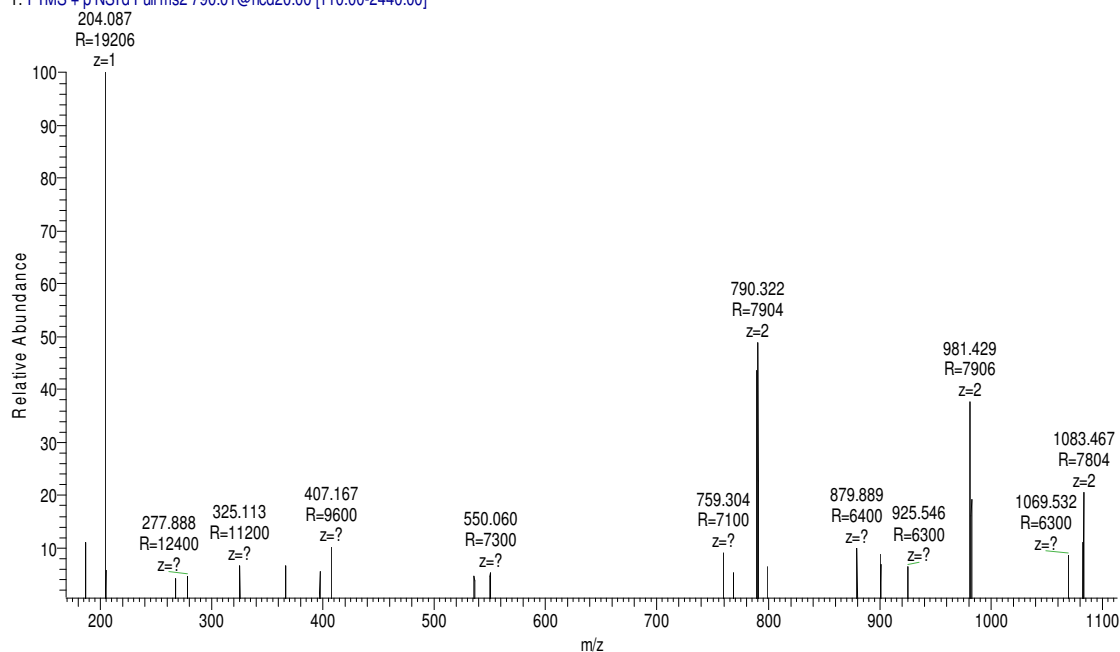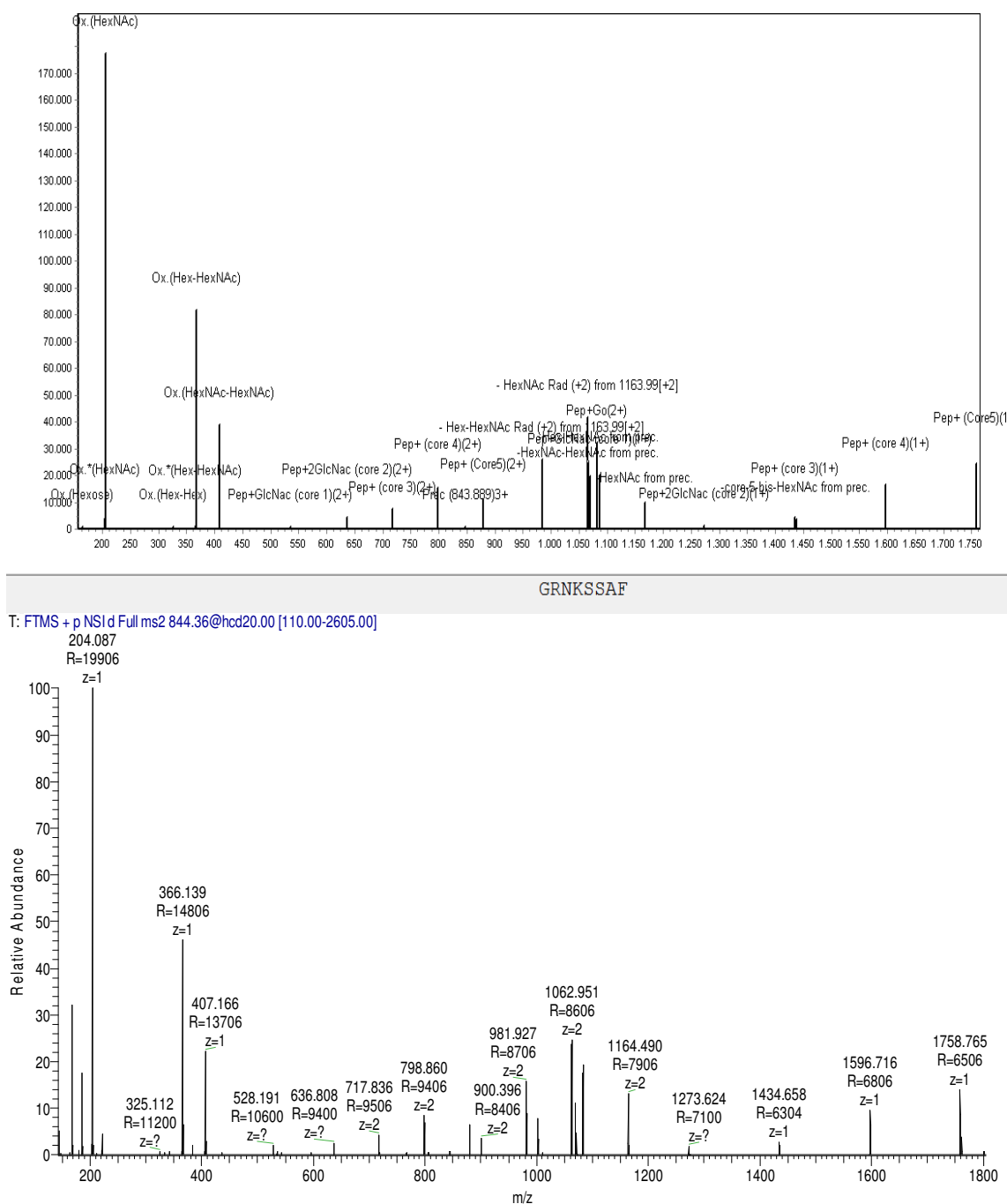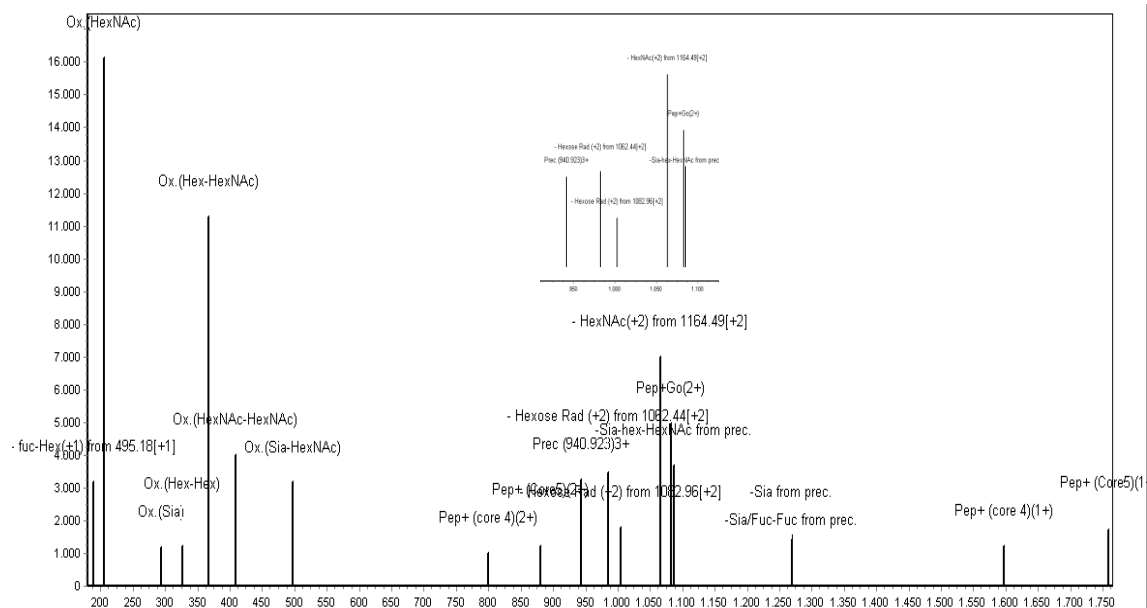
FIG. 8: Bottom: MS/MS spectrum of the doubly charged molecular ions of the glycopeptide HexNAc₃Hex₃+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 1960,805 Da, which corresponds to the theoretical one 1960,838, with an error of 0.032 Da (16 ppm). Top: MassAI identification.
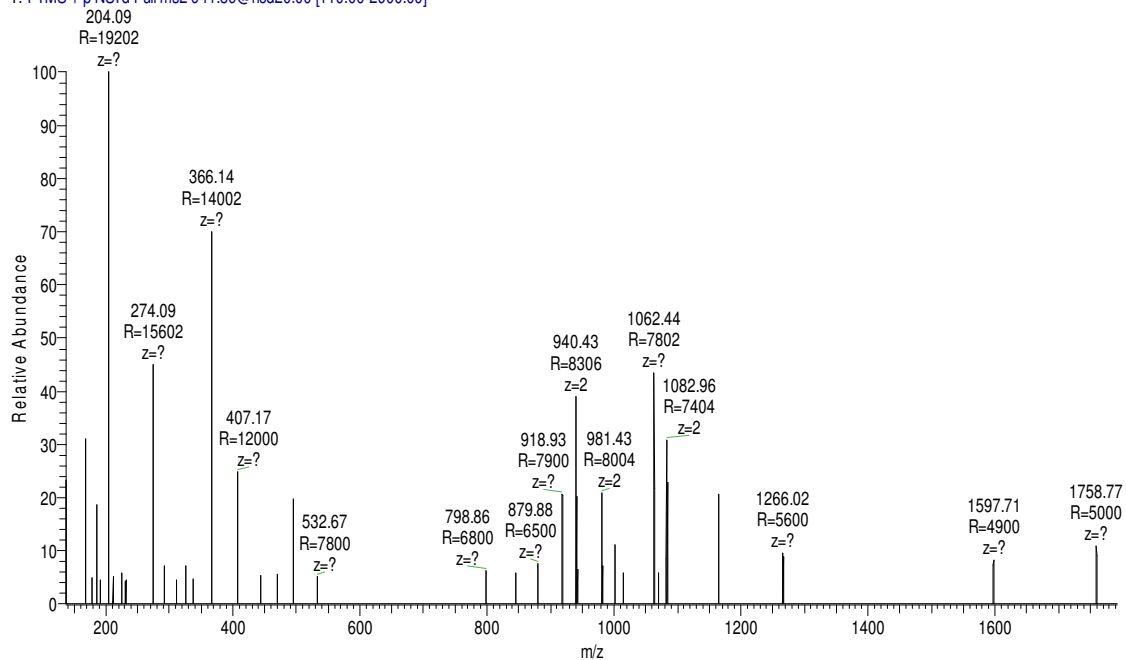
FIG. 9: Bottom: MS/MS spectrum of the triply charged molecular ions of the glycopeptide HexNAc$_3$Hex$_4$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2122.891 Da, which corresponds to the theoretical one 2122.891, with an error of 0.000 Da (0 ppm). Top: MassAI identification.
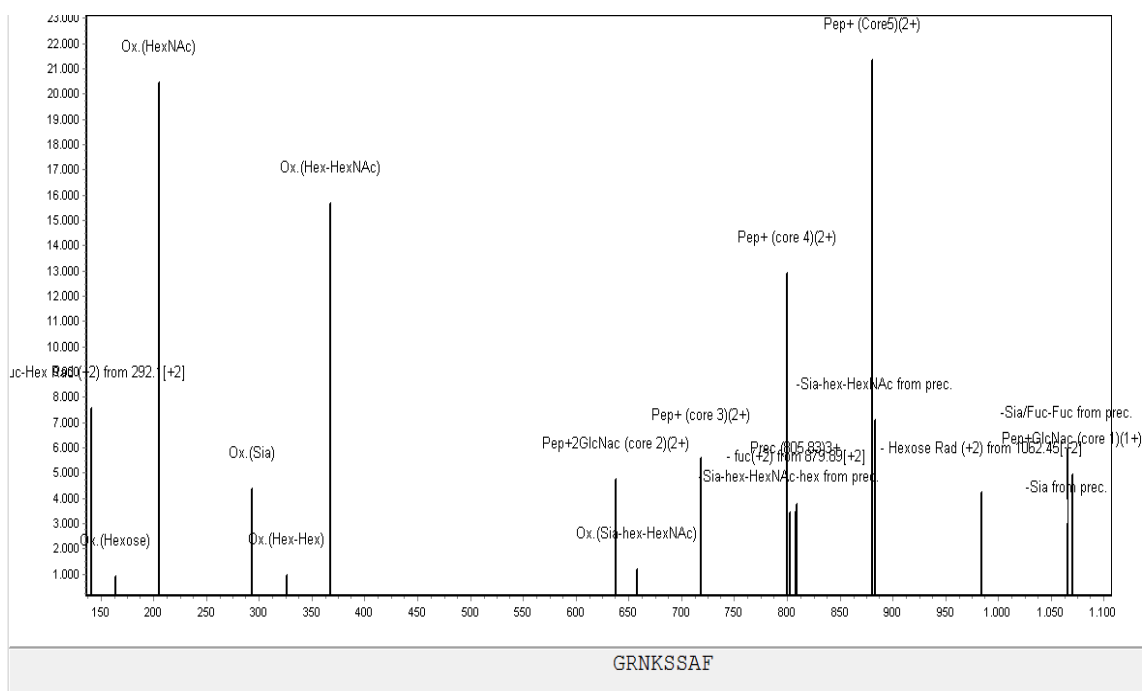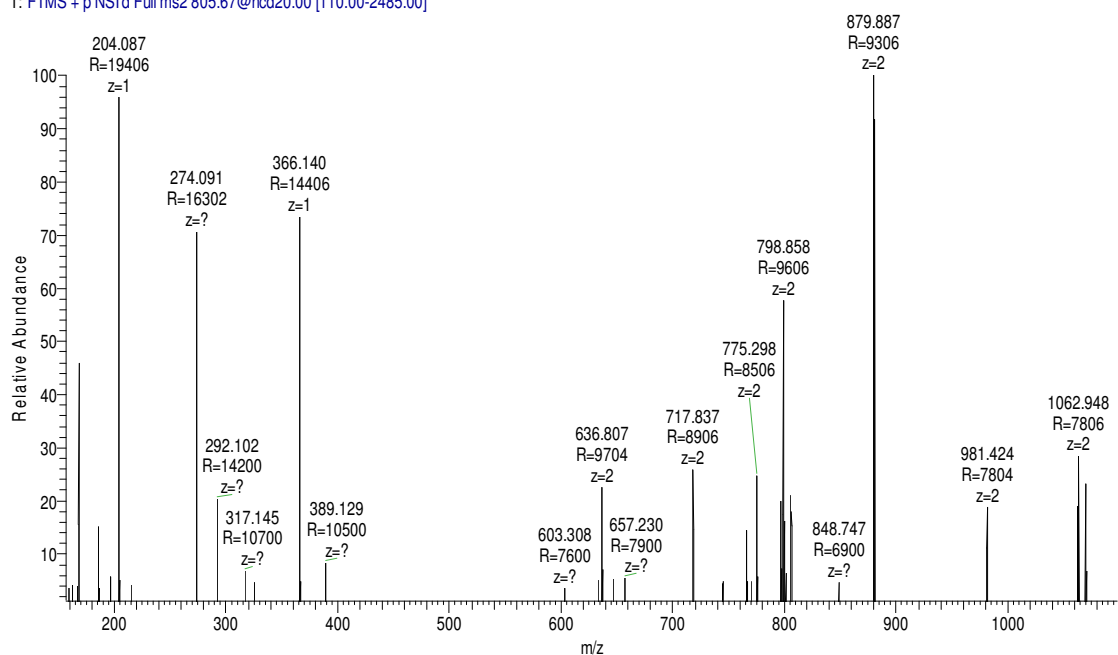
FIG. 10: Bottom MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_3$Hex$_4$NeuAc$_1$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2413.985 Da, which corresponds to the theoretical one 2413.986, with an error of 0.001 Da (0 ppm). **Top:** MassAI identification.
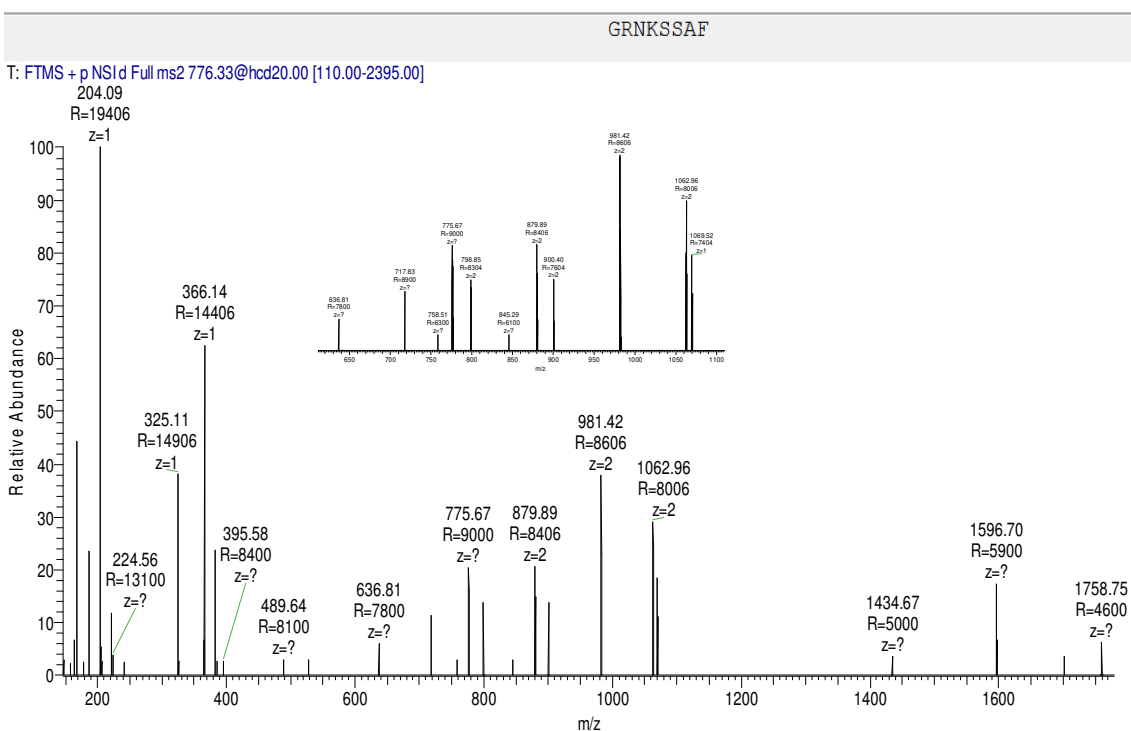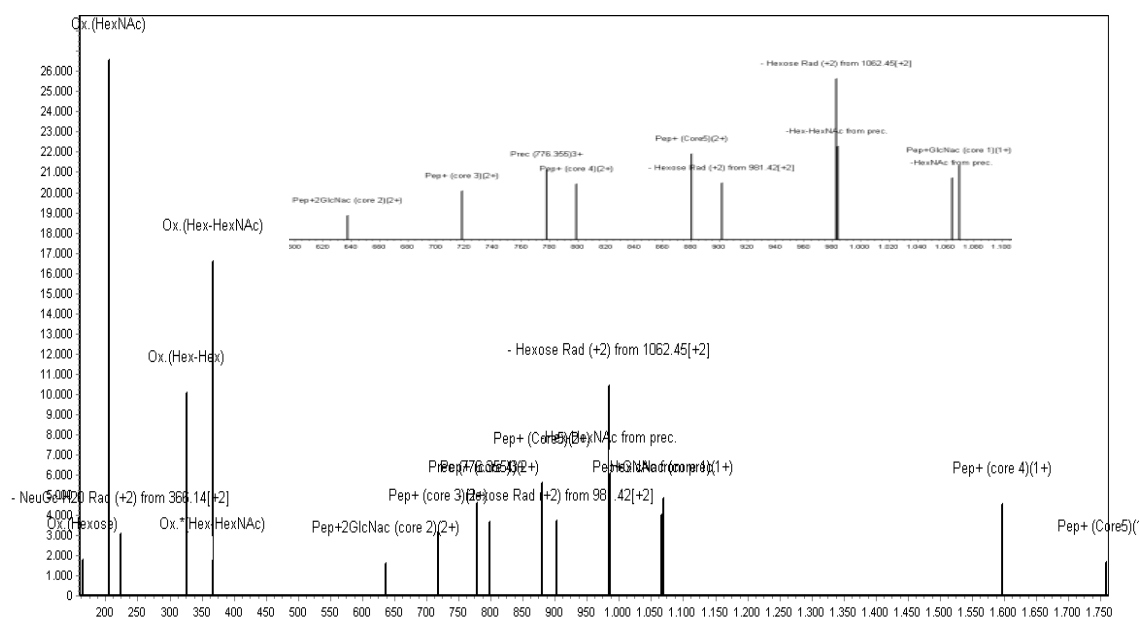
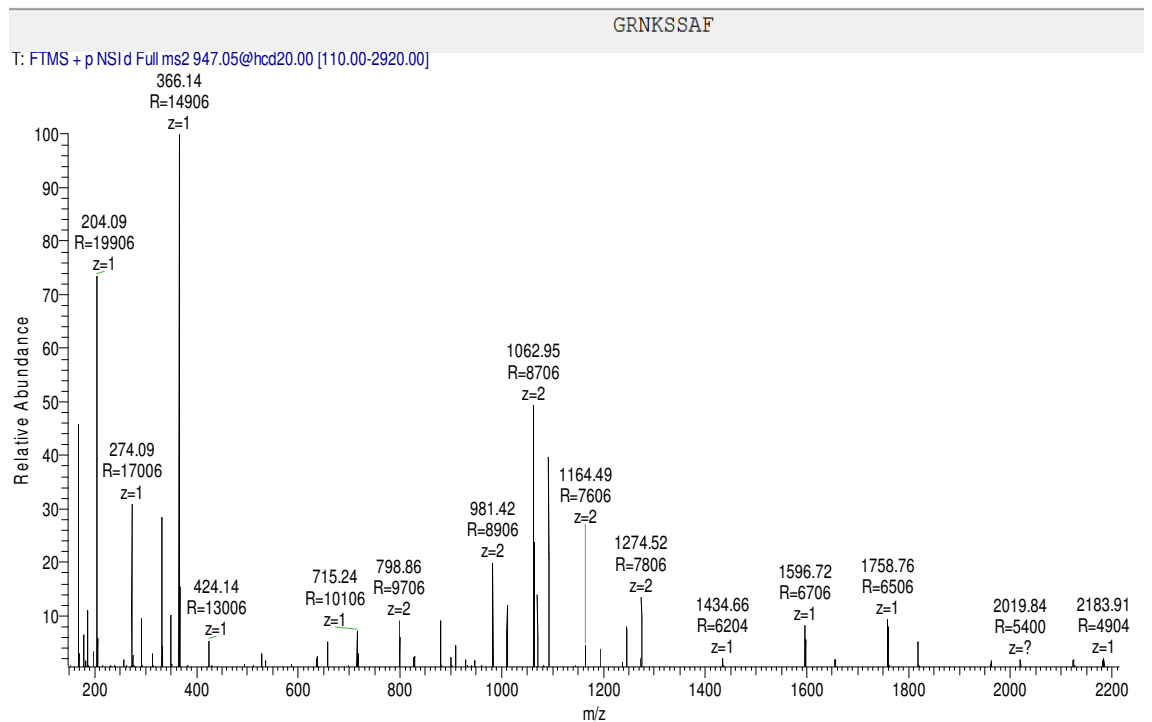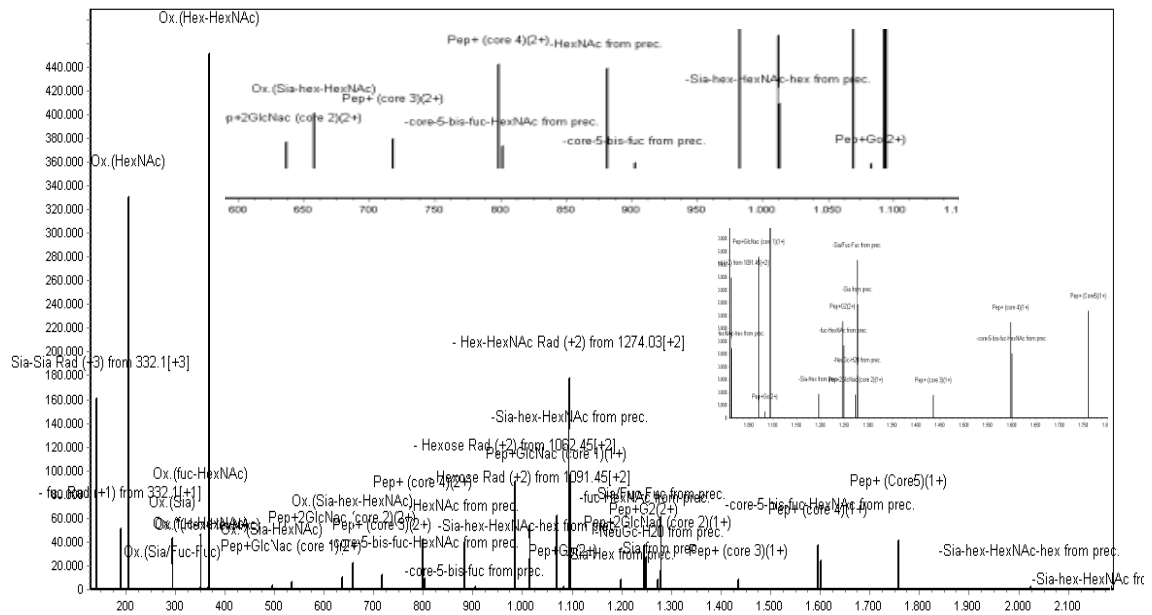**FIG. 11: Bottom MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₄Hex₄+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2325.969 Da, which corresponds to the theoretical one 2325.970, with an error of 0.001 Da (0 ppm). Top: MassAI identification.**
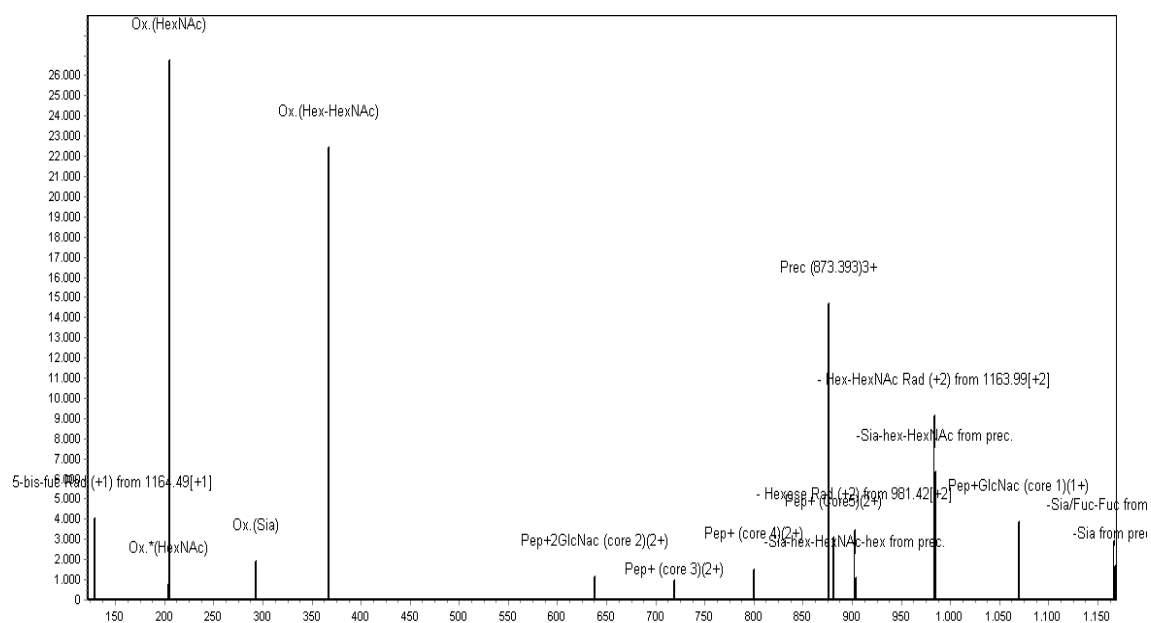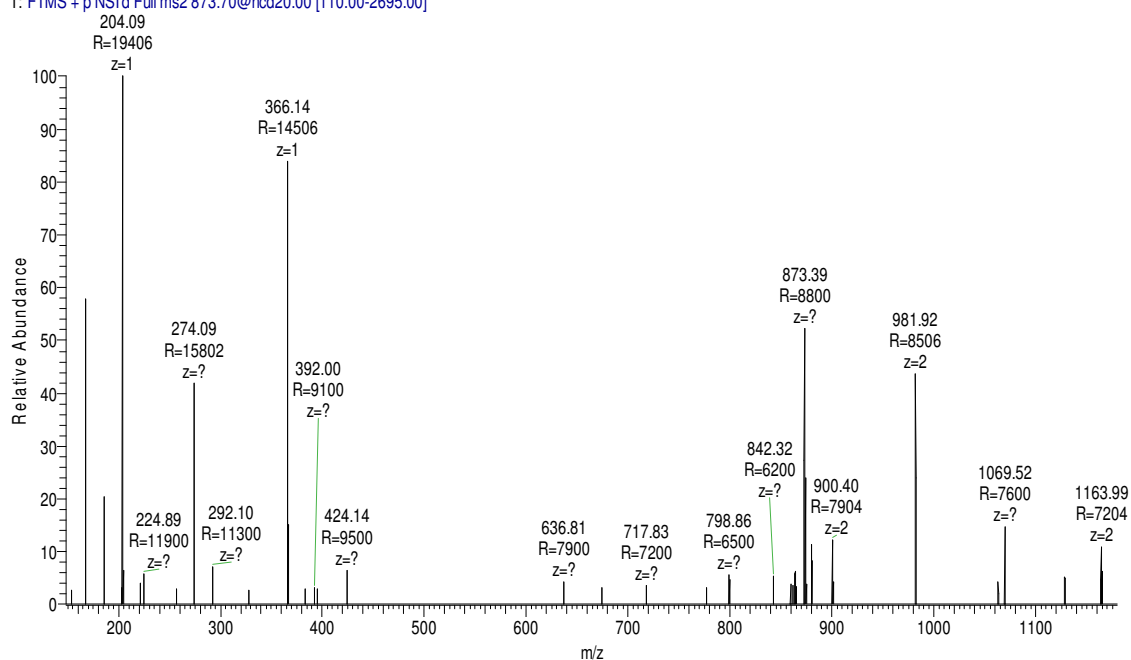
FIG. 12: Bottom MS/MS spectra of triply charged molecular ions of the glycopeptide HexNAc$_4$Hex$_4$NeuAc$_1$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2617.053 Da, which corresponds to the theoretical one 2617.065, with an error of 0.012 Da (5 ppm). **Top:** MassAI identification.
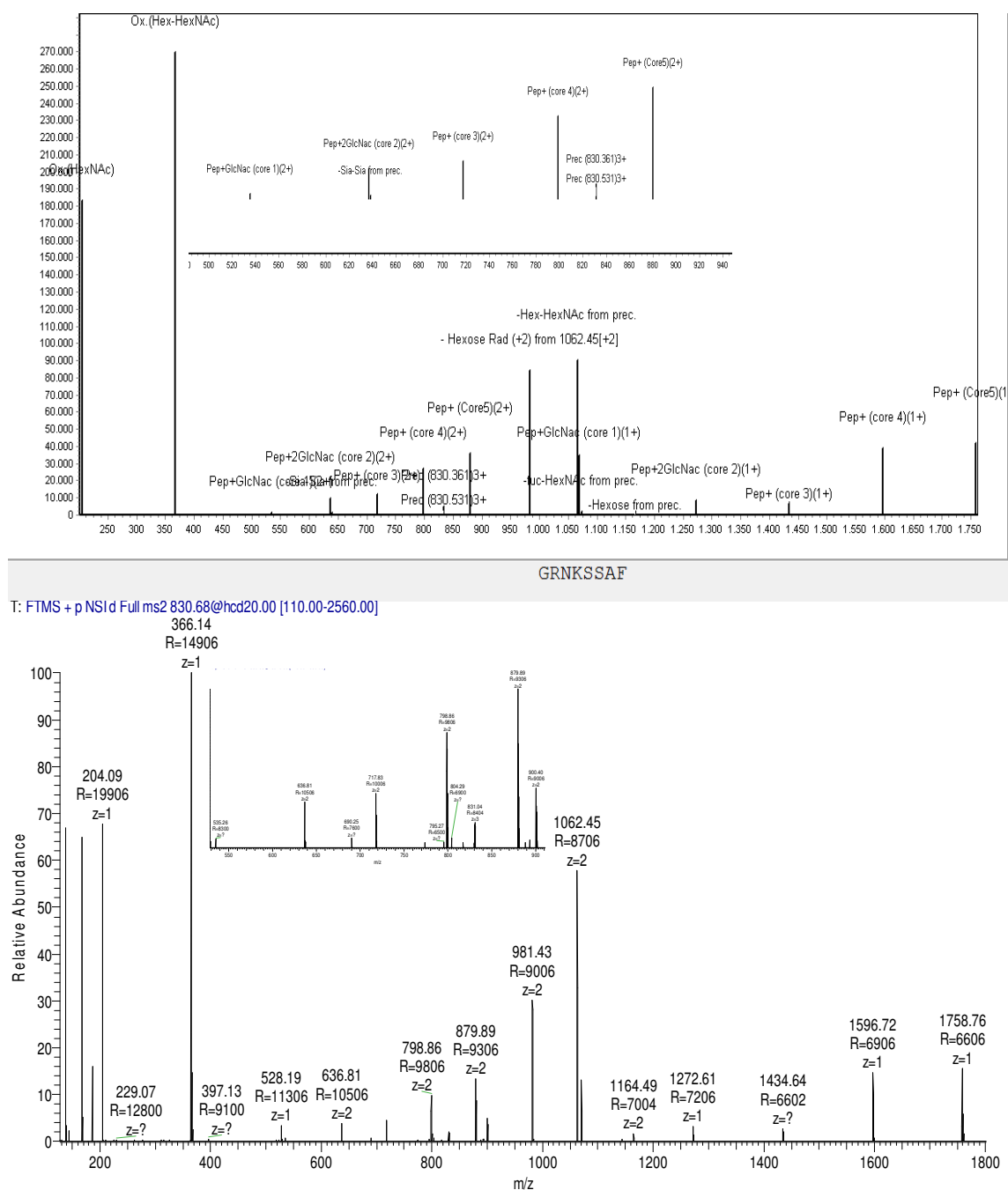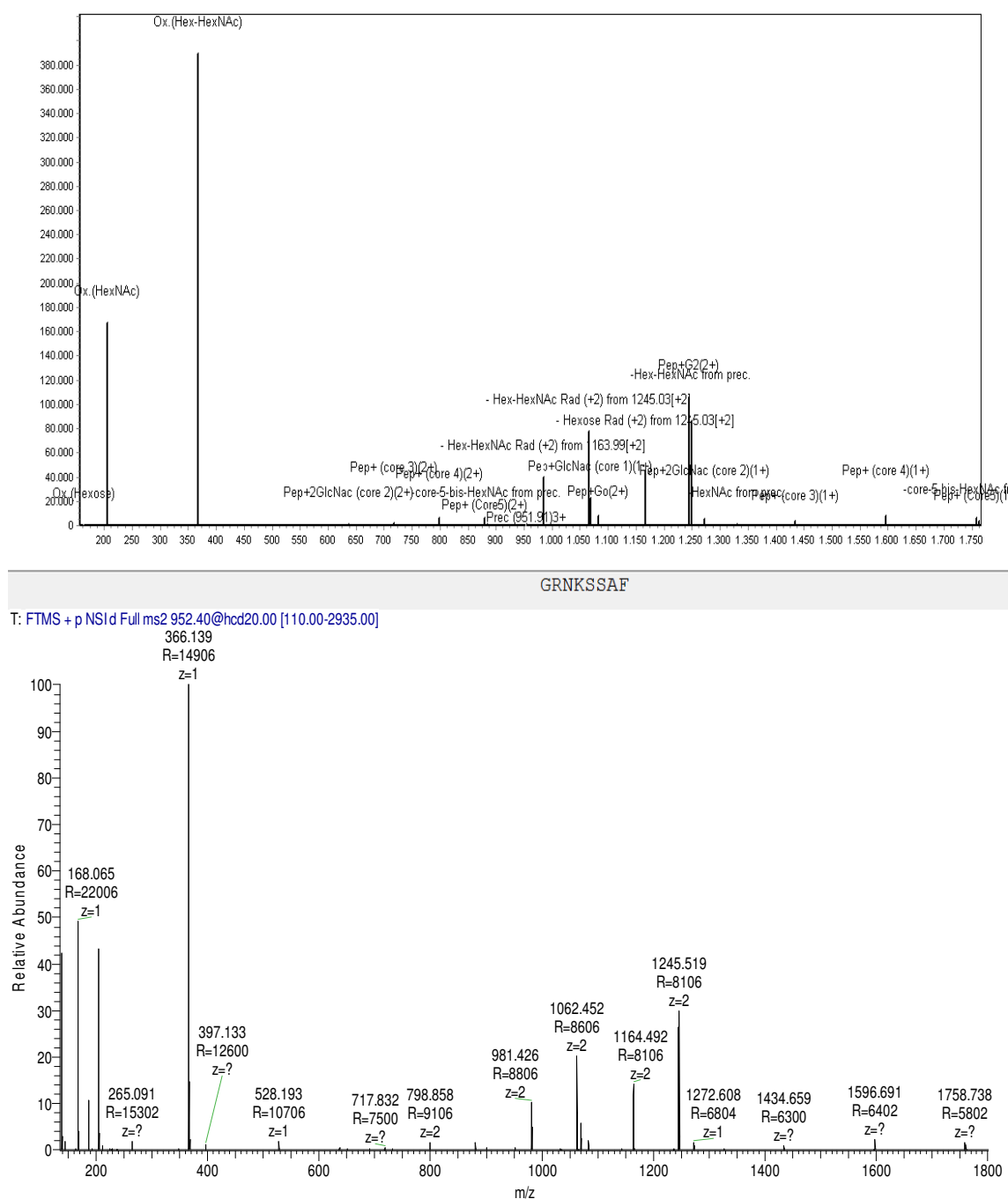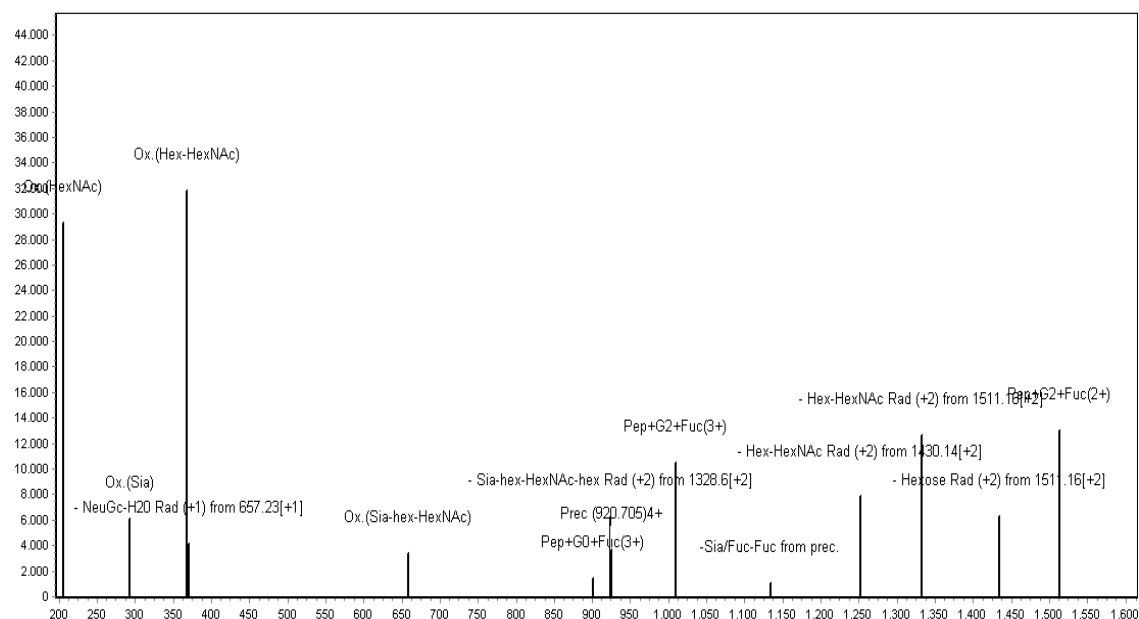
FIG. 13: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₄Hex₅+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2488.023 Da, which corresponds to the theoretical one 2488.023, with an error of 0.000 Da (0 ppm). Top: MassAI identification.
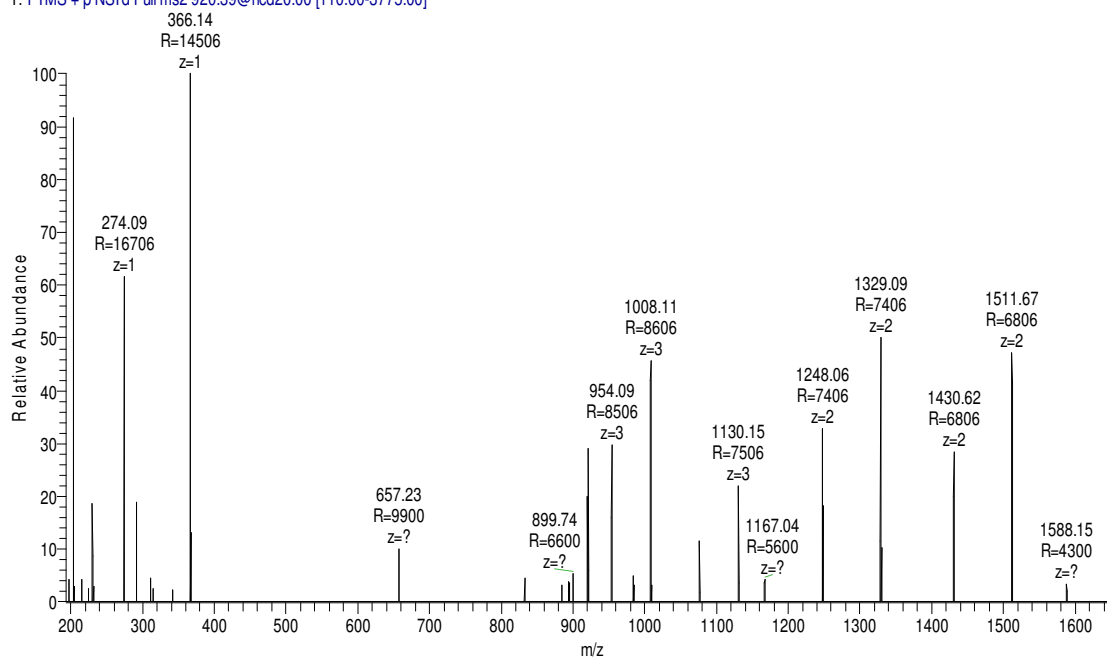
FIG. 14: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_4$Hex$_5$NeuAc$_1$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2779.116 Da, which corresponds to the theoretical one 2779.118, with an error of 0.002 Da (1 ppm). Top: MassAI identification.
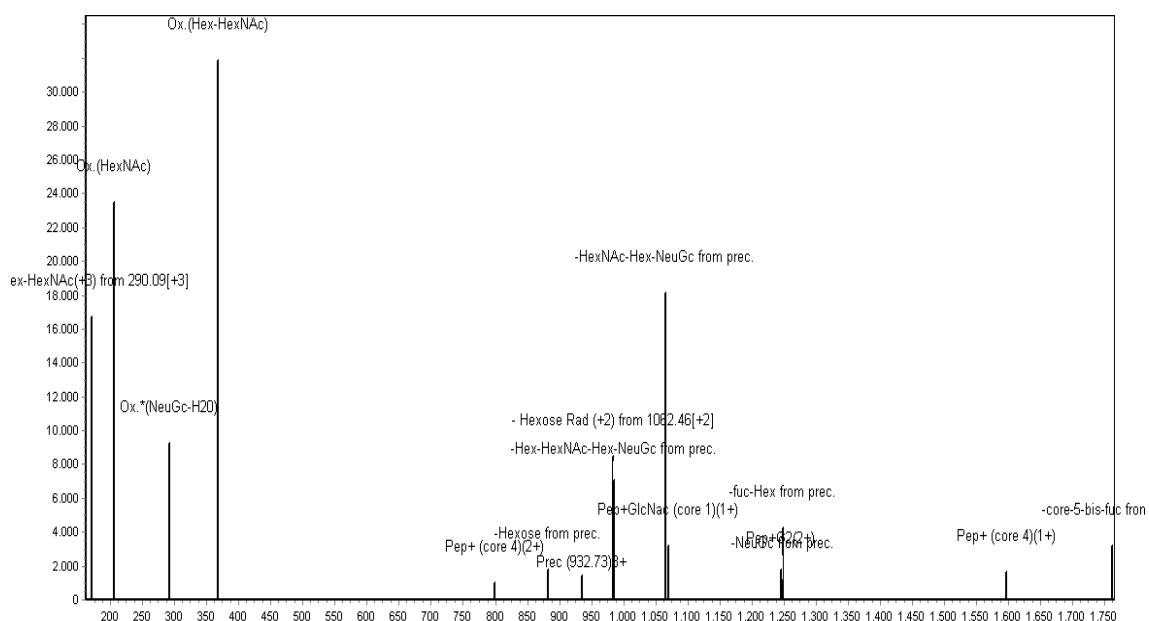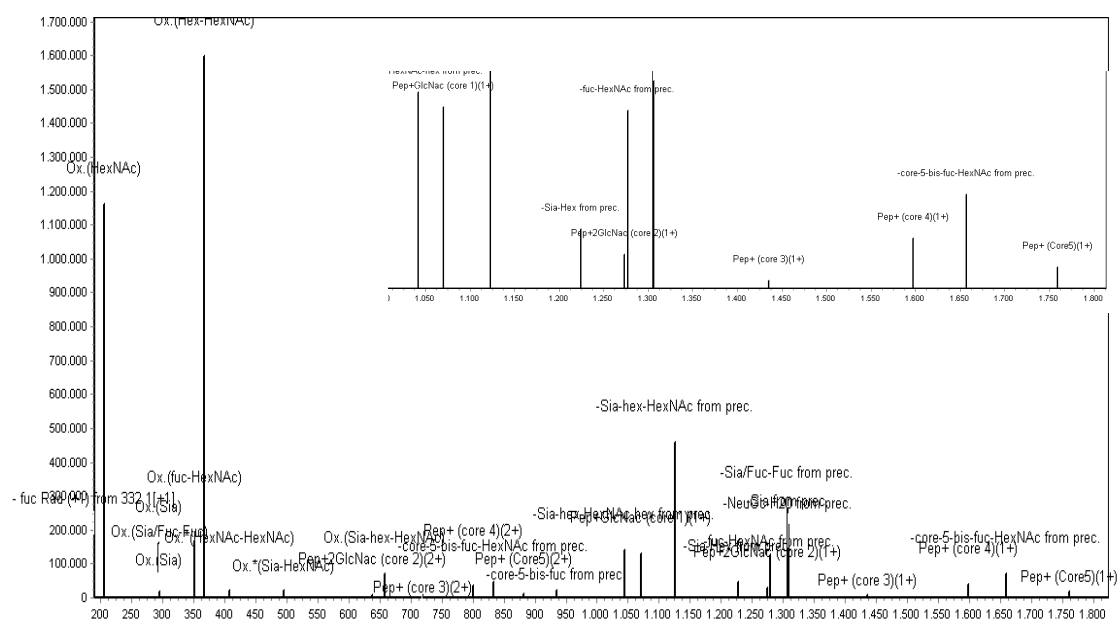
FIG. 15: Bottom: MS/MS spectrum of the triply charged molecular ions of the glycopeptide HexNAc$_4$Hex$_5$NeuGc$_1$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2795.111 Da, which corresponds to the theoretical one 2795.113, with an error of 0.002 Da (1 ppm). Top: MassAI identification.
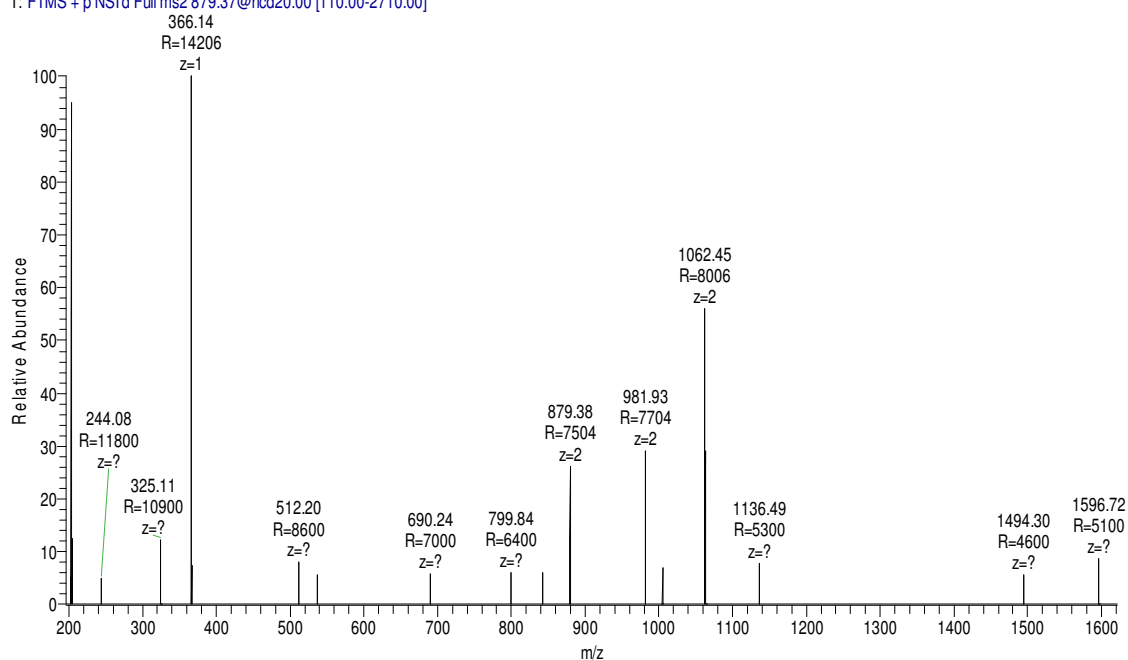
FIG. 16: Bottom: MS/MS spectrum of the triply charged selected molecular ions in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2634.088 Da, which corresponds to the theoretical one: 2634.081, with an error of 0.007 Da (3 ppm). Top: MassAI identification.
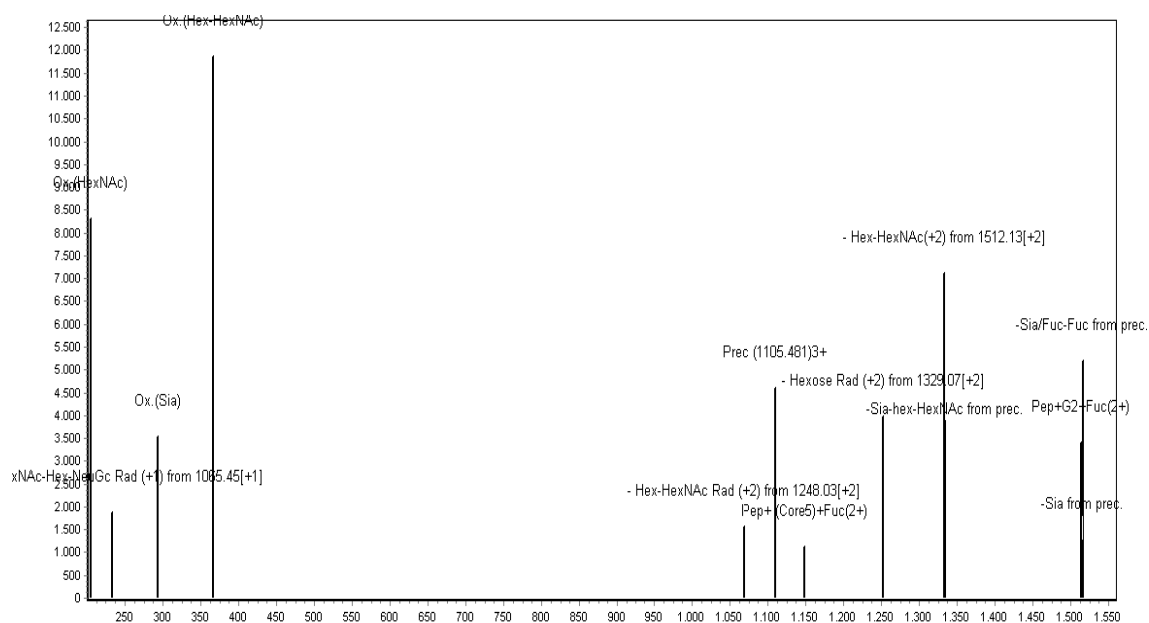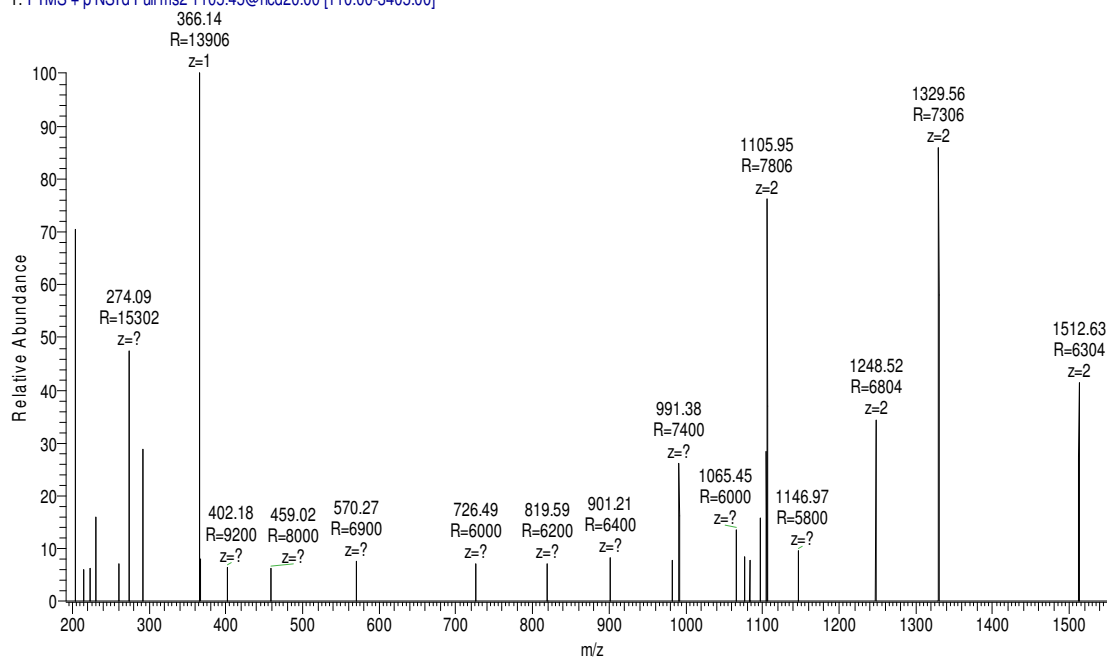
FIG. 17: Bottom MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_4$Hex$_5$Fuc$_1$NeuAc$_1$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3313.341 Da, which corresponds to the theoretical one 3313.387, with an error of 0.046 Da (14 ppm). Top: MassAI identification.
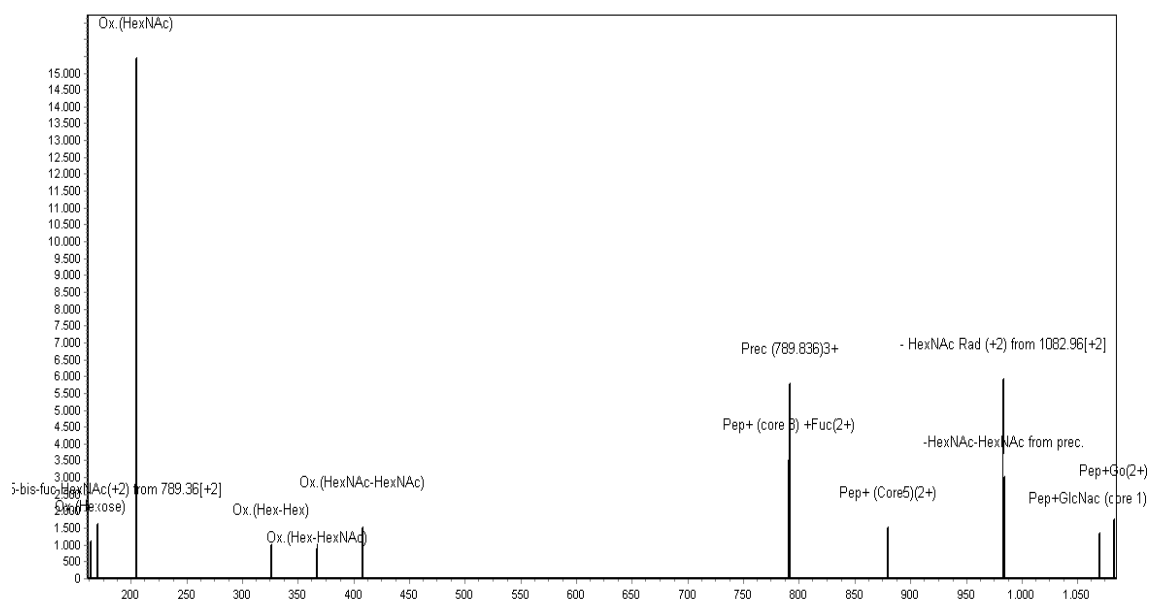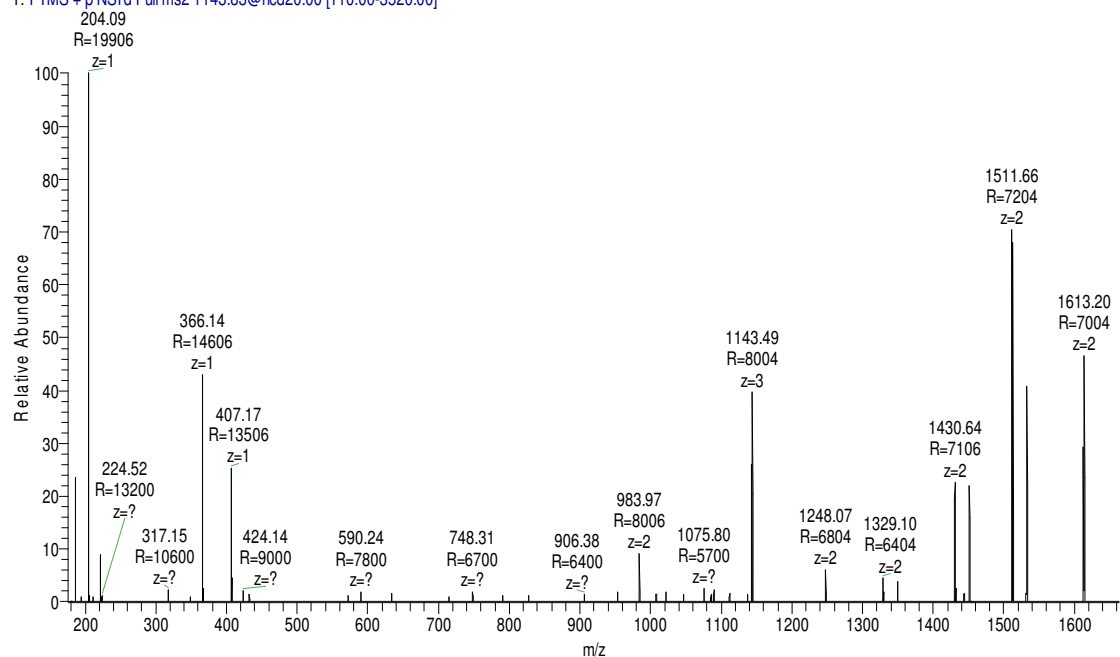
FIG. 18: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₅Hex₃+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2366.995 Da, which corresponds to the theoretical one 2366.996, with an error of 0.001 Da (0 ppm). Top: MassAI identification.
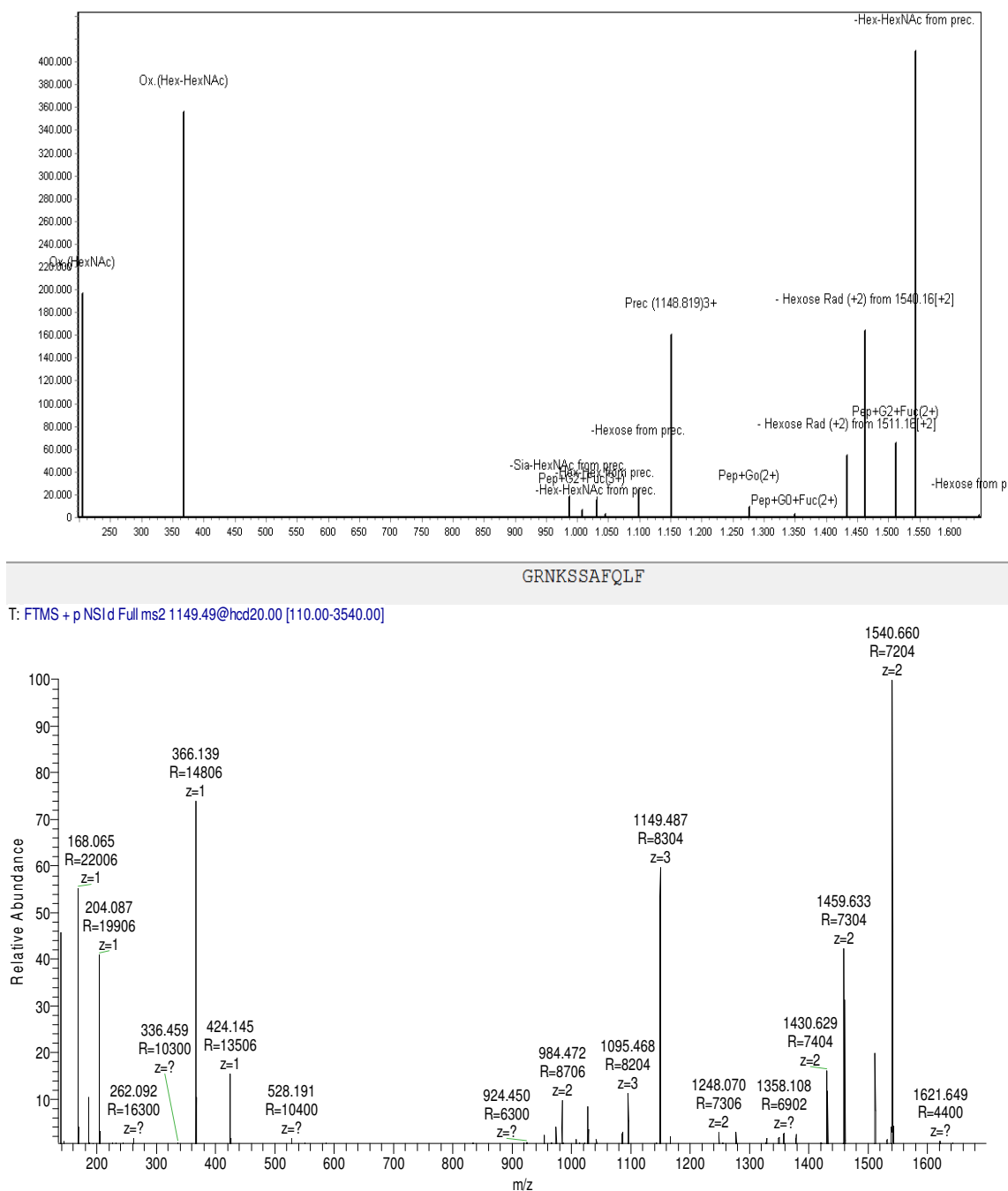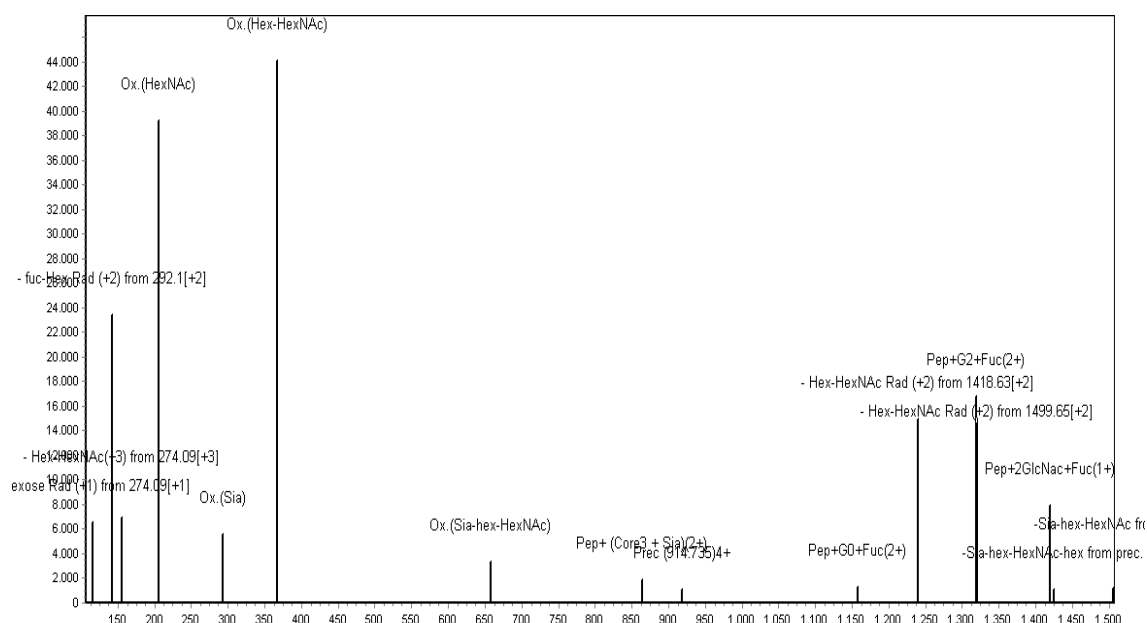
FIG. 19: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₅Hex₄+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2529.049 Da, which corresponds to the theoretical one 2529.049, with an error of 0.000 Da (0 ppm). Top: MassAI identification.

