Journal of Chromatography A, 1227 (2012) 219–233

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Gel-free shotgun proteomic analysis of human milk

Gianluca Picariello[a,*], Pasquale Ferranti[a,b], Gianfranco Mamone[a], Iveta Klouckova[c], Yehia Mechref[c], Milos V. Novotny[c], Francesco Addeo[a,b]

[a] Istituto di Scienze dell'Alimentazione – Consiglio Nazionale delle Ricerche (CNR), Via Roma 64, 83100 Avellino, Italy
[b] Dipartimento di Scienza degli Alimenti, University of Naples "Federico II," Parco Gussone, Portici (NA) 80055, Italy
[c] National Center for Glycomics and Glycoproteomics, Department of Chemistry, Indiana University, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:
Received 1 December 2011
Received in revised form 2 January 2012
Accepted 3 January 2012
Available online 11 January 2012

Keywords:
Human milk
Milk fat globule membrane proteins
Shotgun proteomics
Fourier Transform-Ion Cyclotron Resonance mass spectrometry
Host defense proteins
Lipid droplet proteins

ABSTRACT

The composition of milk has adapted during the evolution of the species to fulfill the specific nutritional needs of the offspring. Currently, it is widely recognized that milk benefits go beyond mere nutrition and serve as a source of a number of functional components to the newborn, particularly host defense effectors. However, the human milk proteome description is still incomplete, primarily because the detection of low-abundance proteins remains challenging. To overcome the limitations of the classical electrophoresis-based approach, previously separated milk fat globule membrane (MFGM) and whey protein fractions were analyzed by nanoflow-high performance liquid chromatography (HPLC)/Fourier Transform-Ion Cyclotron Resonance (FT-ICR) mass spectrometry (MS). This shotgun strategy showed an as yet unmatched potential to profile low-abundance proteins in human milk. Proteins associated with 301 different gene products were identified, some of which could be clustered into subsets of protein isoforms, thus providing one of the largest protein inventories of human milk. The identified proteins, which were derived from multiple metabolic pathways, are involved in different physiological functions, such as membrane trafficking, cell signaling, fat metabolism and transport, metabolite delivery, protein synthesis/proteolysis or folding, and immunity-related actions. Nevertheless, it appears clear from this study that the overall picture of the human milk proteome is still incomplete, although several protein signatures of milk evolution are emerging.

1. Introduction

Milk is a complex biological fluid because it contains a wide range of macro- and micro-nutrients. It has been estimated that the entire mammary gland machinery includes several thousand genes whose evolution can be traced back to more than 300 million years [1]. The astounding preservation of the protein complexity along mammalian phylogenesis signifies the essential role of breast milk for term, pre-term and small-for-date human infants, which goes beyond mere nutritional aspects. Therefore, a basic understanding of the functions accomplished by breast milk, either contingently or in terms of "imprinting", has become a central theme in nutrition research. Many investigations have attempted to comprehensively characterize the proteins in human colostrum and mature milk.

Human milk proteins can be grouped into three major classes: caseins; milk fat globule membrane (MFGM) proteins and predominant whey proteins. Whey contains several low-abundance blood serum-derived proteins in addition to those secreted by neutrophils, by lymphocytes that have infiltrated the mammary gland and by somatic cells. Similar to other biological fluids, the wide range of protein concentrations in the human milk proteome, thought to span eight to ten orders of magnitude, presents a major challenge to systematic cartography. In fact, six or seven polypeptide chains constitute more than 90% of the human milk protein content, and the remaining content is distributed throughout hundreds of cellular proteins [2].

An early proteomic study based on two-dimensional electrophoresis (2DE) and microsequencing identified 22 proteins in human colostrum [3] although as many as 400 spots were observed. Later, through the "classical" 2DE-mass spectrometry (MS) proteomic approach, 107 protein spots, corresponding to 39 gene products, were identified in the human MFGM fraction [4,5]. Indeed, the index of proteins in human milk was increased to 73, including polypeptides identified in the whey fraction [6,7]. Many strategies have been developed for the selective removal of albumin and other highly abundant proteins from human and bovine milk (or colostrum). Nevertheless, only 15 proteins were identified in bovine mature milk and colostrum by 2DE and microsequencing after immunodepletion [8]. Immunoadsorption of the major proteins followed by mono-dimensional (1D) electrophoresis and