FIG. 20: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_5$Hex$_4$NeuAc$_1$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2820.145 Da, which corresponds to the theoretical one 2820.145, with an error of 0.000 Da (0 ppm). Top: MassAI identification.
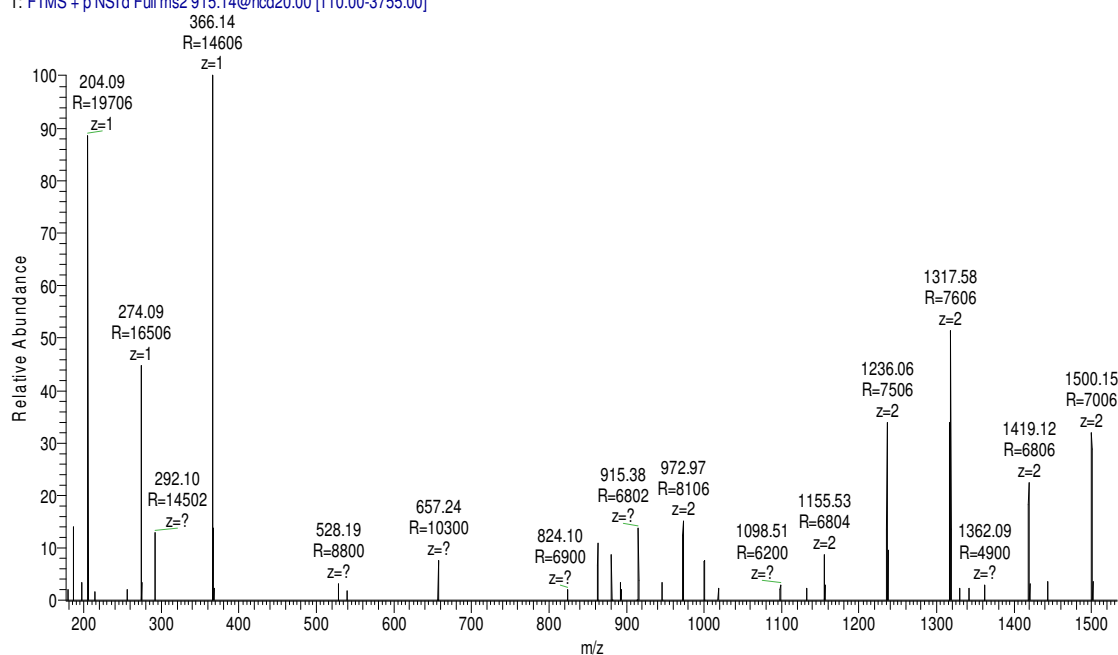
FIG. 21: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₅Hex₅Fuc₁+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2837.126 Da, which corresponds to the theoretical one 2837.160, with an error of 0.036 Da (13 ppm). Top: MassAI identification.

FIG. 22: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_5$Hex$_6$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2853.161 Da, which corresponds to the theoretical one 2853.155, with an error of 0.006 Da (1 ppm). Top: MassAI identification.
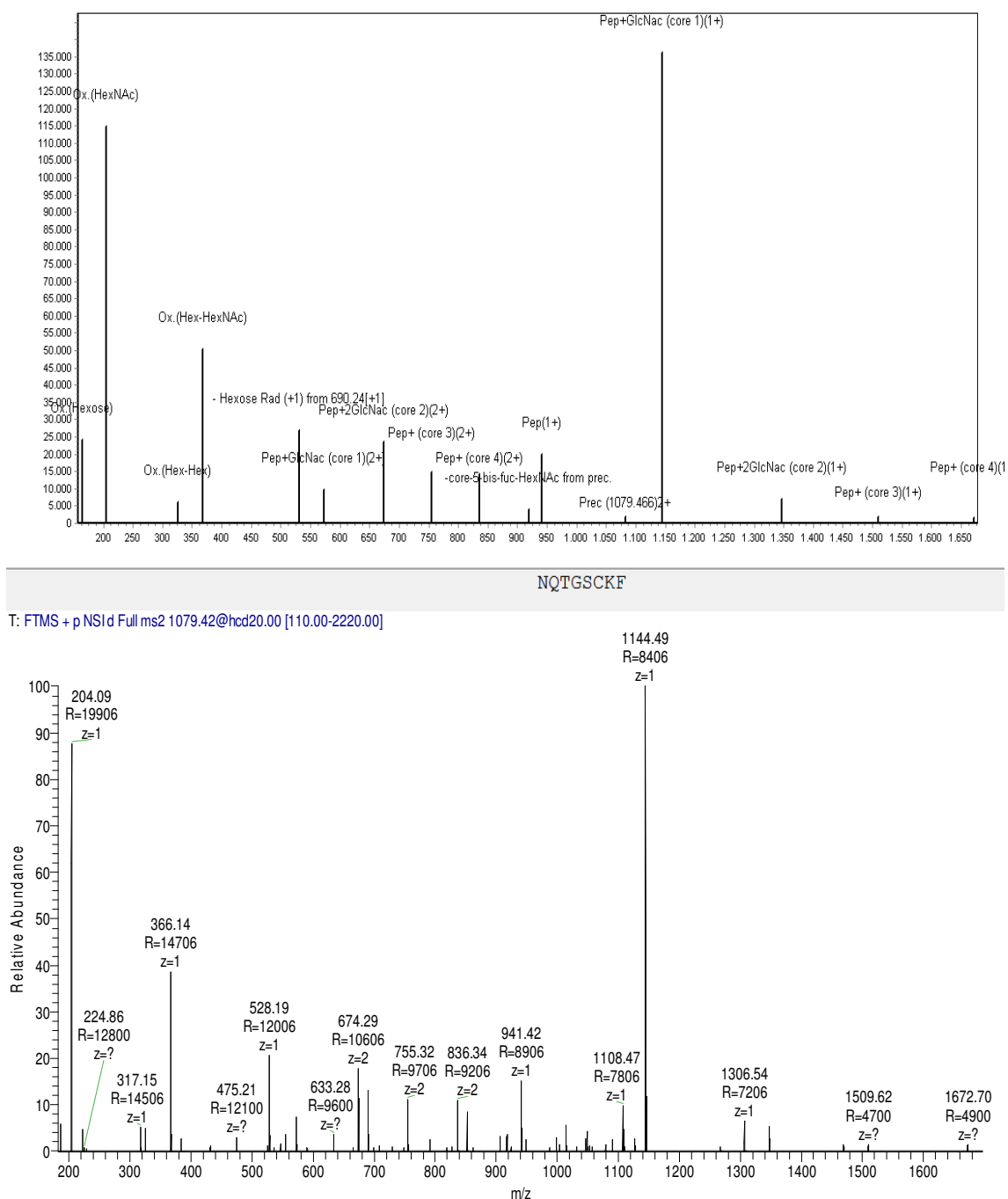
FIG. 23: Bottom: MS/MS spectrum of quadruply charged molecular ions of the glycopeptide HexNAc$_5$Hex$_6$Fuc$_1$NeuAc$_1$+GRNKSSAFQLF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3678.529 Da, which corresponds to the theoretical one 3678.519, with an error of , 0.010 Da (3 ppm). Top: MassAI identification.
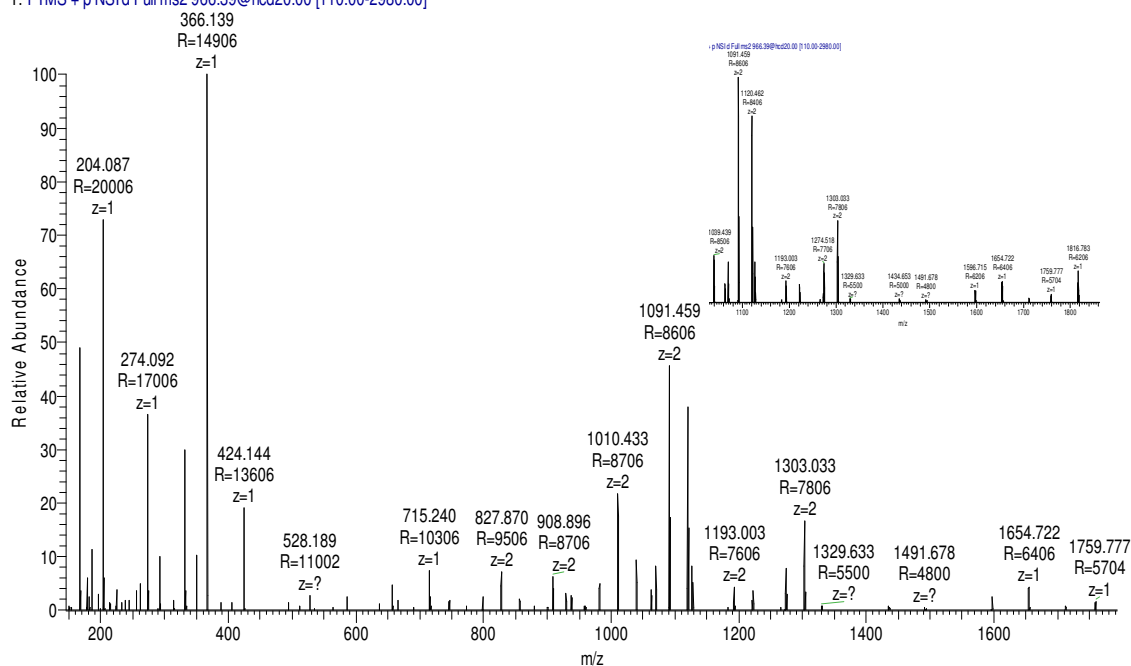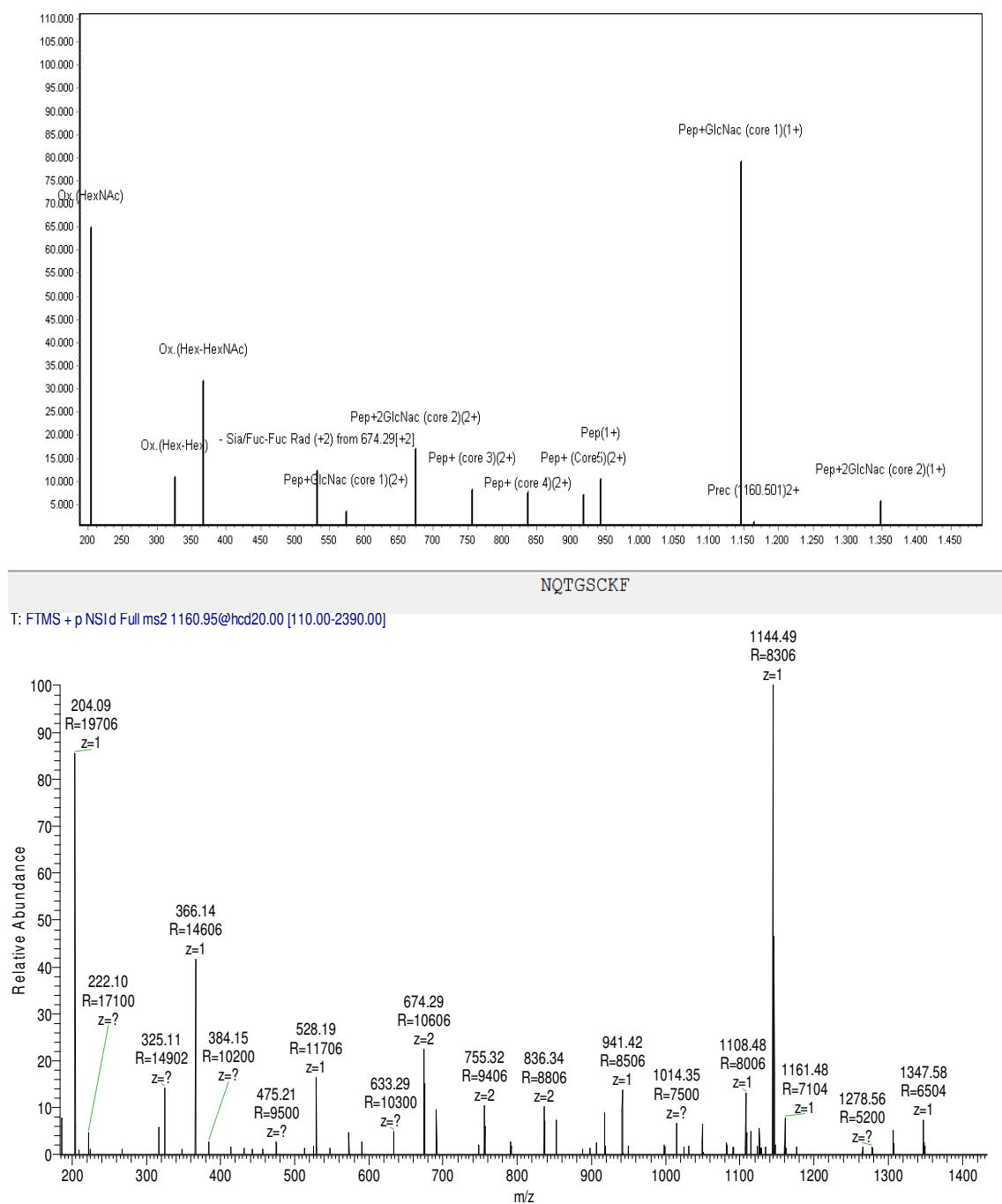
FIG. 24: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_6$Hex$_5$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2894.143 Da, which corresponds to the theoretical one 2894.181, with an error of 0.038 Da (13 ppm). Top: MassAI identification.
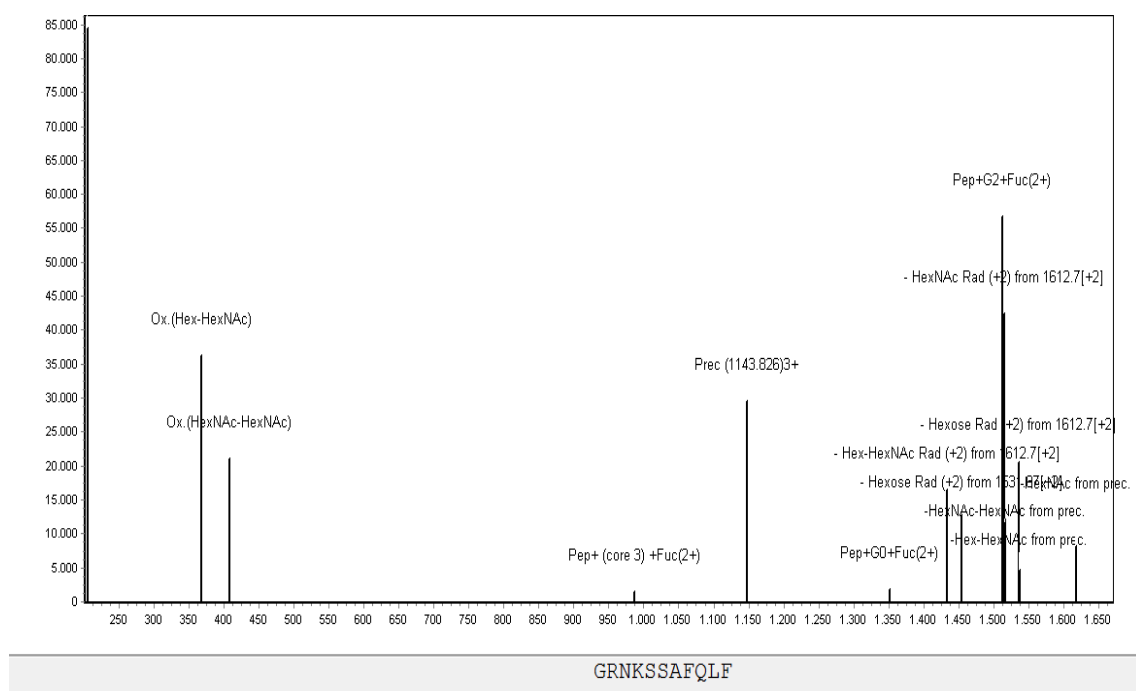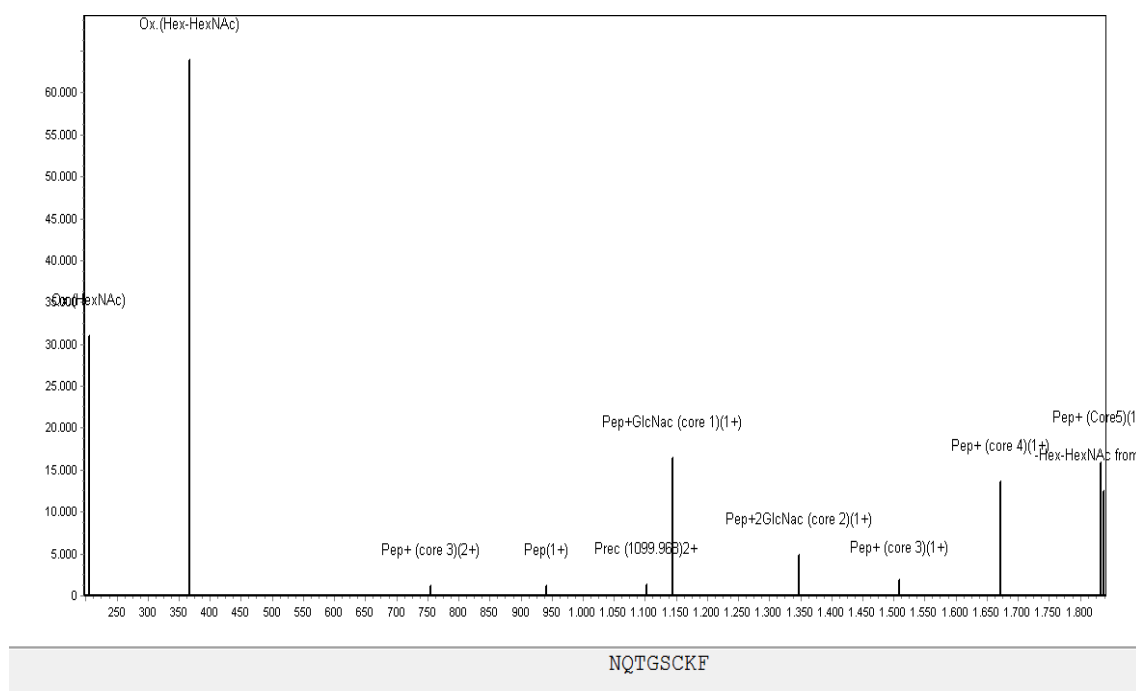
FIG. 25: Bottom: MS/MS spectra of triply charged molecular ions of the glycopeptide HexNAc₆Hex₅Fuc₁+GRNKSSAFQLF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3428.465 Da, which corresponds to the theoretical one 3428.450, with an error of 0.015 Da (4 ppm). Top: MassAI identification.
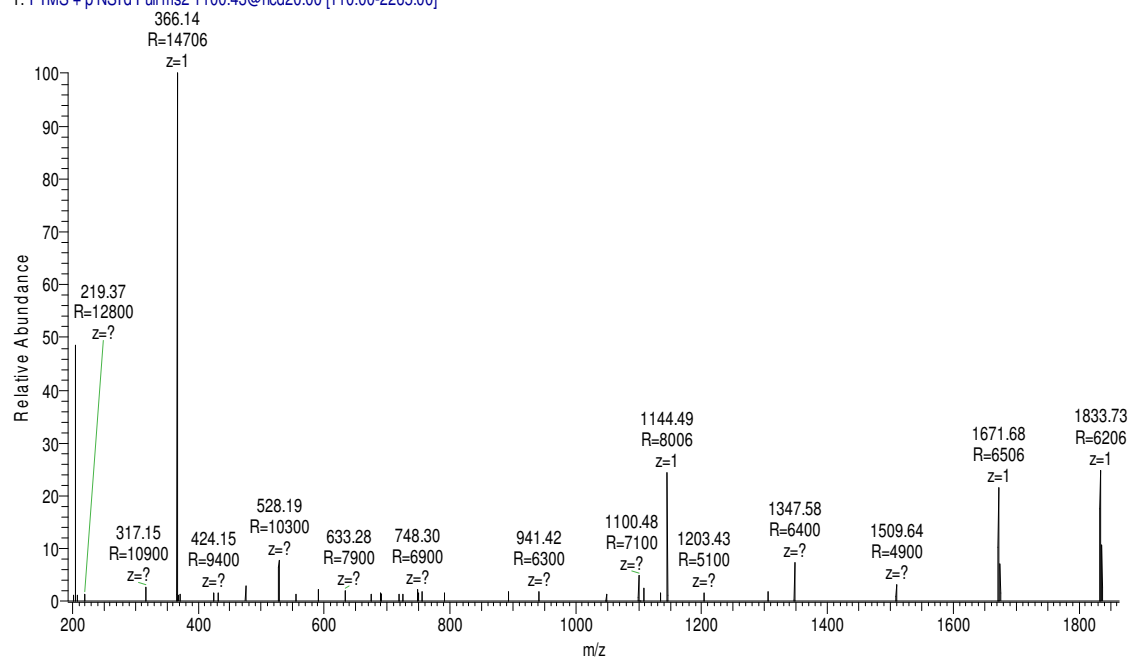
FIG. 26: Bottom:MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_6$Hex$_6$+GRNKSSAFQLF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3444,442 Da, which corresponds to the theoretical one 3444.445, with an error of 0.003 Da (1 ppm). Top: MassAI identification.