* Corresponding author. Tel.: +39 0825 299111; fax: +39 0825 782599.
E-mail address: picariello@isa.cnr.it (G. Picariello).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2012.01.014

MS identified 151 gene products in human whey colostrum, 83 of which were reported for the first time [9]. Using SDS-PAGE coupled with MS, 120 non-redundant gene products were identified in bovine MFGM protein fractions [10], while 95 and 72 distinct gene products were identified in bovine [11] and water buffalo milk [12], respectively, by liquid chromatography (LC)–MS and 2DE-MS. The non-linear compression of the dynamic range via combinatorial peptide ligand libraries (ProteoMiner), though recently debated [13,14], has been definitively demonstrated to be effective in improving the coverage of complex proteomes in shotgun experiments [15]. Thus, equalization through peptide libraries was successful in cataloguing 149 unique proteins in bovine milk [16] and 115 in human whey [17].

Merging together the results of past studies, a total of 285 non-redundant database entries were catalogued in the inventory of the human milk proteome, including a "core" of 106 proteins that are shared with the bovine milk proteome [18].

The quasi-totality of studies aimed at characterizing proteomes at different depths (deep vs. shallow) relies upon an electrophoretic separation step prior to MS analysis. Conventional 2DE for descriptive proteomics suffers from a number of limitations that are particularly critical for characterizing complex mixtures. For example, very low-abundance proteins, as well as those with extreme isoelectric points or molecular weight values, can escape detection on the gels. In spite of the recent advances in 2DE, the detection and identification of very hydrophobic membrane proteins, such as MFGM proteins, is affected by their solubility [19]. The protein chip array of the SELDI-TOF (surface enhanced laser desorption ionization-time of flight) MS approach was also unsatisfactory to resolve the complexity of the human milk proteome, with particular concern for the MFGM protein fraction [2]. The well-known drawbacks of the 2DE-MS approach have prompted the development of gel-free strategies for high-throughput proteomics, especially when membrane proteins are to be analyzed [20,21]. In particular the use of high resolution capillary chromatographic columns coupled to ultra high mass accuracy (less than 1 ppm error tolerance) LTQ-Fourier Transform-Ion Cyclotron Resonance (FT-ICR) MS was shown to be capable of identifying thousands of peptide components in a single 1D high-resolution chromatographic separation [22]. With this technique, as a general procedure, an entire protein extract is digested with trypsin before nanoflow LC-electrospray (ESI) and MS analysis. In the present work, a gel-free shotgun approach has been exploited to penetrate deeper into the human milk proteome. Very recently, with a UPLC-Orbitrap-based shotgun MS approach, targeted to identify host-defense proteins, 268 and 269 gene products were identified in human and bovine milk, respectively. These results confirm that gel-free techniques are more appropriate for analyzing complex proteomes than the classic electrophoresis-based proteomic approaches [23]. Whey and MFGM proteins were separately trypsinized and analyzed by nanoflow reversed-phase HPLC coupled on-line with high-resolution FT-ICR MS. Validation of data sets was performed with automated bioinformatic tools.

2. Materials and methods

2.1. Sample preparation

Human milk samples were collected from five healthy Italian women by manual expression in a sterile tube, 10–15 days after full-term labor, and pooled (5 mL each). To prevent undesired proteolysis, PMSF was added to a final concentration of 1 mM. Milk freeze–thaw cycles were avoided to prevent breakdown of the fat globule membrane; samples were transferred to the laboratory in a refrigerated box and utilized for protein extraction within 2 h. Milk was gently shaken, and aliquots of 15 mL were centrifuged at 3500 rpm for 20 min at 4 °C to separate the fat, which was manually removed. MFGM proteins were extracted as described by Fortunato et al. [4]. The scarce pellet that resulted from the milk centrifugation, containing somatic cells and large casein micelles, was discarded. The skim milk fraction that here is called "milk whey" still contained a significant amount of casein. No attempt was carried out to completely remove casein by isoelectric precipitation in order to avoid the depletion of possible hydrophobically interacting minor proteins. Milk whey was filtered through Millex sterile 0.22 μm membranes (Millipore, Bedford, MA, USA) and precipitated by adding 4 volumes of −20 °C cold acetone/10% (w/v) trichloroacetic acid (TCA) containing 2% 2-mercaptoethanol. The protein pellet was washed three times with −20 °C cold acetone, separated by centrifugation (4 °C, 20 min, 12,000 rpm) and lyophilized.