FIG. 27: Bottom: MS/MS spectrum of quadruply molecular ions of the glycopeptide HexNAc$_6$Hex$_7$Fuc$_1$NeuAc$_1$+GRNKSSAF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3655.520Da, which corresponds to the theoretical one: 3655.440, with an error of 0.080 Da (23 ppm). Top: MassAI identification.
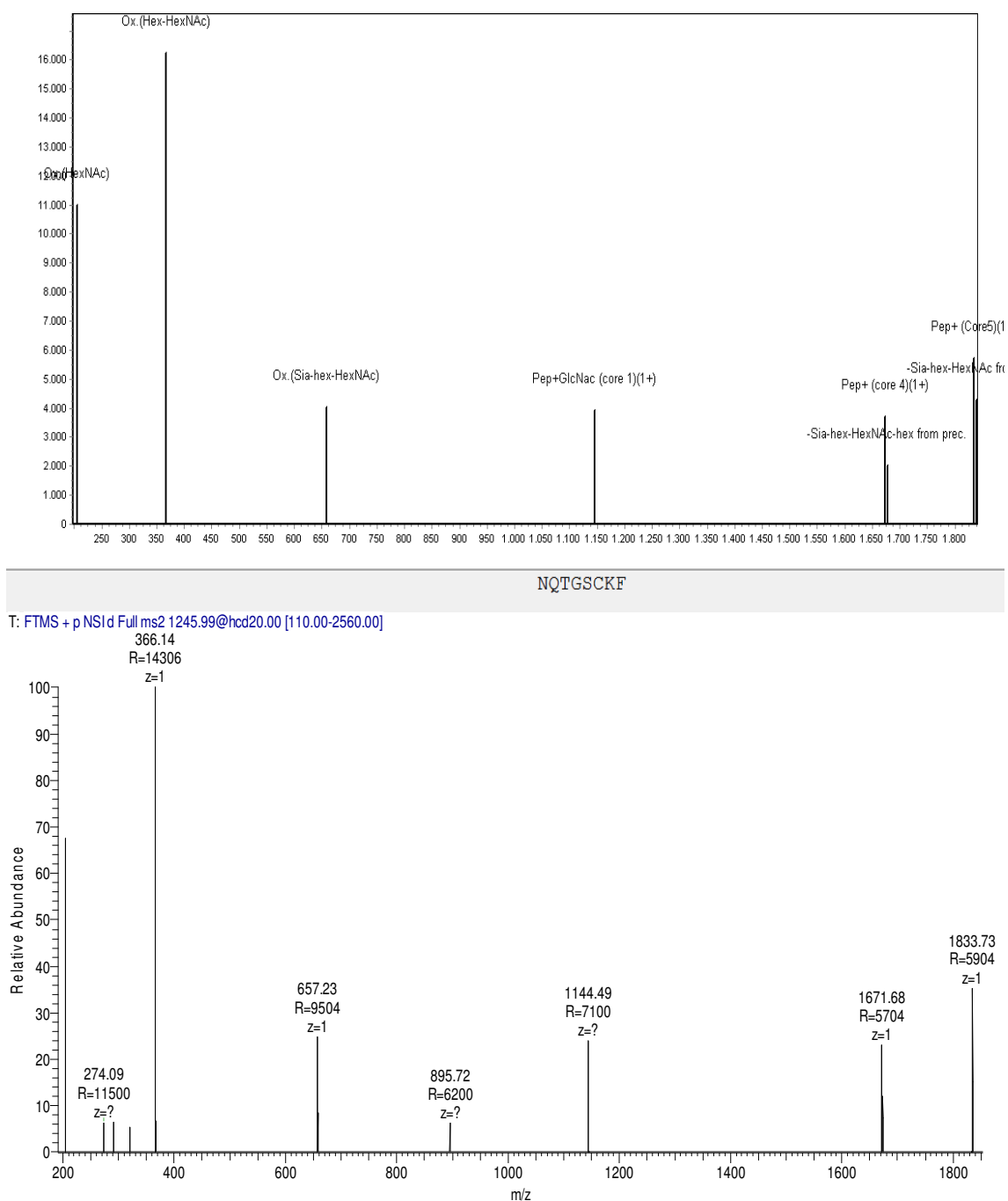
FIG. 28: Bottom: MS/MS spectrum of doubly charged molecular ions of the glycopeptide HexNAc$_2$Hex$_5$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2156.827 Da, which corresponds to the theoretical one 2156.829, with an error of 0.002 Da (1 ppm). Top: MassAI identification.
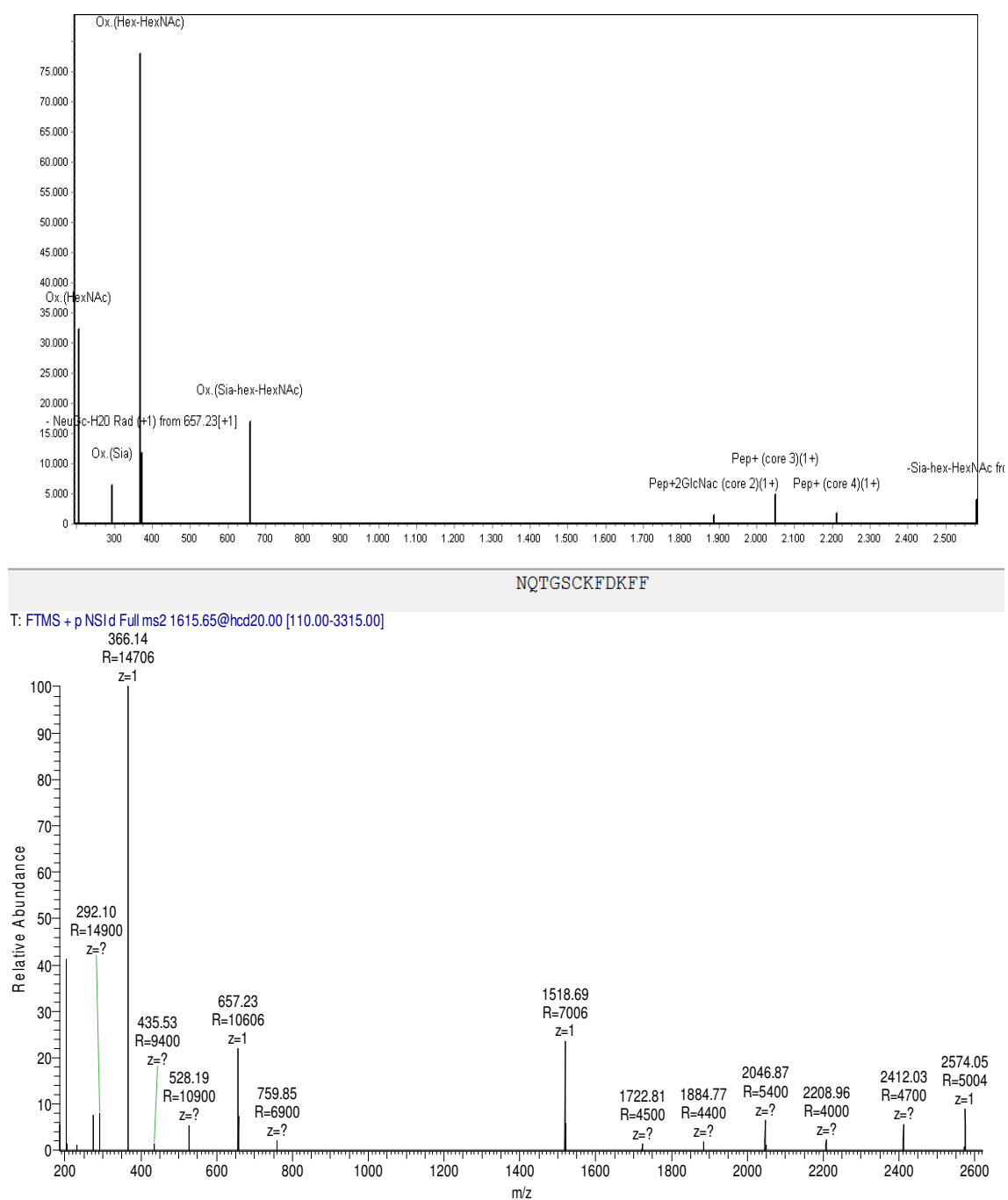
FIG. 29: Bottom: MS/MS spectrum of doubly charged molecular ions of the glycopeptide HexNAc₂Hex₆+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2318.887 Da, which corresponds to the theoretical one 2318.882, with an error of 0.005 Da (2 ppm). Top: MassAI identification.
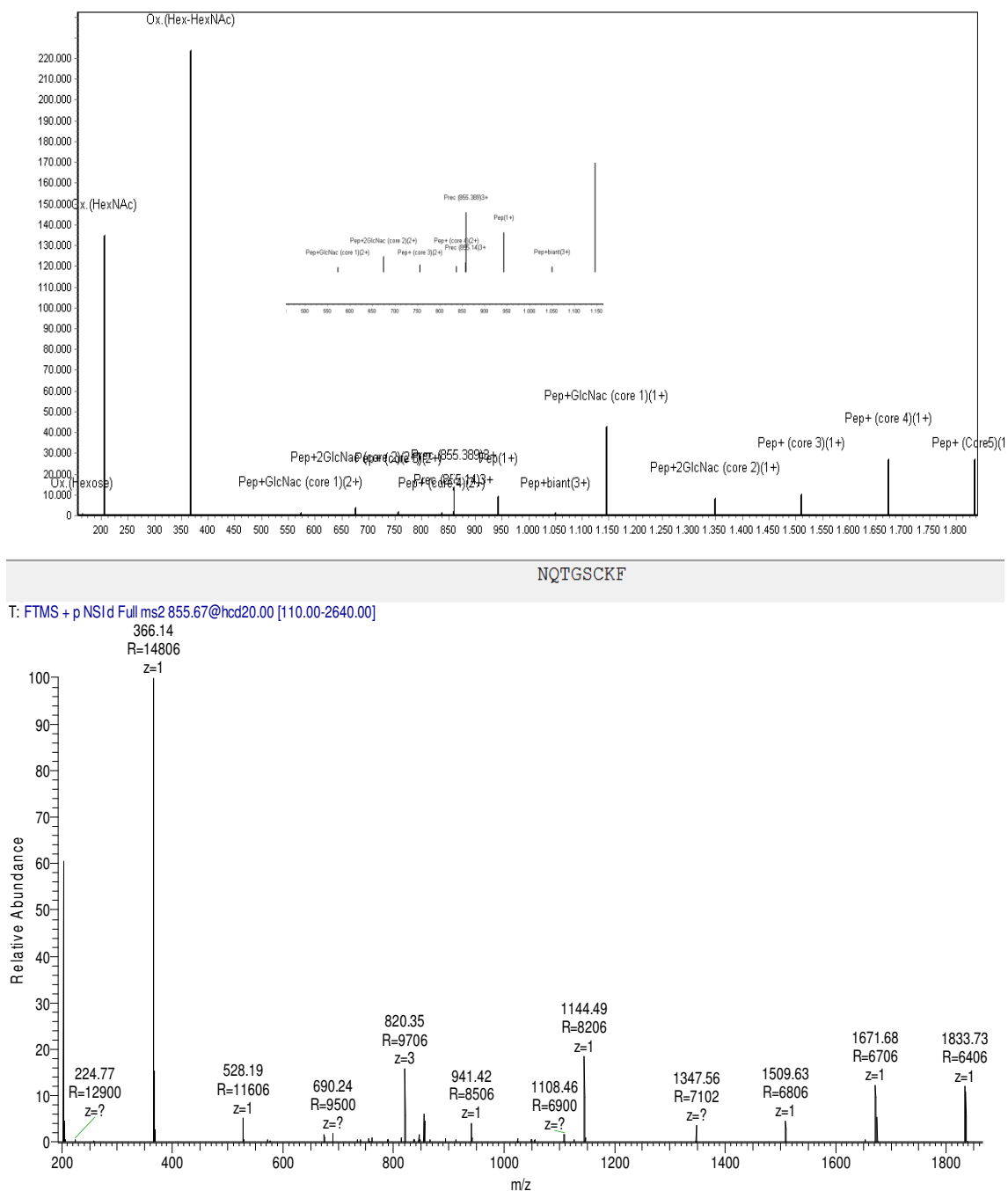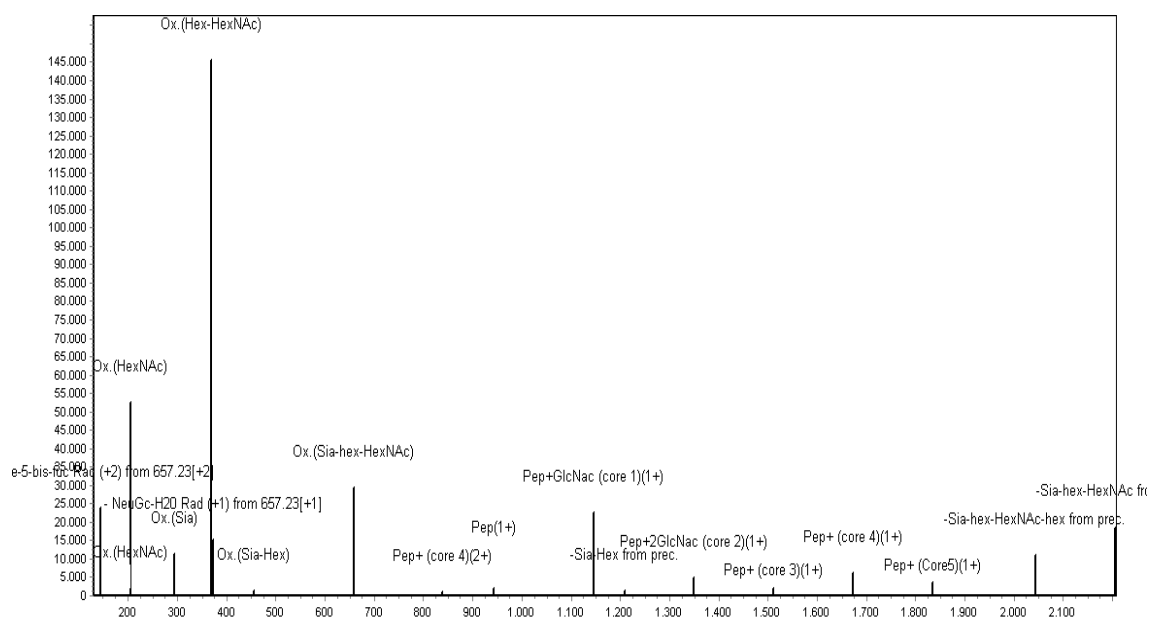
FIG. 30: Bottom: MS/MS spectrum of doubly charged molecular ions of the glycopeptide HexNAc₃Hex₄+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2197.856 Da, which corresponds to the theoretical one 2197.856, with an error of 0.000 Da (0 ppm). Top: MassAI identification.