2.2. Reduction and alkylation of MFGM and milk whey proteins

Disulphide bridges were reduced by incubating approximately 300 μg of both human MFGM and milk whey proteins in 1 mL of a denaturing/reducing buffer (6 M guanidine HCl, 100 mM Tris–HCl, 1 mM EDTA, 10 mM DTT, pH 8.0) for 1 h at 56 °C. Cysteine residues were alkylated by incubation for 30 min with iodoacetamide at a final concentration of 55 mM at room temperature in the dark. To quench the alkylation reaction, proteins were immediately desalted by PD-10 columns using 50 mM ammonium bicarbonate (AMBIC), pH 8.5, as the eluent. Proteins were quantified by the Bradford assay using BSA as the standard and incubated overnight at 37 °C with modified proteomic grade trypsin (Promega, Madison, WI, USA) using a 50:1 (w/w) protein-to-trypsin ratio. The reaction was stopped by freeze-drying. To prevent the missed detection of peptides due to phosphorylation, digests were re-dissolved in 0.5 mL of AMBIC at pH 8.5 and dephosphorylated overnight at 37 °C with 5 enzymatic units of alkaline phosphatase (Roche, Basel, Switzerland). Finally, to remove AMBIC, peptide solutions were repeatedly lyophilized after the addition of a few drops of aqueous 0.1% TFA. Alkaline phosphatase was discarded from the list of identified proteins in human milk.

To enlarge the proteome coverage, aliquots of the peptides derived from MFGM and milk whey proteins were subjected to N-deglycosylation (18 h, 37 °C, 2 mU/100 μg, in AMBIC) by PNGase F (Roche) and analyzed separately.

2.3. LC–MS/MS analysis

Tryptic digests were separated using an UltiMate 3000 nano-LC system (Dionex, Sunnydale, CA, USA). Peptides derived from 3 μg of the protein fractions were loaded by the autosampler on a Pep-Map300 C18 trapping cartridge (5 μm, 200 Å, Dionex). The nanocolumn used for nano-LC-ESI/MS/MS runs was prepared by packing a slurry of a 90 Å Jupiter C12 bound phase (Phenomenex, Torrance, CA, USA) into a 150 mm × 75 μm i.d. pulled-tip fused silica (Polymicro Technologies, Phoenix, AZ) using pressurized air. The column was washed with 80% acetonitrile mobile phase containing 0.1% formic acid and subsequently equilibrated with 3% acetonitrile with 0.1% formic acid at a 250 nL/min flow rate.

Solvent A was 3% acetonitrile in HPLC-grade water containing 0.1% formic acid, while solvent B was HPLC-grade acetonitrile containing 0.1% formic acid. The separation was carried out at a flow rate of 250 nL/min with a linear gradient from 3% to 55% of solution B over 150 min following 10 min of isocratic elution at 3% of solution B.

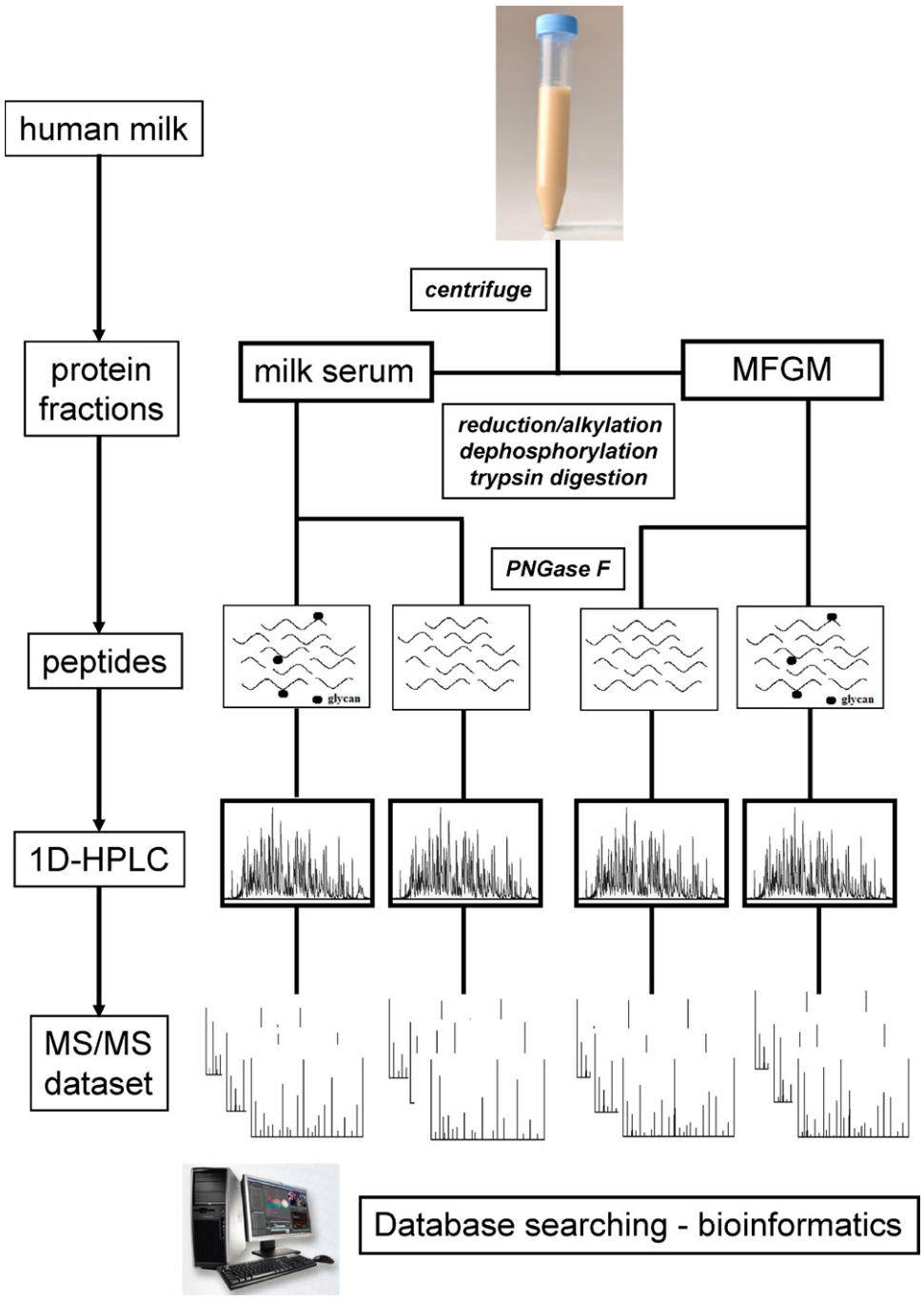

Fig. 1. Schematic overview of the shotgun proteomic strategy adopted for the analysis of human milk.

The nanoLC capillary column was directly interfaced with the electrospray source of an LTQ-FT mass spectrometer (Thermo-Electron, San Jose, CA, USA) operating in the positive-ion mode. The instruments recorded full FT mass spectra (m/z 250–2500). The precursor ions were isolated using the data-dependent acquisition mode. The 5 most intense ions in the survey scan were selected for MS/MS and fragmented in the instrument ion trap at a normalized collision energy of 35%. A dynamic exclusion window of 30 s was set up to prevent dominant precursor ions from being selected more than once for CID, thus enabling the fragmentation of less abundant precursors. The Q activation values for all MS^n were set to 0.250, while the activation times were set to 75 ms. Under such conditions a MS mass accuracy of <5 ppm was typically achieved. Each sample was run in triplicate.