FIG. 31: Bottom: MS/MS spectrum of doubly charged molecular ions of the glycopeptide HexNAc$_3$Hex$_4$NeuAc$_1$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2488.957 Da, which corresponds to the theoretical one: 2488.951, with an error of 0.006 Da (2 ppm). Top: MassAI identification.
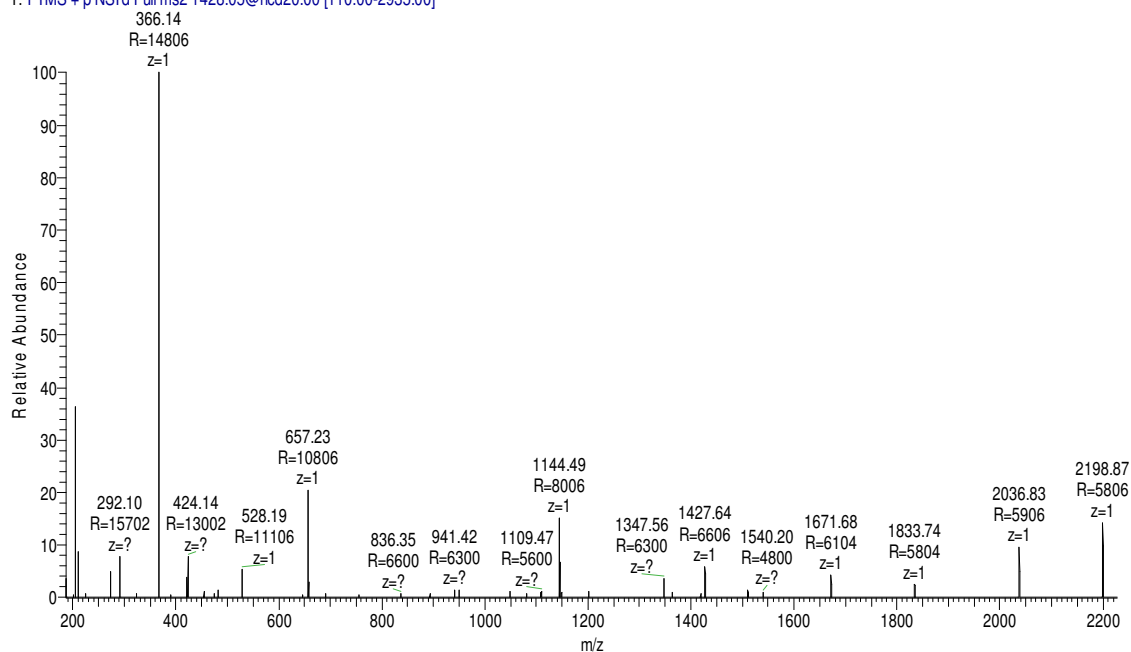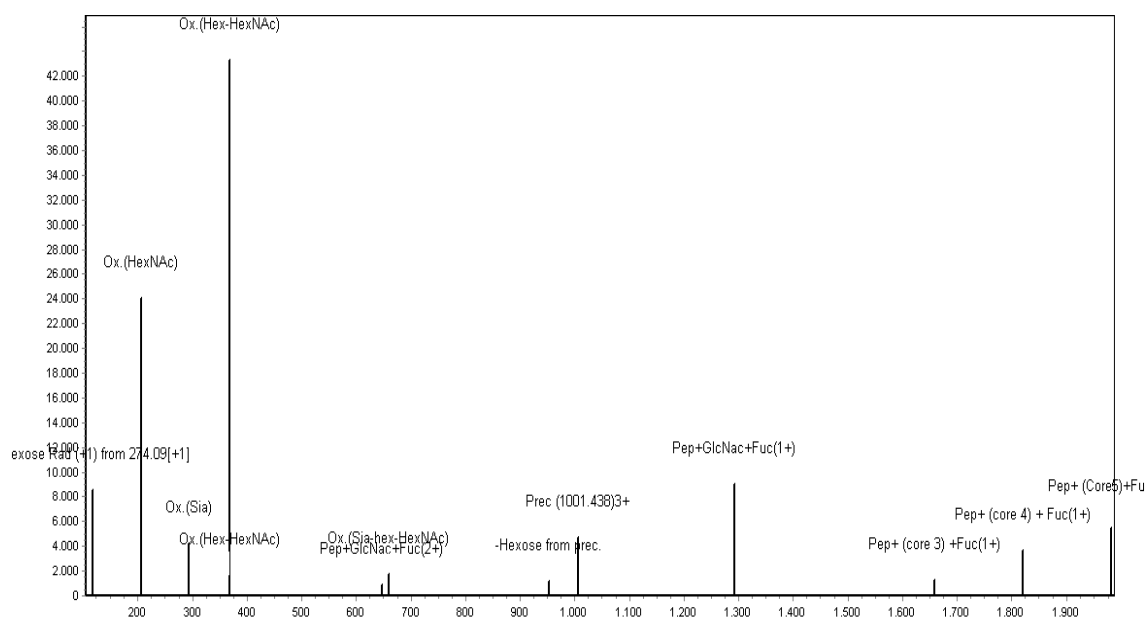
FIG. 32: Bottom: MS/MS spectrum of doubly charged molecular ions of the glycopeptide HexNAc₄Hex₄+NQTGSCKFDKFF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2938.181 Da, which corresponds to the theoretical one 2938.193, with an error of 0.012 Da (5 ppm). Top: MassAI identification.

FIG. 33: Bottom: MS/MS spectrum of doubly charged molecular ions of the glycopeptide HexNAc$_4$Hex$_4$NeuAc$_1$+NQTGSCKFDKFF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3229.285 Da, which corresponds to the theoretical one: 3229.288, with an error of 0.003 Da (1 ppm). Top: MassAI identification.

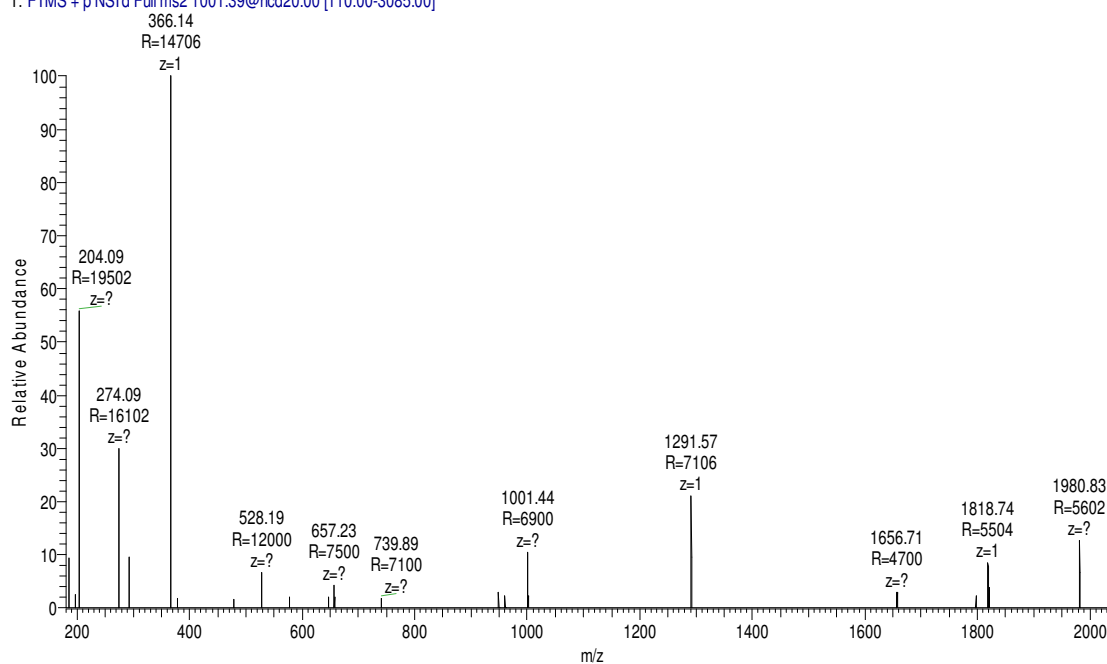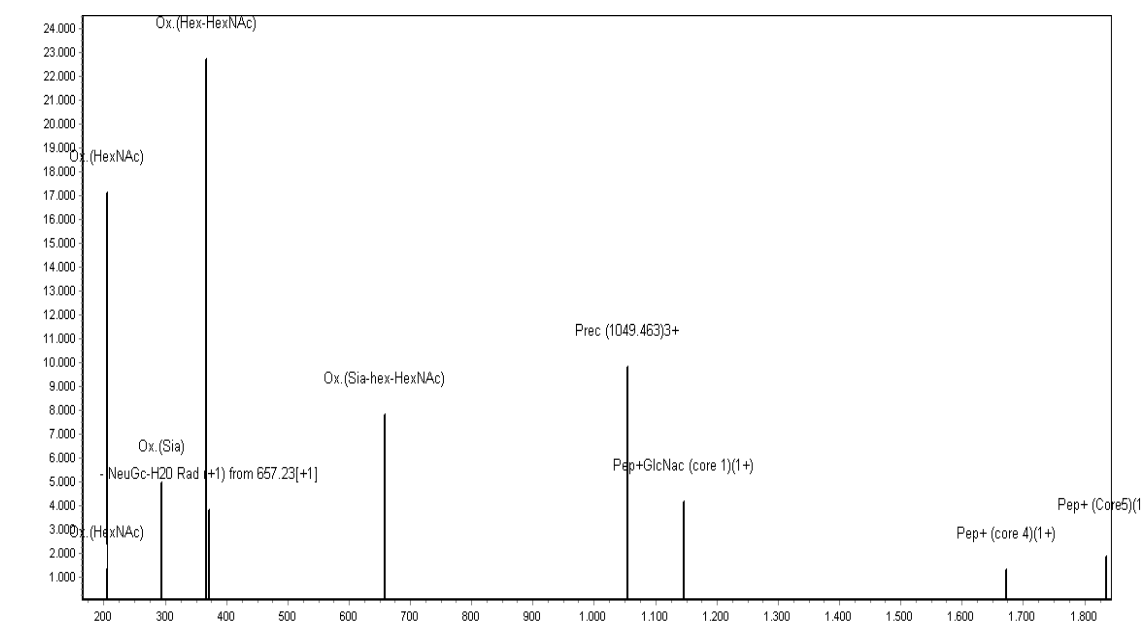**FIG. 34: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₄Hex₅+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2562.987 Da, which corresponds to the theoretical one: 2562.988, with an error of 0.001 Da (0 ppm). Top: MassAI identification.**

FIG. 35: Bottom: MS/MS spectrum of doubly charged molecular ions of the glycopeptide HexNAc$_4$Hex$_5$NeuAc$_1$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2854.082 Da, which corresponds to the theoretical one 2854.083, with an error of 0.001 Da (0 ppm). Top: MassAI identification.
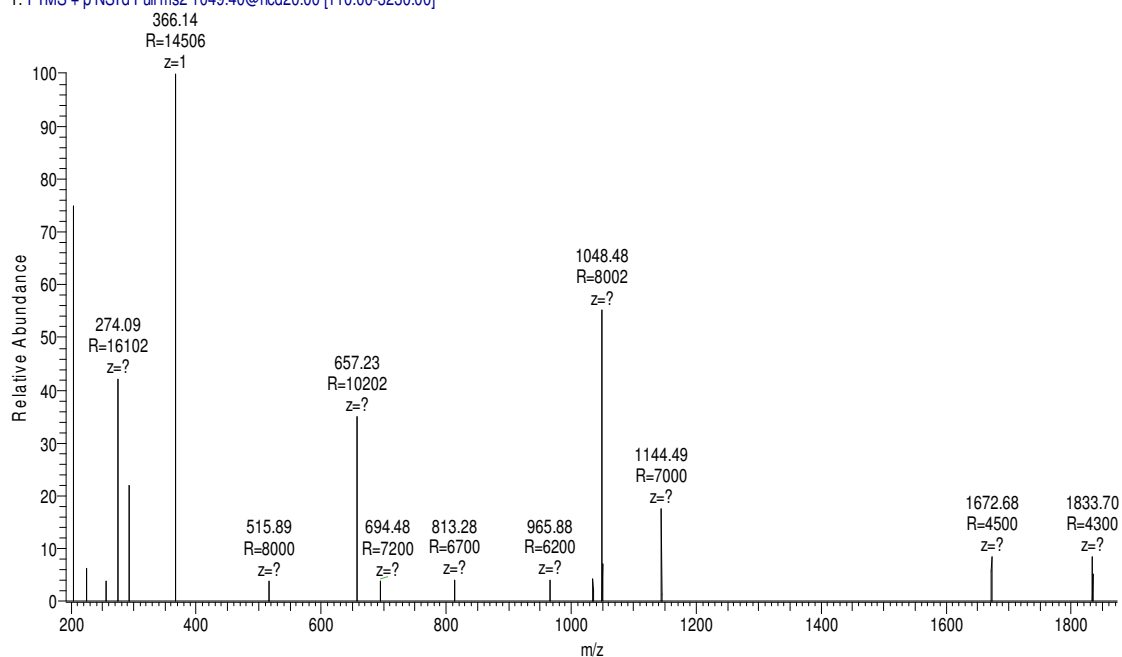
FIG. 36: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc₄Hex₅Fuc₁NeuAc₁+NQTGSCKF(Q->E) in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3001.151 Da, which corresponds to the theoretical one 3001.125, with an error of 0.026 Da (9 ppm). Top: MassAI identification.

FIG. 37: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_4$Hex$_5$NeuAc$_2$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3145.176 Da, which corresponds to the theoretical one 3145,178, with an error of 0.002 Da (1 ppm). **Top:** MassAI identification.
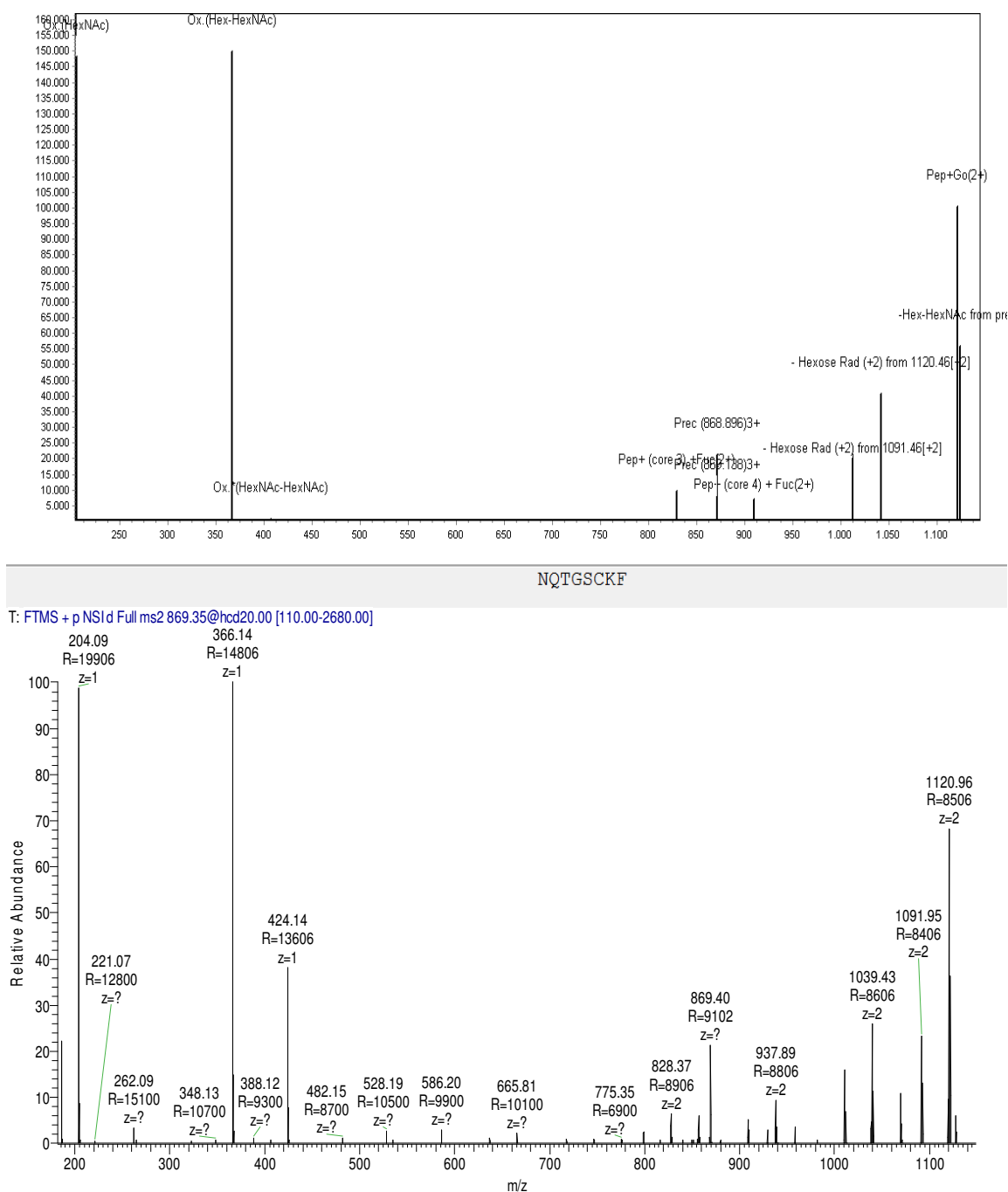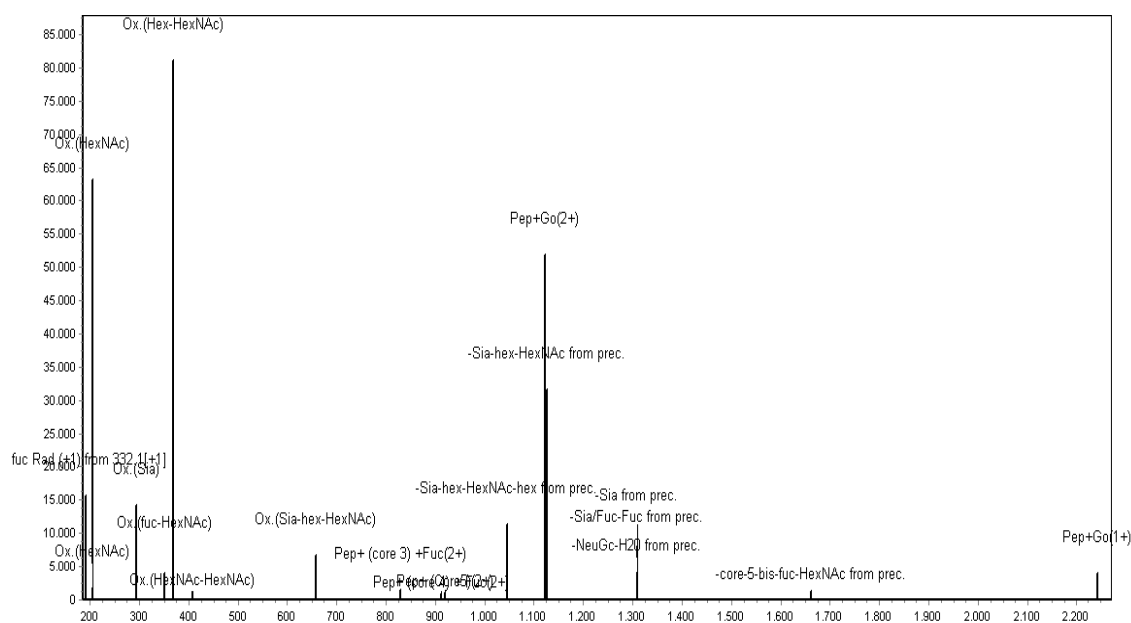
FIG. 38: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_5$Hex$_4$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2604.034 Da, which corresponds to the theoretical one 2604,014 with an error of 0.020 Da (8 ppm). Top: MassAI identification.
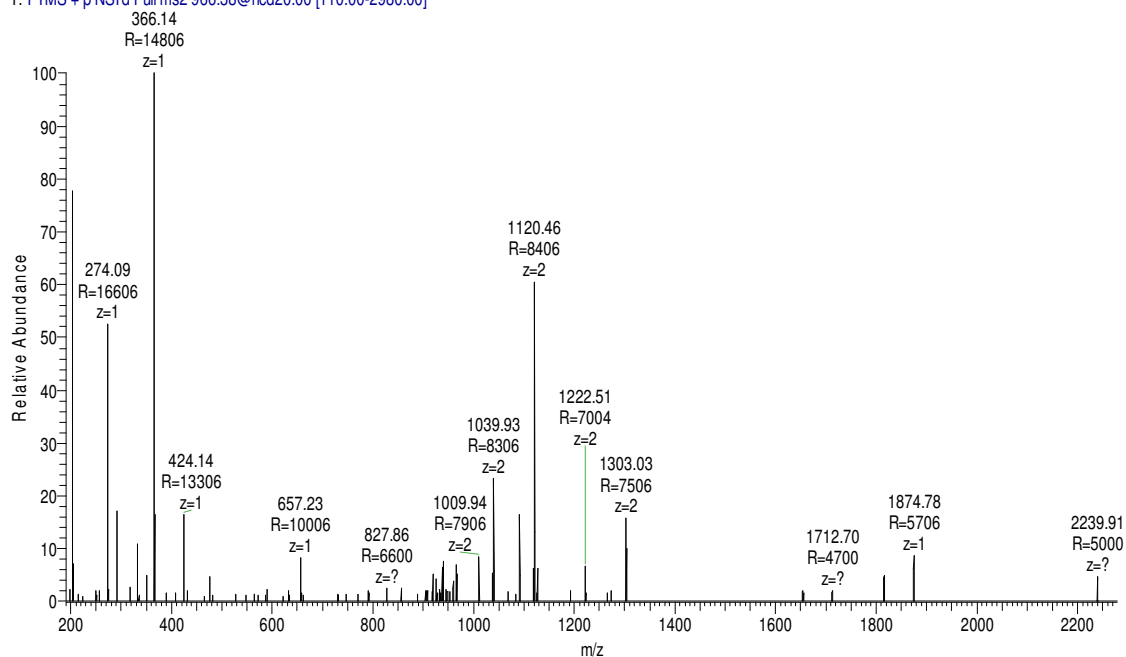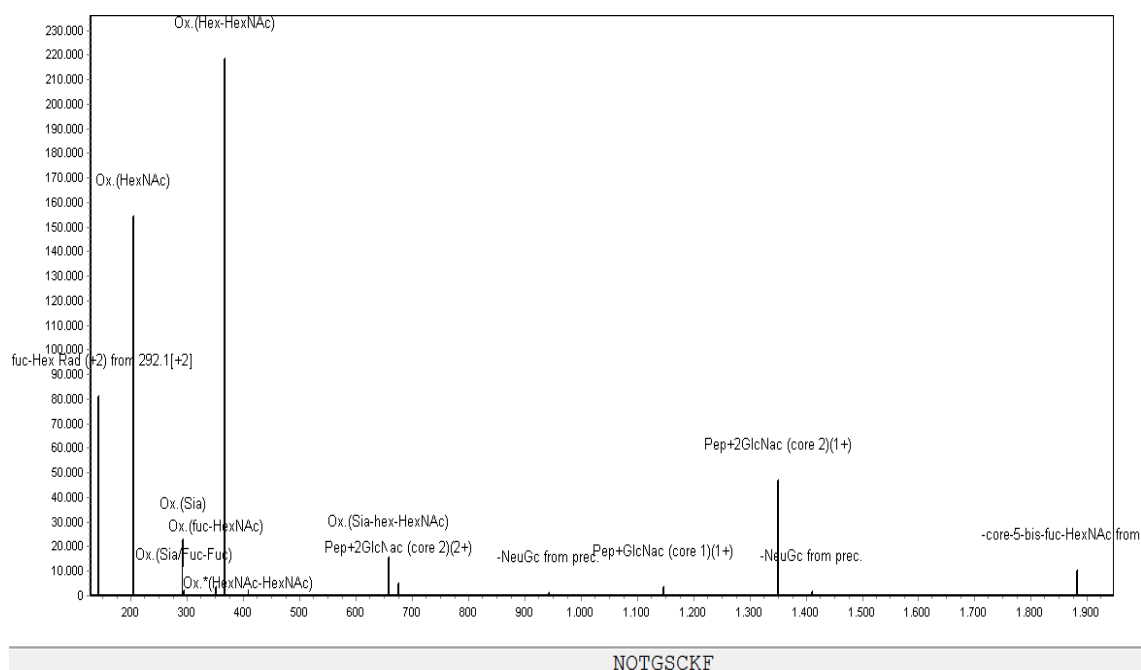
FIG. 39: Bottom: MS/MS spectrum of triply charged selected molecular ions of the glycopeptide HexNAc$_5$Hex$_4$NeuAc$_1$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2895.110 Da, which corresponds to the theoretical one: 2895.110, with an error of 0.000 Da (0 ppm). Top: MassAI identification.

FIG. 40: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_5$Hex$_5$Fuc$_1$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2912.093 Da, which corresponds to the theoretical one: 2912.125, with an error of 0.032 Da (11 ppm). Top: MassAI identification.
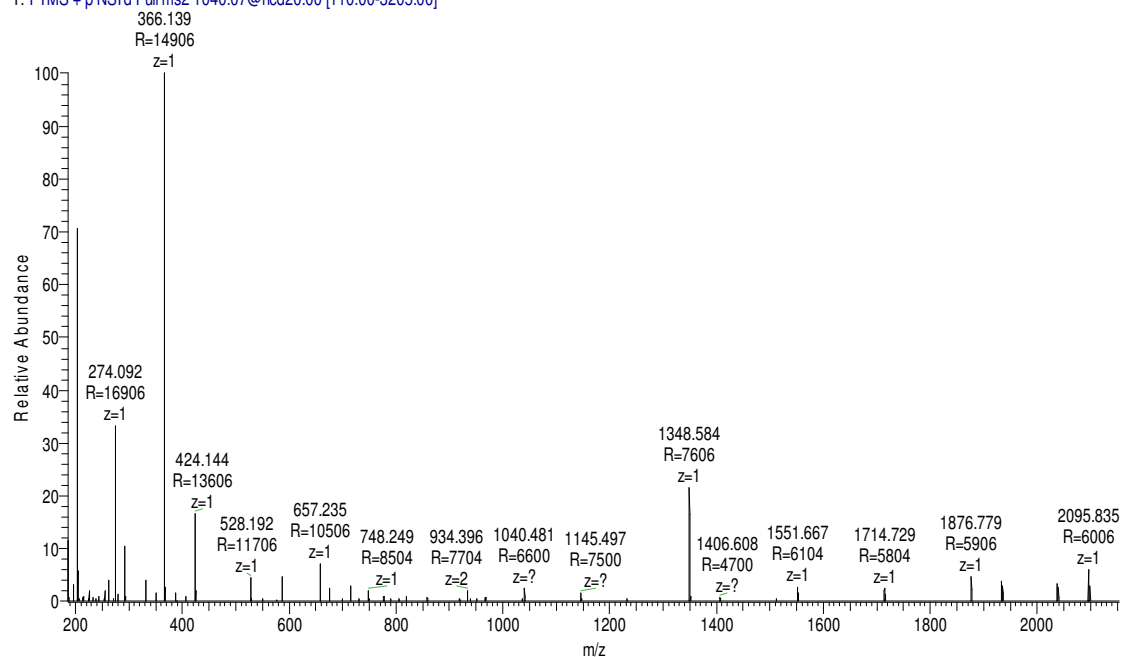
FIG. 41: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_5$Hex$_6$Fuc$_1$+NQTGSCKF (Q->E) in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3116.187 Da, which corresponds to the theoretical one 3116.188, with an error of 0.001 Da (0 ppm). Top: MassAI identification.
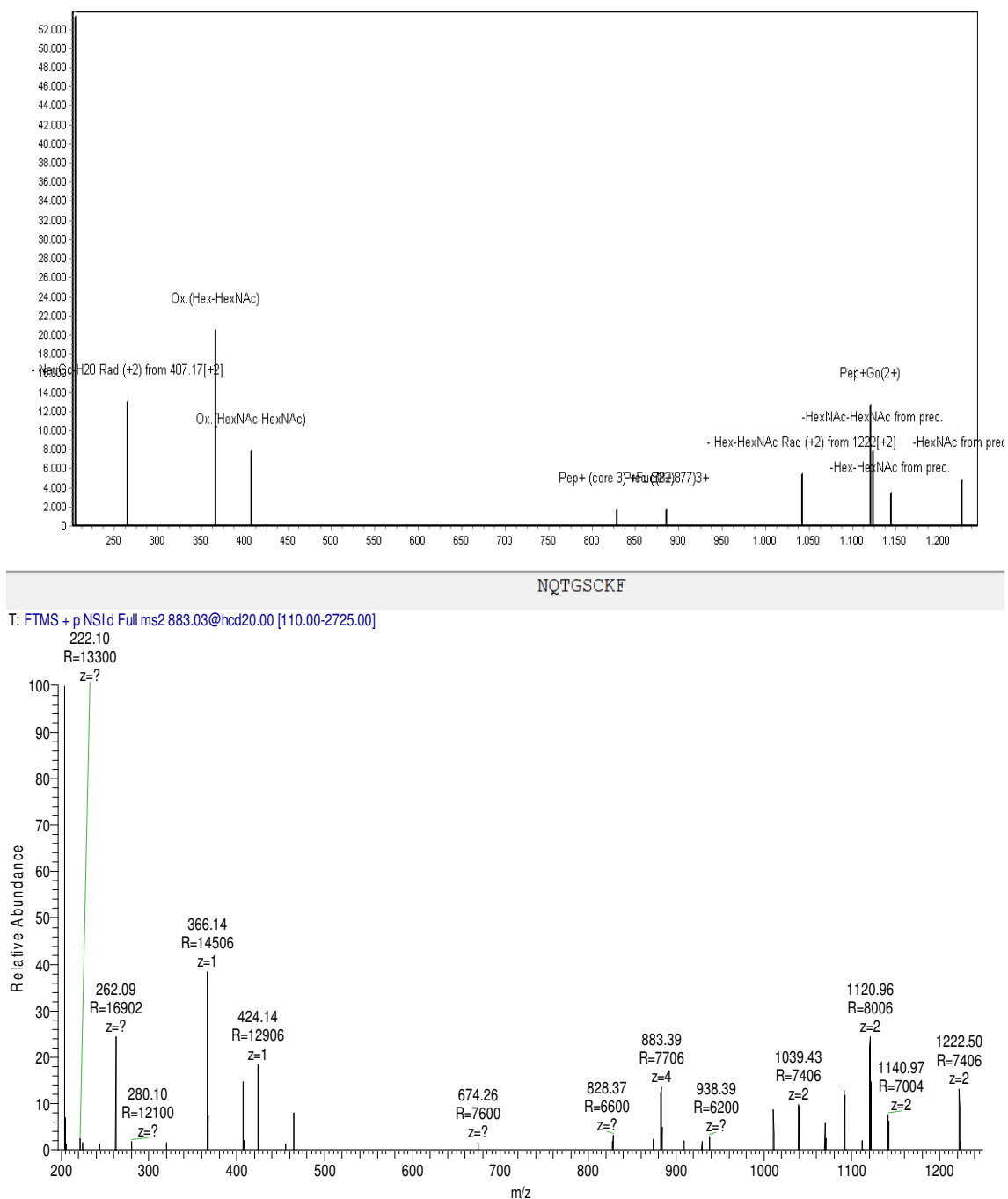
FIG. 42: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_6$Hex$_3$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2645.062 Da, which corresponds to the theoretical one 2645.041, with an error of 0.021 Da (8 ppm). Top: MassAI identification.
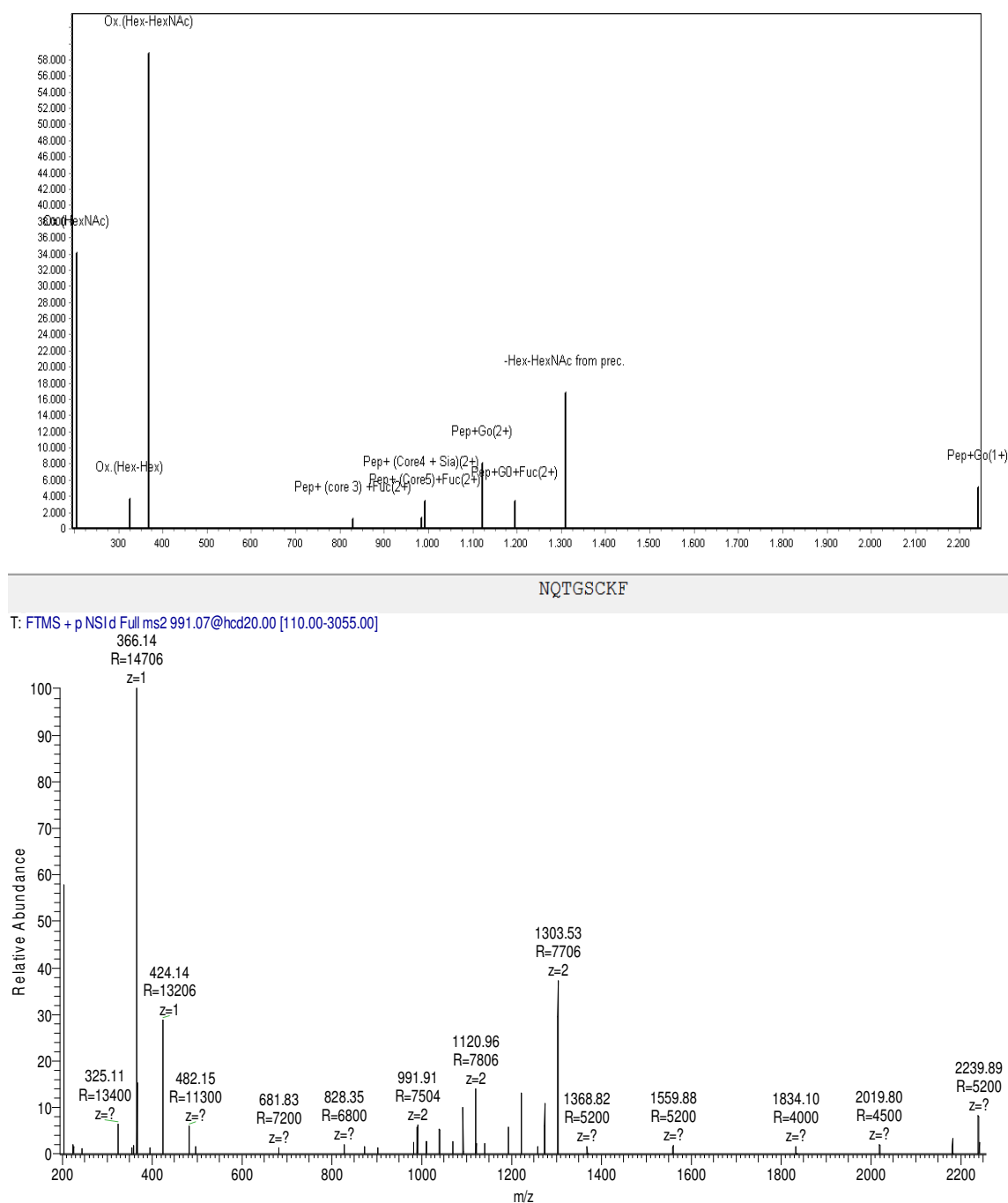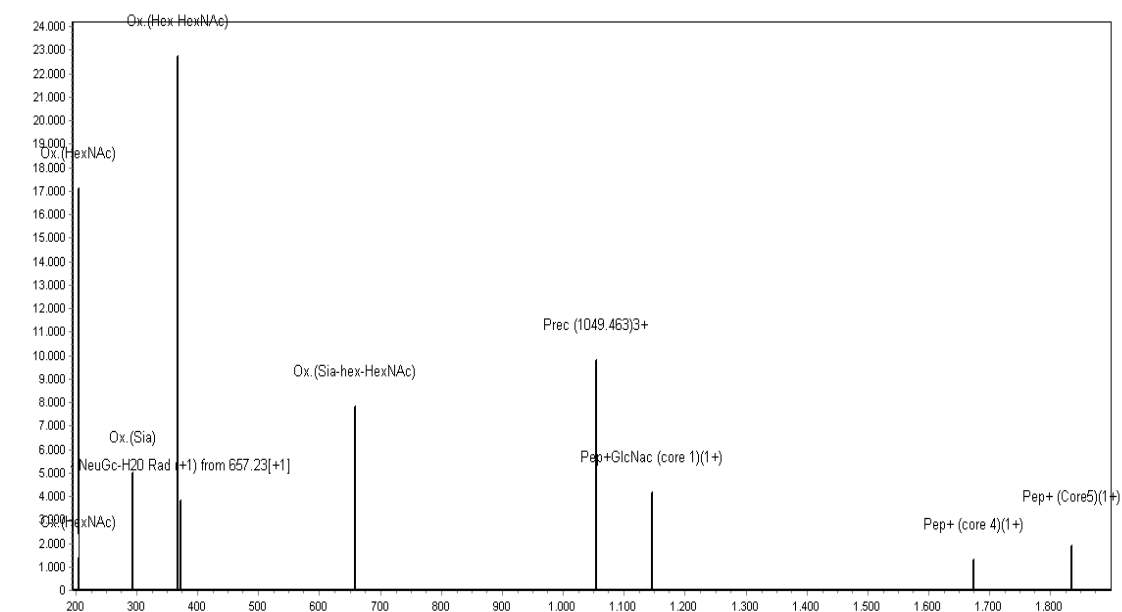
Top spectrum labels: Ox.(Hex-HexNAc), Ox.(HexNAc), Ox.(Hex-Hex), Pep+ (core 3) +Fuc(2+), Pep+ (Core4 + Sia)(2+), Pep+ (Core5)+Fuc(2+), Pep+Go(2+), Pep+G0+Fuc(2+), -Hex-HexNAc from prec., Pep+Go(1+)

NQTGSCKF

T: FTMS + p NSI d Full ms2 991.07@hcd20.00 [110.00-3055.00]

FIG. 43: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_6$Hex$_5$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 2969.170 Da, which corresponds to the theoretical one 2969.146, with an error of 0.024 Da (8 ppm). Top: MassAI identification.
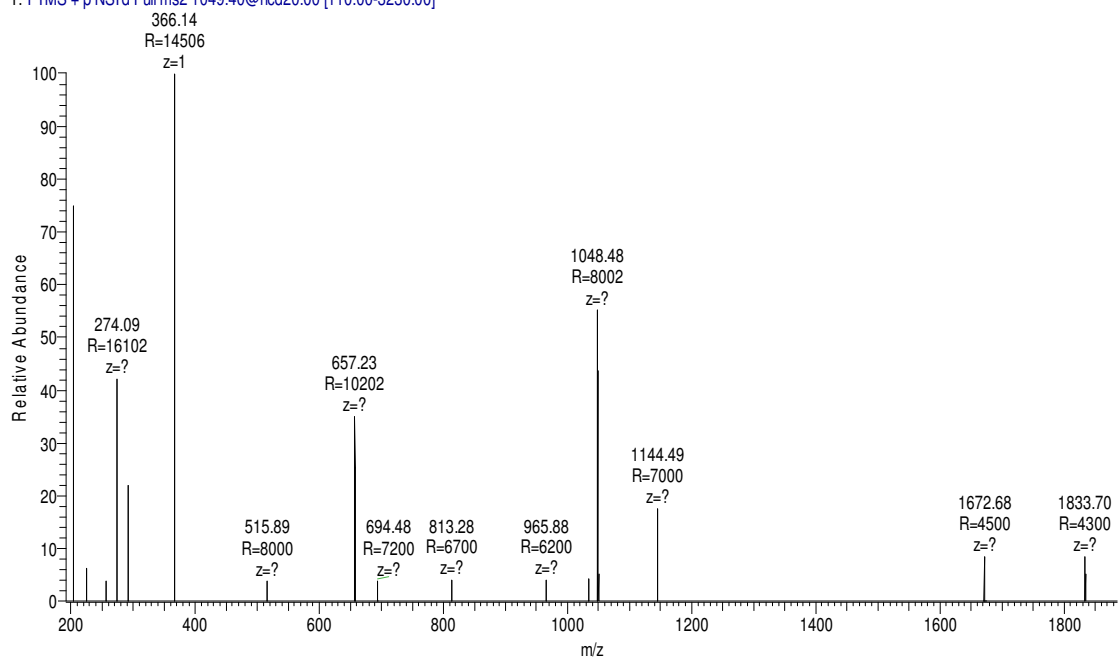
FIG. 44: Bottom: MS/MS spectrum of triply charged molecular ions of the glycopeptide HexNAc$_6$Hex$_5$Fuc$_1$+NQTGSCKF in the RP-HPLC/ESI-MS/MS analysis of the α-chymotryptic digest of donkey lactoferrin. The experimental determined molecular mass of the glycopeptide is 3145.176 Da, which corresponds to the theoretical one 3145.178, with an error of 0.002 Da (1 ppm). Top: MassAI identification.

PUBLICATION LIST

V. Cunsolo , E. Fasoli , R. Saletti , V. Muccilli , <u>S. Gallina</u>, PG. Righetti , S. Foti (2015). "*Zeus, Aesculapius, Amalthea And The Proteome Of Goat Milk*". Journal of Proteomics 128 (2015) 69–82.

<u>S. Gallina</u>, V. Cunsolo, R. Saletti, V. Muccilli, A. Di Francesco, S. Foti, A. M. Lorenzten, P. Roepstorff. "*Sequence Characterization Of Donkey's Milk Lactoferrin By Multiple Enzyme Digestions And Mass Spectrometry*" Manuscript submitted for publication (2015).

<u>S.Gallina</u> , R. Saletti, V. Cunsolo, V.Muccilli, S.Foti, P. Roepstorff.
"*Site-Specific Donkey Milk Lactoferrin Glycosylation Analysis By High Resolution Mass Spectrometry*" Manuscript in preparation

CONGRESS COMUNICATIONS

<u>S. Gallina</u>, R. Saletti, V. Cunsolo, V. Muccilli, A. Di Francesco, S. Foti, P. Roepstorff "*High resolution mass spectrometry in the analysis of site-specific donkey milk lactoferrin glycosylation*" XXXVI Congresso Nazionale della Divisione di Chimica Organica Bologna 13-17 September 2015 (Proceedings, p. 207).

V. Cunsolo, V. Muccilli, R. Saletti, <u>S. Gallina</u>, A. Di Francesco, S. Foti "*Integrated bottom-up/top-down ms-based approach for sequence determination of a donkey' CSN1S2 duplicate gene products*" XXXVI Congresso Nazionale della Divisione di Chimica Organica, 13-17 September 2015 (Proceedings, p. 208).

V. Cunsolo, E. Fasoli, V. Muccilli, S. Gallina, R. Saletti, P.G. Righetti, S. Foti "*A Deep exploration of goat milk proteome via Combinatorial Peptide Ligand Library technology and high resolution mass spectrometry*" 9th Annual Congress of European Proteomics Association – EuPA, 23-28 June 2015, Milan (Proceedings, p.152).

S. Gallina, R. Saletti, V. Cunsolo, V. Muccilli, S. Foti, P. Roepstorff "*Site-Specific Donkey Milk Lactoferrin Glycosylation Analysis by High Resolution Mass Spectrometry*" Convegno MASSA 2015, Alghero 10-12 June 2015

V. Cunsolo, V. Muccilli, S. Gallina, R. Saletti, S. Foti, "*Characterization of Donkey's CSN1S2 Duplicate Gene Products by High Resolution Mass Spectrometry*" 32nd IMMS, Balatonszárszó (Ungheria) 11-14 May 2014 (Proceedings, p. 138).

V. Cunsolo , S. Gallina, V. Muccilli, R. Saletti , S. Foti, Ho JTC. "*Sequence Determination Of A Donkey's Csn1s2 Duplicate Gene Product By High Resolution Mass Spectrometry*". In: IX ItPA Annual Congress. Naples, 24-27 June (2014), (Proceedings, p. 118).

S. Gallina, V. Cunsolo , V. Muccilli, R. Saletti, S. Foti , (2013). "*Sequence Determination Of A Donkey's Csn1s2 Duplicate Gene Product By High Resolution Mass Spectrometry*". In: Convegno Congiunto delle Sezioni Calabria e Sicilia 2013. Catania, 2-3 December 2013, (Proceedings, p.56).

ACKNOWLEDGMENTS

I am grateful to Dr D. Franchina and Dr K.Torrisi (ASILAT srl farm, Milo, Catania) for the gift of the raw donkey milk.

Last but not the least, I would like to thank my family, for supporting me spiritually throughout writing this thesis and my life in general.

REFERENCES LIST

1. H.E. Swaisgood, "Chemistry of the caseins", in *Advanced Dairy Chemistry*, Ed by P.F. Fox. Elsevier, London, **63** (1992).

2. B. Lönnerdal, "Nutritional and physiologic significance of human milk proteins" *Am. J. Clin. Nutr.,* **77**, 1537S (2003).

3. M. Malacarne, F. Martuzzi, A. Summer and P. Mariani, "Protein and fat composition of mare's milk: some nutritional remarks with reference to human and cow's milk", *Int. Dairy J.*,**12**, 869 (2002).

4 . H.Y. Guo, K. Pang, X.Y. Zhang, L. Zhao, S.W. Chen, M.L. Dong et al. "Composition, physicochemical properties, nitrogen fraction distribution, and amino acid profile of donkey milk", *J. Dairy Sci.*, **90**, 1635 (2007).

5 . A.M. Caroli, S. Chessa and G.J. Erhardt, "Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition", *J. Dairy Sci.*, **92**, 5335 (2009).

6. P. Martin, M. Szymanowska, L. Zwierzchowski and C. Leroux, "The impact of genetic polymorphisms on the protein composition of ruminant milks", *Reprod. Nutr. Dev.*, **42**, 433 (2002).

7. K. F. Ng-Kwai Hang, F. Grosclaude, "Genetic polymorphisms of milk proteins", in *Advanced Dairy Chemistry*, Ed by P.F. Fox. Elsevier, London, p. 405 (1992).

8. W.N. Eigel, C.J. Hofmann, B.A. Chibber, J.M. Tomich, T.W. Keenan and E.T. Mertz, "Plasmin-mediated proteolysis of casein in bovine milk", *Proc. Natl. Acad. Sci. USA*, **76**, 2244 (1979).

9. E.D. Bastian and R.J. Brown, "Plasmin in milk and dairy products: an update", *Int. Dairy J.*, **6**, 435 (1996).

10. S.S. Nielsen, "Plasmin system and microbial proteases in milk: characteristics, roles, and relationship", *J. Agric. Food Chem.*, **50**, 6628 (2002).

11. A.T. Andrews, "Proteinases in normal bovine milk and their action on caseins", *J. Dairy Res.*, **50**, 45 (1983).

12. A. D'Alessandro, L. Zolla, A. Scaloni, "The bovine milk proteome: cherishing, nourishing and fostering molecular complexity", [1] D'Alessandro, A., Zolla, L., Scaloni, A., *Molecular Biosystems* 2011, *7*, 579-597.

13. R.J. Merritt, M. Carter, M. Haight and L.D. Eisenberg, "Whey protein hydrolysate formula for infants with gastrointestinal intolerance to cow milk and soy protein in infant formulas", *J. Pediatr. Gastroenterol. Nutr.*, **11**, 78 (1990).

14. D.A. Osborn and J. Sinn, "Formulas containing hydrolysed protein for prevention of allergy and food intolerance in infants", *Cochrane Database Syst. Rev.* **18**, CD003664 (2006).

15 . M. Cavaletto, M.G. Giuffrida and A. Conti, "Milk Fat Globule Membrane Components – A Proteomic Approach", *Adv. Exp. Med. Biol.*, **606**, 129 (2008).

16. H. Meisel, "Biochemical properties of peptides encrypted in bovine milk proteins", *Curr. Med. Chem.*, **12**, 1905 (2005).

17. J.M. Wal, "Bovine milk allergenicity", *Ann. Allergy Asthma Immunol.*, **93**, S2 (2004).

18. L. Monaci, V. Tregoat, A.J. van Hengel ans E. Anklam, "Milk allergens, their characteristics and their detection in food: A review", *Eur. Food Res. Technol.*, **223**, 149 (2006).

19. M. Natale, C. Bisson, G. Monti, A. Peltran, L.P. Garoffo, S. Valentini, C. Fabris, E. Bertino, A. Coscia and A. Conti, "Cow's milk allergens identification by two-dimensional immunoblotting and mass spectrometry", *Mol. Nutr. Food Res.*, **48**, 363 (2004).

20. A. Carroccio, F. Cavataio, G. Montaldo, D. D'Amico, L. Alabrese and G. Iacono, "Intolerance to hydrolysed cow's milk proteins in infants: clinical characteristics and dietary treatment", *Clin. Exp. Allergy*, **30**, 1597 (2000).

21. J. Walker-Smith, "Hypoallergenic formulas: Are they really hypoallergenic?", *Ann. Allergy Asthma Immunol.*, **90**, 112 (2003).

22. K.M. Järvinen, P. Chatchatee. "Mammalian milk allergy: clinical suspicion, cross-reactivities and diagnosis", *Curr. Opin. Allergy Clin. Immunol.* **9**, 251 (2009).

23. E. Bertino, D. Gastaldi, G. Monti, C. Baro, D. Fortunato, L. Perono Garoffo, A. Coscia, C. Fabris, M. Mussap, C. Conti, "Detailed proteomic analysis on DM: insight into its hypoallergenicity", *Front. Biosci.*, **1**, 526 (2010).

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

25. V. Cunsolo, V. Muccilli, E. Fasoli, R. Saletti, P.G. Righetti, S. Foti, "Poppea's bath liquor: the secret proteome of she-donkey's milk", *J. Proteomics*, **74**, 2083 (2011).

26. V. Cunsolo, E. Cairone, R. Saletti, V. Muccilli, S. Foti. Sequence and phosphorylation level determination of two donkey β-caseins by mass spectrometry. *Rapid Commun. Mass Spectrom.*, **23**, 1907 (2009).

27. V. Cunsolo, E. Cairone, D. Fontanini, A. Criscione, V. Muccilli, R. Saletti, S. Foti, "Sequence determination of α_{S1}-casein isoforms from donkey by mass spectrometric methods", *J. Mass Spectrom.*, **44**, 1742 (2009)

28. R. Saletti, V. Muccilli, V. Cunsolo, D. Fontanini, A. Capocchi, S. Foti, "MS-based characterization of α_{S2}-casein isoforms in donkey's milk. *J. Mass Spectrom.* **47**, 1150 (2012).

29. V. Cunsolo, V. Muccilli, R. Saletti, S. Foti, "Applications of mass spectrometry techniques in the investigation of milk proteome" Review, *Eur. J. Mass Spectrom.*, **17**, 305 (2011).

30. A. Criscione, V. Cunsolo, S. Bordonaro, A. M. Guastella, R. Saletti, A. Zuccaro, G. D'Urso, and D. Marletta. "Donkey's milk protein fraction investigated by electrophoretic methods and mass spectrometric analysis", *Int. Dairy J.,* **19**, 190, (2009).

31. F. Lara-Villoslada, M. Olivares, J. Xaus. "The balance between caseins and whey proteins in cow's milk determines its allergenicity" *J. Dairy Sci.*, **88**, 1654e, (2005).

32. M. G. Giuffrida, A. Cantisani, L. Napolitano, A. Conti, J. Godovac-Zimmerman, "The amino-acid sequence of two isoforms of α-lactalbumin from donkey (Equus asinus) milk is identical", *Biol. Chem. Hoppe-Seyler*, **373**, 931, (1992).

33. A. Conti, L. Napolitano, P. Lai, W. Pinna, J. Godovac-Zimmerman, "Isolation of donkey whey proteins and N-terminal amino acid sequence of α-lactoalbumins A and B, β-lactoglobulins I and II and lysozyme", *Milchwissenschaft*, **44**, 138, (1989).

34. V. Cunsolo, A. Costa, R. Saletti, V. Muccilli, S. Foti, "Detection and sequence determination of a new variant β-lactoglobulin II from donkey", *Rapid Commun. Mass Spectrom.*, **21**, 1438, (2007).

35. L. Chianese, C. De Simone, P. Ferranti, R. Mauriello, A. Costanzo, M. Quarto, G. Garro, G. Picariello, G. Mamone, L. Ramunno, "Occurrence of qualitative and quantitative polymorphism at donkey β-lactoglobulin II locus", *Food Res. Intern.*, **54**, 1273, (2013).

36. M. Herrouin, D. Molle, J. Fauquant, F. Ballestra, J. L. Maubois, J. Leonil, "New Genetic Variants Identified in Donkey's Milk Whey Proteins", *J. Protein Chem.*, **19**, 105,(2000).

37. V. Cunsolo, R. Saletti, V. Muccilli, S. Foti, "Characterization of the Protein Profile of Donkey's Milk Whey Fractio", *J. Mass Spectrom.*, **42**, 1162, (2007).

38. X. Y. Zhang, L. Zhao, L. Jiang, M. L. Dong, F. Z. Ren, "The antimicrobial activity of donkey milk and its microflora changes during storage" *Food Control*, **19**, 1191. (2008).

39 . D. Legrand, A. Pierce, E. Elass, M. Carpentier, C. Mariller, J. Mazurier, "Lactoferrin structure and functions", *Adv. Exp. Med. Biol.*, **606**, 163, (2008).

40. M. Soerensen, S. P. L. Soerensen, "The proteins in whey", C. R. Trav. Lab. Carlsberg **23**, 55-99, (1939).

41. B. G. Johansson, "Isolation of an iron containing red protein from human milk", Acta Chem. Scand., **14,** 510-512, (1960).

42 . I. A. Garcia-Montoya, T. S. Cendon, S. Arevalo-Gallegos, Q. Rascon-Cruz, "Lactoferrin a multiple bioactive protein: An overview", *Biochim. Biophys. Acta*. **1820,** 226-236, (2011).

43. B. Lönnerdal, "Nutritional and physiologic significance of human milk proteins", *Am J Clin Nutr.* , **77**,1537S-1543S, (2003).

44 . T. Nagasawa, I. Kiyosawa, K. Kuwahara, "Amounts of lactoferrin in human colostrum and milk", *J Dairy Sci.*, **55**, 1651-1659, (1972).

45. J. T. Smilowitz, S. M. Totten, J. Huang, D. Grapov, H. A. Durham, C. J. Lammi-Keefe, C. Lebrilla, J. B. German, "Human Milk Secretory Immunoglobulin A and Lactoferrin N-Glycans Are Altered in Women with Gestational Diabetes Mellitus", *J. Nutr.*, **143**, 1906-1912, (2013).

46. R. J. Harmon, F. L. Schanbacher, L. C. Ferguson, K. L. Smith, "Concentration of lactoferrin in milk of normal lactating cows and changes occurring during mastitis", *Am. J. Vet. Res*., **36**, 1001-1007. (1975).

47 . S. Hiss, T. Meyer, H. Sauerwein, "Lactoferrin concentrations in goat milk throughout lactation", *Small Ruminant Res*., **80** ,87-90, (2008).

48. J. W. Froehlich, E. D. Dodds, M. Barboza, E. L. McJimpsey, R. R. Seipert, J. Francis, H. J. An HJ, S. Freeman, J. B. German JB, C. Lebrilla. "Glycoprotein Expression in Human Milk during Lactation", *J. Agric. Food. Chem*., **58**, 6440-6448, (2010).

49. P. F. Levay, M. Viljoen, "Lactoferrin: a general review", *Haematologica* **80,** 252-267, (1995).

50. B. Lonnerdal, S. Iyer, "Lactoferrin: molecular structure and biological function", *Annu. Rev. Nutr*. **15**, 93–110, (1995).

51. P. L. Masson, J. F. Heremans, C. Dive, "An iron-binding protein common to many external secretions", *Clin. Chim. Acta*, **14**, 735-739, (1966).

52 . M. Baggiolini, C. De Duve, P. L. Masson, J. F. Heremans, "Association of lactoferrin with specific granules in rabbit heterophil leukocytes", *J. Exp. Med*. **131**, 559–570, (1970).

53. E. N. Baker, H. M. Baker, "Molecular structure, binding properties and dynamics of lactoferrin", *Cell Mol. Life Sci.* **62**, 2531-2539, (2005).

54. M. E. Redwan, V. N. Uversky , E. M. El-Fakharany , H. Al-Mehdar, "Potential lactoferrin activity against pathogenic viruses", *Comptes Rendus Biologies* **337**, 581-595, (2014).

55. B. F. Anderson, H. M. Baker, G. E. Norris, D. W. Rice, E. N. Baker, "Structure of human lactoferrin: crystallographic analysis and refinement at 2.8Å resolution", *J. Mol. Biol.* **209**, 711-734, (1989).

56. M. Haridas, B. F. Anderson, E. N. Baker, "Structure of human diferric lactoferrin refined at 2.2 Å resolution", *Acta Crystallogr*., **D51**, 629-646, (1995).

57. S. A. Moore, B. F. Anderson, C. R. Groom, M. Haridas and E. N. Baker Three dimensional structure of diferric bovine lactoferrin at 2.8 Å resolution. *J. Mol. Biol.* **274**, 222-236, (1997).

58. S. Karthikeyan, M. Paramasivam, S. Yadav, R. Srinivasan, T. P. Singh, "Structure of buffalo lactoferrin at 2.5 Å re solution using crystals grown at 303 K shows different orien tations of the N and C lobes", *Acta Crystallogr.*, **D55**, 1805-1813, (1999).

59. A. K. Sharma, M. Paramasivam, A. Srinivasan, M. P. Yadav, T. P. Singh, "Three-dimensional structure of mare diferric lactoferrin at 2.6 Å resolution", *J. Mol. Biol.*, **289**, 303-317, (1998).

60. J. A. Khan, P. Kumar, M. Paramasivam, R. S. Yadav, M. S. Sahani, S. Sharma, "Camel lactoferrin – a transferrincum-lactoferrin: crystal structure of camel apolactoferrin at 2.6 Å resolution and structural basis of its dual role", *J. Mol. Biol.*, **309,** 751-761, (2001).

61. S. Sharma, M. Sinha, S. Kaushik, P. Kaur, T. P. Singh, "C-Lobe of Lactoferrin: The Whole Story of the Half-Molecule", *Biochem. Res.Int.*, **Article ID 271641**, (2013).

62 . S. Baveye, E. Elass, J. Mazurier, G. Spik, D. Legrand, "Lactoferrin: a multifunctional glycoprotein involved in the modulation of the in- flammatory process", *Clin. Chem. Lab. Med.*, **37,** 281-286, (1999).

63. B. F. Anderson, H. M. Baker, E. J. Dodson, G. E. Norris, S. V. Rumball, J. M. Waters, E. N. Baker, "Structure of human lactoferrin at 3.2 A resolution", *Proc. Natl. Acad. Sci. USA,* **84**, 1769-1773, (1987).

64. E. N. Baker, H. M. Baker, "A structural framework for understanding the multifunctional character of lactoferrin", *Biochimie*, **91**, 3-10, (2009).

65. E. N. Baker, B. F. Anderson, H. M. Baker, C. L. Day, M. Haridas, G. E. Norris, S. V. Rumball, C. A. Smith, D. H. Thomas, "Three-dimensional structure of lactoferrin in various functional states", *Adv. Exp. Med. Biol.*, **357**, 1-12, (1994).

66. E. N. Baker, "Structure and reactivity of transferrins", *Adv. Inorg. Chem.*, **41**, 389-463 , (1994).

67. P.F. Levay, M. Viljoen, "Lactoferrin: a general review", *Haematologica*, **80**, 252-267, (1995).

68. B. Lonnerdal, S. Iyer, "Lactoferrin: molecular structure and biological function", *Annu. Rev. Nutr.*, **15**, 93–110, (1995).

69. J. He, P. Furmanski, "Sequence specificity and transcriptional activation in the binding of lactoferrin to DNA", *Nature*, **373**, 721-724, (1995).

70. E. Elass-Rochard, A. Roseanu, D. Legrand, M. Trif, V. Salmon, C. Motas, J. Montreuil, G. Spik, "Lactoferrin-lipopolysaccharide interaction: involvement of the 28-

34 loop region of human lactoferrin in the high affinity binding to Escherichia coli 055B5 lipopolysaccharide", *Biochem. J.*, **312**, (Pt 3), 839-845, (1995).

71. R. R. Arnold, J. E. Russell, W. J. Champion, M. Brewer, J. J. Gauthier, "Bactericidal activity of human lactoferrin: differentiation from the stasis of iron deprivation", *Infect. Immunol.*, **35**, 792-799, (1982).

72. A. Aguila, A. Herrera, W. Velzquez, "Isolation and structure-functional characterization of human colostral lactoferrin", *Biotecnologia Aplicada*, **17**, 177-182, (2000).

73. B.W. Van der Strate, L. Beljaars, G. Molema, M. C. Harmsen, D. K. Meijer, "Antiviral activities of lactoferrin", *Antiviral Res.*, **52**, 225-239, (2001).

74. S. P. Crouch, K. J. Slater, J. Fletcher, "Regulation of cytokine release from mononuclear cells by the iron-binding protein lactoferrin", *Blood*, **80**, 235-240, (1992).

75. L. Sanchez, M. Calvo, J. H. Brock, "Biological role of lactoferrin", *Arch. Dis. Child*, **67**, 657-661, (1992).

76. S. Hashizume, K. Kuroda, H. Murakami, "Identification of lactoferrin as an essential growth factor for human lymphocytic cell lines in serumfree medium", *Biochim. Biophys. Acta*, **763**, 377-382, (1983).

77. C. Mariller, S. Hardiville, E. Hoedt, I. Huvent, S. Pina-Canseco, A. Pierce, "Delta-lactoferrin, an intracellular lactoferrin isoform that acts as a transcription factor", *Biochem. Cell Biol.*, **90**, 307-319, (2012).

78. S. Hardiville, A. Escobar-Ramirez, S. Pina-Canceco, E. Elass, A. Pierce, "Delta-lactoferrin induces cell death via the mitochondrial death signaling pathway by upregulating bax expression", *Biometals*, **27 (5)**, 875-889, (2014).

79. L. Sanchez, M. Calvo, J. H. Brock, "Biological role of lactoferrin", *Arch. Dis. Child*, **67**, 657-661, (1992).

80. S. Farnaud, R. W. Evans, "Lactoferrin – a multifunctional protein with antimicrobial properties", *Mol. Immunol.*, **40**, 395-405, (2003).

81. P. Valenti, M. Marchetti, F. Superti, M. G. Ammendolia, P. Puddu, S. Gessani, P. Borghi, F. Belardelli, G. Antonini, L. Seganti, "Antiviral activity of lactoferrin", *Advances in Lactoferrin Research,* **443**,199-203, (1998).

82. P. Valenti, G. Antonini, "Lactoferrin: an important host defence against microbial and viral attack", *Cell. Mol. Life Sci.*, **62**, 2576-2587, (2005).

83. S. Farnaud, R. W. Evans, "Lactoferrin – a multifunctional protein with antimicrobial properties", *Mol. Immunol.*, **40**, 395-405, (2003).

84. C. A. Bortner, R. R. Arnold, R. D. Miller, "Bactericidal effect of lactoferrin on Legionella pneumophila: effect of the physiological state of the organism", *Can. J. Microbiol.*, **35**, 1048-1051, (1989).

85. C. A. Bortner, R. D. Miller, R. R. Arnold, "Bactericidal effect of lactoferrin on Legionella pneumophila", *Infect. Immunol.*, **51**, 373-377, (1986).

86. C. H. Kirkpatrick, I. Green, R. R. Rich, A. L. Schade, "Inhibition of growth of Candida albicans by iron-unsaturated lactoferrin: relation to hostdefense mechanisms in chronic mucocutaneous candidiasis", *J. Infect. Dis.*, **124**, 539-544, (1971).

87. D. A. Rodriguez, L. Velazquez, G. Ramos, "Antimicrobial mechanisms and potencial clinical application of lactoferrin", *Rev. Latinoam. Microbiol.*, **47**, 102-111, (2005).

88. S. M. E. Drago, "Actividades antibacterianas de la lactoferrina", *Enf. Inf. Microbiol.*, **26 (2)**, 58-63, (2006).

89. K. Yamauchi, H. Wakabayashi, K. Shin, M. Takase, "Bovine lactoferrin: benefits and mechanism of action against infections", *Biochem. Cell Biol.*, **84**, 291-296, (2006).

90. M. P. Sherman, S. H. Bennett, F. F. Hwang, C. Yu, "Neonatal small bowel epithelia: enhancing anti-bacterial defense with lactoferrin and Lactobacillus GG", *Biometals*, **17**, 285-289, (2004).

91. M. Tomita, W. Bellamy, M. Takase, K. Yamauchi, H. Wakabayashi, K. Kawase, "Potent antibacterial peptides generated by pepsin digestion of bovine lactoferrin", *J. Dairy Sci.*, **74**, 4137-4142, (1991).

92. W. Bellamy, M. Takase, K. Yamauchi, H. Wakabayashi, K. Kawase, M. Tomita, "Identification of the bactericidal domain of lactoferrin", *Biochim. Biophys. Acta*, 1121:130–6, (1992).

93. M. E. Redwan, V. N. Uversky , E. M. El-Fakharany , H. Al-Mehdar, "Potential lactoferrin activity against pathogenic viruses", *Comptes Rendus Biologies* **337**, 581-595, (2014).

94. K. Hasegawa, W. Motsuchi, S. Tanaka, S. Dosako, "Inhibition with lactoferrin of in vitro infection with human herpes virus", *Jpn. J. Med. Sci. Biol.*, **47**, 73-85, (1994).

95. K. Hara, M. Ikeda, S. Saito, S. Matsumoto, K. Numata, N. Kato, K. Tanaka, H. Sekihara, "Lactoferrin inhibits hepatitis B virus infection in cultured human hepatocytes", *Hepatol. Res.*, **24**, 228, (2002).

96. J. H. Andersen, H. Jenssen, K. Sandvik, T. J. Gutteberg, "Anti-HSV activity of lactoferrin and lactoferricin is dependent on the presence of heparan sulphate at the cell surface", *J. Med. Virol.*, **74**, 262-271, (2004).

97. B. L. Waarts, O. J. Aneke, J. M. Smit, K. Kimata, R. Bittman, D. K. Meijer, J. Wilschut, "Antiviral activity of human lactoferrin: inhibition of alphavirus interaction with heparan sulfate", *Virology*, **333**, 284-292, (2005).

98. R. M. Redwan, A. Tabll, "Camel lactoferrin markedly inhibits hepatitis C virus genotype 4 infection of human peripheral blood leukocytes", *J. Immunoassay Immunochem.*, **28**, 267-277, (2007).

99. J. H. Nuijens, P. H. van Berkel, M. E. Geerts, P. P. Hartevelt, H. A. de Boer, H. A. van Veen, F. R. Pieper, "Characterization of recombinant human lactoferrin secreted in milk of transgenic mice", *J. Biol. Chem.*, **272**, 8802-8807, (1997).

100. E. M. El-Fakharany, A. Tabll, A. Abd El-Wahab, B. M. Haroun, E. M. Redwan, "Potential activity of camel milk-amylase and lactoferrin against hepatitis C virus infectivity in HepG2 and lymphocytes", *Hepat. Mon.*, **8**, 101-109, (2008).

101. M. C. Harmsen, P. J. Swart, M. P. de Bethune, R. Pauwels, E. De Clercq, T. H. The, D. K. Meijer, "Antiviral effects of plasma and milk proteins: lactoferrin shows potent activity against both human immunodefi- ciency virus and human cytomegalovirus replication in vitro", *J. Infect. Dis.*, **172**, 380–388, (1995).

102. P. Puddu, P. Borghi, S. Gessani, P. Valenti, F. Belardelli, L. Seganti, "Antiviral effect of bovine lactoferrin saturated with metal ions on early steps of human immunodeficiency virus type 1 infection", *Int. J. Biochem. Cell Biol.*, **30**, 1055-1062, (1998).

103. P. J. Swart, M. C. Harmsen, M. E. Kuipers, A. A. Van Dijk, B. W. Van Der Strate, P. H. Van Berkel, J. H. Nuijens, C. Smit, M. Witvrouw, E. De Clercq, "Charge modification of plasma and milk proteins results in antiviral active compounds", *J. Pept. Sci.*, **5**, 563-576, (1999).

104. M. C. Harmsen, P. J. Swart, M. P. de Bethune, R. Pauwels, E. De Clercq, T. H. The, D. K. Meijer, "Antiviral effects of plasma and milk proteins: lactoferrin shows

potent activity against both human immunodefi- ciency virus and human cytomegalovirus replication in vitro", *J. Infect. Dis.*, **172**, 380-388, (1995).

105. P. Puddu, P. Borghi, S. Gessani, P. Valenti, F. Belardelli, L. Seganti, "Antiviral effect of bovine lactoferrin saturated with metal ions on early steps of human immunodeficiency virus type 1 infection", *Int. J. Biochem. Cell Biol.*, **30**, 1055-1062, (1998).

106. P. J. Swart, M. C. Harmsen, M. E. Kuipers, A. A. Van Dijk, B. W. Van Der Strate, P. H. Van Berkel, J. H. Nuijens, C. Smit, M. Witvrouw, E. De Clercq, "Charge modification of plasma and milk proteins results in antiviral active compounds", *J. Pept. Sci.*, **5**, 563-576, (1999).

107. H. Choe, K. A. Martin, M. Farzan, J. Sodroski, N. P. Gerard, C. Gerard, "Structural interactions between chemokine receptors, gp120 Env and CD4", *Semin. Immunol.*, **10**, 249-257, (1998).

108. S. Genoud, F. Kajumo, Y. Guo, D. Thompson, T. Dragic, "CCR5-Mediated human immunodeficiency virus entry depends on an amino-terminal gp120-binding site and on the conformational integrity of all four extracellular domains", *J. Virol.*, **73**, 1645-1648, (1999).

109. S.L. Kozak, S.E. Kuhmann, E.J. Platt, D. Kabat, "Roles of CD4 and coreceptors in binding, endocytosis, and proteolysis of gp120 envelope glycoproteins derived from human immunodeficiency virus type 1", *J. Biol. Chem.* **274**, 23499–23507, (1999).

110. F. Groot, T. B. Geijtenbeek, R. W. Sanders, C. E. Baldwin, M. Sanchez Hernandez, R. Floris, Y. van Kooyk, E. C. de Jong, B. Berkhout, "Lactoferrin prevents dendritic cell-mediated human immunodeficiency virus type 1 transmission by blocking the DC-SIGN – gp120 interaction", *J. Virol.*, **79**, 3009-3015, (2005).

111. M. Marchetti, F. Superti, M. G. Ammendolia, P. Rossi, P. Valenti, L. Seganti, "Inhibition of poliovirus type 1 infection by iron-, manganese and zinc-saturated lactoferrin", *Med. Microbiol. Immunol.*, **187**, 199-204, (1999).

112. C. A. Carvalho, I. P. Sousa Jr., J. L. Silva, A. C. Oliveira, R. B. Goncalves, A. M. Gomes, "Inhibition of Mayaro virus infection by bovine lactoferrin", *Virology*, **452–453**, 297–302, (2014).

113. M. Zimecki, M. Kocieba, M. Kruzel, "Immunoregulatory activities of lactoferrin in the delayed type hypersensitivity in mice are mediated by a receptor with affinity to mannose", *Immunobiology*, **205**, 120-131, (2002).

114. M. Zimecki, J. Mazurier, M. Machnicki, Z. Wieczorek, J. Montreuil, G. Spik, "Immunostimulatory activity of lactotransferrin and maturation of CD4– CD8– murine thymocytes", *Immunol. Lett.*, **30**, 119-123, (1991).

115. M. Zimecki, J. Mazurier, G. Spik, J. A. Kapp, "Human lactoferrin induces phenotypic and functional changes in murine splenic B cells", *Immunology*, **86**, 122-127, (1995).

116. H. Korhonen, "Milk-derived bioactive peptides: from science to applications", *J. Funct. Foods*, **1**,177-187.(2009).

117. M. Aito-Inoue, D. Lackeyram, M. Z. Fan, K. Sato, Y. Mine, "Transport of a tripeptide, Gly-Pro-Hyp, across the porcine intestinal brush-border membrane", *J. Pept. Sci.*, **13**, 468-474.(2007).

118. M. Heyman, J. F. Desjeux, "Significance of intestinal food protein transport", *J. Pediatr. Gastroenterol. Nutr.*, **15**, 48–57, (1992).

119. M. Shimizu, M. Tsunogai, S. Arai, "Transepithelial transport of oligopeptides in the human intestinal cell, Caco-2", *Peptides*,**18**, 681-687. (1997).

120. M. Satake, M. Enjoh, Y. Nakamura, T. Takano, Y. Kawamura, S. Arai, "Transepithelial transport of the bioactive tripeptide, Val-Pro-Pro, in human intestinal Caco-2 cell monolayers", *Biosci. Biotechnol. Biochem.*, **66**, 378-384.(2002).

121. A. Quirós, A. Dávalos, M. A. Lasunción, M. Ramos, I. Recio, "Bioavailability of the antihypertensive peptide LHLPLP: transepithelial flux of HLPLP", *Int. Dairy J.*, **18**,279-286, (2008).

122. P. Ruiz-Giménez, J. B. Salom, J. F. Marcos, S. Vallés, D. Martínez-Maqueda, I. Recio, G. Torregrosa, E. Alborch, P. Manzanares, "Antihypertensive effect of a bovine lactoferrin pepsin hydrolysate: Identification of novel active peptides", *Food Chemstry*, **131**, 266-273, (2012).

123. N. Y. Lee, J. T. Cheng, T. Enemoto, I. Nakamura, "The Antihypertensive Activity of Angiotensin-Converting Enzyme Inhibitory Peptide Containing in Bovine Lactoferrin", *Chin. J. Physiol.* **49**(2), 67-73, (2006).

124. R. Fernandes-Musoles, J. B. Salom, D. Martinez-Maqueda "Antihypertensive effects of lactoferrin hydrolyzates: Inhibition of angiotensin- and endothelin-converting enzymes", *Food Chem.,***193**, 994-1000, (2013).

125. D. Legrand, A. Pierce, E. Elass, M. Carpentier, C. Mariller, J. Mazurier, "Lactoferrin structure and functions", *Adv. Exp. Med. Biol.*, **606**, 163-194, (2008).

126. C. Liepke, K. Adermann, M. Raida, H. J. Ma¨gert, W. G. Forssmann, H. D. Zucht, "Human milk provides peptides highly stimulating the growth of bifidobacteria", *Eur. J. Biochem.*,**269**, 712-718. (2002).

127. A. K. Murthy, B. K. R. Chaganty, T. Troutman, M. N. Guentzel, J. J. Yu, S. K. Ali, C. M. Lauriano, J. P. Chambers, K. E. Klose, B. P. Arulanandam, "Mannose-containing oligosaccharides of non-specific human secretory immunoglobulin A mediate inhibition of Vibrio cholerae biofilm formation", *PLoS ONE*, **6**, e16847, (2011).

128. H. Schroten, C. Stapper, R. Plogmann, H. Ko¨hler, J. Hacker, F. G. Hanisch, "Fab-independent antiadhesion effects of secretory immunoglobulin A on S-fimbriated Escherichia coli are mediated by sialyloligosaccharides", *Infect. Immun.*, **66**, 3971-3973, (1998).

129. T. Yu , C. Guo , J. Wang , P. Hao P, S. Sui, X. Chen , R. Zhang, P. Wang, G. Yu, L. Zhang , Y. Dai , N. Li N. "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, (2011).

130. M. Barboza, J. Pinzon, S. Wickramasinghe, J. W. Froehlich, I. Moeller, J. T. Smilowitz, L. R. Ruhaak, J. Huang, B. Loennerdal, J. B. German, J. F. Medrano, B. C Weimer, C. B. Lebrilla, "Glycosilation of human milk lactoferrin exhibits dynamic changes during early lactation", *Molecular & Cellular Proteomics,* **11**,(2012).

131. C. C. Nwosu, J. S. Strum, H. J. An, C. B. Lebrilla, "Enhanced Detection and Identification of Glycopeptides in Negative Ion Mode Mass Spectrometry", *Anal Chem.*, **82**, 9654-9662, (2010).

132. C. C. Nwosu, J. Huang, D. L. Aldredge, J. S. Strum, S. Hua, R. R. Seipert, C. B. Lebrilla, "In-Gel Nonspecific Proteolysis for Elucidating Glycoproteins: A Method for Targeted Protein-Specific Glycosylation Analysis in Complex Protein Mixtures", *Anal Chem.*, **85**, 956-963, (2012).

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

134. G. Spik, G.Strecker, B. Fournet, S. Bouquelet, J. Montreuil, L. Dorland, H. van Halbeek, J. F. Vliegenthart, "Primary structure of the glycans from human lactotransferrin", *Eur. J. Biochem.,* **121**, 413-419, (1982).

135. G. Picariello, P. Ferranti, G. Mamone, P. Roepstorff, F. Addeo, "Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry", *Proteomics,* **8**, 3833-3847, (2008).

136. G. Spik, G. Strecker, B. Fournet, S. Bouquelet, J. Montreuil, L. Dorland, H. van Halbeek, J. F. Vliegenthart, "Primary structure of the glycans from human lactotransferrin", Eur. J. Biochem. / FEBS, **121**, 413-419, (1982).

137. P. H. C. vanBerkel, H. A. vanVeen, M. E. J. Geerts, H. A. deBoer, J. H. Nuijens, "Heterogeneity in utilization of N-glycosylation sites Asn(624) and Asn(138) in human lactoferrin: A study with glycosylation-site mutants", *Biochemical Journal,* **319**, 117-122, (1996).

138 . C. C. Nwosu, D. L. Aldredge, H. Lee, L. A. Lerno, A. M. Zivkovic, J. B. German, C. B. Lebrilla, "Comparison of the Human and Bovine Milk N-Glycome via High-Performance Microfluidic Chip Liquid Chromatography and Tandem Mass Spectrometry", *J. Proteome Res*., **11**, 2912-2924, (2012).

139 . R. Apweiler, H. Hermjakob, N. Sharon, "On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database", *Biochim. Biophys. Acta,* **1473**, 4-8, (1999).

140. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler, *Essentials of glycobiology*, 2nd edn. Cold Spring Harbor, New York (2010).

141 . R. Apweiler, H. Hermjakob, N. Sharon, "On the frequency of protein glycosylation, as deduced from analysis of the swiss-prot database", *Biochim. Biophys. Acta.*, **1473**, 4-8, (1999).

142. N. Taniguchi, Human disease glycomics/ proteome initiative (HGPI). *Mol Cell Proteomics*,**7**, 626-627, (2008).

143. R. D. Marshall, "Glycoproteins", *Annu. Rev. Biochem.*, **41**, 673-702, (1972).

144. R. Kornfeld, S. Kornfeld, "Assembly of asparagine-linked oligosaccharides" *Annu. Rev. Biochem.*, **54**, 631-664, (1985).

145. C. Kirmiz, B. Li, H. J. An, B. H. Clowers, H. K. Chew, K. S. Lam, A. Ferrige, R. Alecio, A. D. Borowsky, S.Sulaimon, C. B. Lebrilla, S. Miyamoto, "A serum glycomics approach to breast cancer biomarker", *Mol. Cell. Proteomics*, **6**, 43-55, (2007).

146. G. S. Leiserowitz, C. Lebrilla, S. Miyamoto, H. J. An, H. Duong, C. Kirmiz, B. Ll, H. Liu, K. S. Lam, *"Glycomics analysis of serum: a potential new biomarker for ovarian cancer?"*, *Int. J. Gynecol. Cancer*, **18**, 470-475 (2008).

147. Z. Kyselova, *Y.* Mechref, P. Kang, J.A. Goetz, L. E. Dobrolecki, G. W. Sledge, L. Schnaper, R. J. Hickey, J. Robert, L. H. Malkas, M. V. Novotny, *"Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. Clin. Chem.* **54**, 1166–75 (2008).

148. J. Goetz, Y. Mechref, P. Kang, M. Jeng, M. V. "Glycomic profiling of invasive and non-invasive breast cancer cells", *Glycoconj. J.* **26**, 117-31 (2009).

149. K. S. Nandakumar, M. Collin, A. Olsen, F. Nimmerjahn, A. M. Blom, J. V. Ravetch, R. Holmdahl, "Endoglycosidase treatment abrogates IgG arthritogenicity: importance of IgG glycosylation in arthritis", *Eur. J. Immunol.* **37,** 2973-2982 (2007).

150. P. Van den Steen, P. M. Rudd, R. Dwek, G. Opdenakker, "Concepts and principles of O-linked glycosylation". *Crit. Rev. Biochem. Mol. Biol.*, **33**, 151-208 (1998).

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

152. H. Geyer, R. Geyer, "Strategies for analysis of glycoprotein glycosylation", *Biochim. Biophys. Acta, Proteins Proteomics*, **1764**, 1853-1869, (2006).

153. A. J. Alpert, "Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds", *J. Chromatogr. A*, **499**, 177-196 (1990).

154. P. Hemstrom, K. Irgum, "Hydrophilic interaction chromatography", *J. Sep. Sci.*, **29**, 1784-1821, (2006).

155. M. Gilar, P. Olivova, A. E. Daly, J. C. Gebler, "Orthogonality of separation in two-dimensional liquid chromatography", *Anal. Chem.*, **77**, 6426-6434. (2005).

156. P. Jandera, "Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts", *J. Sep. Sci.*, **31**, 1421-1437. (2008).

157. P. Jandera, "Stationary and mobile phases in hydrophilic interaction chromatography: a review", *Anal. Chim. Acta* , **692**, 1-25,(2011).

158. P. Jandera, T. Hajek, , V. Skerıkova, J. Soukup, "Dual hydrophilic interaction-RP retention mechanism on polar columns: Structural correlations and implementation for 2-D separations on a single column", *J. Sep. Sci.*, **33**, 841-852, (2010).

159. Y. Kawachi, T. Ikegami, H. Takubo, Y. Ikegami, M. Miyamoto, N. Tanaka, "Chromatographic characterization of hydrophilic interaction liquid chromatography stationary phases: Hydrophilicity, charge effects, structural selectivity, and separation efficiency", *J. Chromatogr. A*, **1218**, 5903-5919, (2011).

160. T. Ikegami, K. Tomomatsu, H. Takubo, K. Horie, N. Tanaka, "Separation efficiencies in hydrophilic interaction chromatography", *J. Chromatogr. A*, **1184**, 474-503, (2008)

161. Y. Guo, S. Gaiki, "Retention and selectivity of stationary phases for hydrophilic interaction chromatography", *J. Chromatogr. A*, **1218**, 5920-5938, (2011).

162. R. I. Chirita, C. West, S. Zubrzycki, A. L. Finaru, C. Elfakir, "Investigations on the chromatographic behaviour of zwitterionic stationary phases used in hydrophilic interaction chromatography", *J. Chromatogr. A*, **1218**, 5939-5963, (2011).

163. A. J. Alpert, "Hydrophilic Interaction Chromatography", *J. Chromatogr. A*, **1218**, 5879, (2011).

164. L. R. Snyder, J. J. Kirkland, J. W. Dolan, "Introduction to Modern Liquid Chromatography", Wiley, Hoboken, NJ, (2010).

165. M. W. H. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms, A. J. R. Heck, "Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns", *Anal. Chem.*, **76**, 3935-3943, (2004).

166. M.R. Larsen, S. S. Jensen, L. A. Jakobsen, N. H. H. Heegaard "Exploring the sialiome using titanium dioxide chromatography and mass spectrometry", *Mol Cell Proteomics*, **6**(10), 1778-1787,(2007).

167. T. Miyagi, K. Yamaguchi, "Mammalian sialidases: Physiological and pathological roles in cellular functions", *Glycobiology*, **22**(7), 880-896, (2012).

168. B.L. Parker, P. Gupta, S. J. Cordwell, M. R. Larsen, G.Palmisano, "Purification and Identification of O-GlcNAc-Modified Peptides Using Phosphate-Based Alkyne

CLICK Chemistry in Combination with Titanium Dioxide Chromatography and Mass Spectrometry", *Journal of Proteome Research*, **10**(4), 1449-1458, (2011).

169. B. Zhang, Q. Sheng, X. Li, Q. Liang, J. X. Liang, "Selective enrichment of glycopeptides for mass spectrometry analysis using C18 fractionation and titanium dioxide chromatography", *J. Sep. Sci.* **34**(19), 2745-2750, (2011).

170. G. Palmisano, D. Antonacci, M.R. Larsen, "Glycoproteomic Profile in Wine: A 'Sweet' Molecular Renaissance", *J. Proteome Res.*, **9**(12), 6148-6159, (2010).

171. Y. L.Hsieh, T. H. Chen, C.Y. Liu, "Capillary electrochromatographic separation of proteins on a column coated with titanium dioxide nanoparticles", *Electrophoresis*, **27**(21), 4288-4294, (2006).

172. M.W.H. Pinkse, S. Mohammed, L. W. Gouw, B. van Breukelen, R. Vos, Harmjan, A. J. R. Heck, "Highly robust, automated, and sensitive online tio$_2$-based phosphoproteomics applied to study endogenous phosphorylation in drosophila melanogaster", *J. Proteome Res.*, **7**(2), 687-697, (2007).

173. Y.Y. Zeng, H. J. Chen, K. J. Shiau, S. U. Hung, Y. S. Wang, C. C. Wu, "Efficient enrichment of phosphopeptides by magnetic TiO2-coated carbonen capsulated iron nanoparticles" *Proteomics*, **12**(3), 380-390, (2012).

174. M. Karas, D. Bachmann, F. Hillenkamp, "Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules", *Anal. Chem.* **57**, 2935-2939 (1985).

175. K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, "Protein and Polymer Analyses up to mlz 100000 by Laser Ionization Time-of-flight Mass Spectrometry", *Rapid Commun. Mass Spectrom.* **2,** 151-153 (1988).

176. K. Dreisewerd "The desorption process in MALDI", *Chemical reviews*, **103,** 395-426, (2003).

177. R. D. Smith, K. J. Light-Wahl, "The observation of non-covalent interactions in solution by ESI-MS: promise, pitfall and prognosis". *Biol. Mass Spectrom.*, **22**, 493, (1993).

178. F. W. Rollgen, U. Luttgens, T. Dulcks, U. Giessmann, " Electrospray mass spectrometry of biomacromolecular complexes with non covalent interactions: new analytical perspectives for supramolecular chemistry and molecular recognition processes". *Conf. Am. Soc. Mass Spectrom.*, **35** (8), 806-826, (1993).

179. J.B. Fenn, "Ion formation from charged droplets: roles of geometry, energy, and time", *J. Am. Soc. Mass Spectrom.*,**4**, 524, (1993).

180. L. Sleno, D. Volmer, "Ion activation methods for tandem mass spectrometry", *J. Mass Spectrom.* **39**, 1091–112 (2004).

181. P. Roepstorff, "Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides", *Biom. mass spectrum.*,**11**, 601,(1984).

182. M. Wuhrer, M. I. Catalina, A. M. Deelder, C. H. Hokke, "Glycoproteomics based on tandem mass spectrometry of glycopeptides", *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.*, **849**, 115-128 (2007).

183 . B. Domon, C. E. Costello, "A Systematic Nomenclature for Carbohydrate Fragmentation in Fab-Ms Ms Spectra of Glycoconjugate" *Glycoconj. J.* **5**, 397-409, (1988).

184. K. Gevaert, H. Demol, L. Martens, B. Hoorelbeke, M. Puype, M. Goethals, J. V. Damme, S. D. Boeck, J. Vandekerckhove, "Protein identification based on matrix assisted laser desorption/ionizationpost source decay-mass spectrometry", *Electrophoresis* ,**22**, 1645, (2001).

185. J. Gobom, E. Nordhoff, E. Mirgorodskaya, R. Ekman, P. Roepstorff, "Sample purification and preparation technique based on nanoscale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry", *J. Mass Spectrom.* **34**, 105, **(1999).**

186.P. Hagglund, J. Bunkenborg, F. Elortza, O. N. Jensen, P. Roepsorff, "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**, 556, (2004)*.*

187. MassAI-Bioinfornatics. MassAI.at<http://massai.dk/index.html>.

188 . L. N. Marekov, P. M. Steinert, "Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", *J Mass Spectrom* **38**(4), 373-377, (2003).

189 D. X. Wang, S. R. Kalb, R. J. "Cotter Improved procedures for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing", *Rapid Commun Mass Spectrom* **18**(1), 96-102, (2004).

190. N. E. Robinson, "Protein deamidation", *Proc Natl Acad Sci U S A*, **99**(8), 5283-5288, (2002).

191. H. T. "Wright Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins", *Crit Rev Biochem Mol Biol* **26**(1), 1-52, (1991).

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