2.4. Database searching and protein identification

Raw files generated by the XCalibur software v2.0.7 (Thermo Scientific) were used to generate peak lists in mascot generic files (.mgf) for the Uniprot database searching with a licensed copy of MASCOT (Matrix Science, Boston, MA, USA). Search results were filtered with a ProteinParser according to the following parameters: one tryptic missed cleavage allowed, ±15 ppm tolerance for precursor ions, +2 and +3 charges, carbamidomethylation of cysteines as a fixed modification, oxidation of methionine and, in the analysis of deglycosylated peptides, also deamidation of Asn as variable modifications, ion score ≥30 and a minimum peptide mass of 600.00 Da. The .mgf files were searched against a randomized version of the UniProt database that was previously generated using the decoy

database generator utility available from Matrix Science and filtered using the same parameters as described earlier for the regular searches. In these conditions, the false discovery rate was estimated below 0.1%. Only peptide matches occurring in all the three analyses of each sample were accepted.

3. Results and discussion

3.1. Protein identification

Typically, the protein content of mature human milk ranges between 10 and 20 mg/mL and reaches a minimum value after 1 week of lactation. Although the concentration of each protein varies as a function of several factors during lactation [24], a recent proteomic study has demonstrated a strict qualitative consistency between the proteomes of bovine colostrum and mature milk [25]. Thus, the descriptive proteome of human milk should also not significantly change during lactation.

Similar to previous studies, sample pre-fractionation appeared to be a key issue in the deep investigation of the milk proteome. Therefore, the MFGM fraction, which is primarily composed of hydrophobic membrane or membrane-associated proteins, was analyzed separately from the whey proteins. Because the extractability and solubility of the MFGM proteins is critical [26], a denaturing/reducing buffer was used to dissolve extracted MFGM proteins. In our experimental plan, the possible loss of low-abundance proteins was minimized by avoiding protein enrichment or depletion processes, as schematized in Fig. 1. The identification of MFGM and whey proteins, also including caseins, has been accomplished by coupling high resolution nano-flow reversed phase chromatography separation to high resolution LTQ-FT-ICR MS and MS/MS analysis of tryptic peptides. Although the analytical complexity of the fractions is greatly increased by proteolysis, the multiplication of protein-derived moieties enhances the probability of detecting and confidently identifying at least one peptide of the mother protein. In general, MS exhibits higher sensitivity in the detection of peptides compared to proteins. Thus, peptide-centric methodologies, also referred to as "shotgun" or "bottom-up" proteomics, combined with the bioinformatic tools, have been demonstrated to enhance proteome coverage and achieve larger access compared to techniques for targeting full-length proteins [27].

Due to the high levels of plasminogen activators, plasmin can hydrolyze up to 50–60% of the human milk proteins, especially the casein component [28]. Proteolysis represents a serious drawback of the electrophoresis-based analysis of human milk proteins. Conversely, the gel-free "shotgun" approach is only marginally affected by protein hydrolysis, although the information about the integrity of the native proteins and the possibility of distinguishing isoforms are partially lost [29].

We unambiguously identified 301 proteins, including 165 exclusively from the MFGM fraction. Indeed, numerous related proteins that contain similar peptides, such as the Ig κ or Ig λ isoforms, could not be differentiated based on MS/MS analysis alone and should be grouped together to satisfy the principles of parsimony. A complete list of all of the identified proteins in human whey and MFGM fractions is shown in Table 1. Several proteins have not been described before in human colostrum or milk, although the exact number of these proteins is undefined because of the multiplicity of previously published proteome data, which also include many single protein-focused studies. The annotation of human milk proteins has been upgraded even very recently [30]. It should also be noted that many novel proteins have homologous cognates already described in the milk of other species including the cow, buffalo, and tammar wallaby.

Because many milk proteins are phosphorylated and/or glycosylated, peptide mass dispersion was reduced by dephosphorylation before tryptic digestion. There were 26 and 21 gene products, including MUC5B and plasma membrane SYTL1, CD97 and BSCL2, which were identified exclusively after deglycosylation in the whey and MFGM protein fractions, respectively. However, some of the proteins identified after deglycosylation were not N-glycosylated, and the protein coverage was only occasionally increased by the identification of formerly N-glycosylated peptides. Some selected MASCOT-based identifications of formerly N-glycosylated peptides, deamidated by PNGase F, are reported in the Supplementary Data.

An estimate of protein abundance is provided by the relative score value, although this is only a rough indication because an accurate normalization of the matching to the expected tryptic peptides would be required for a reliable label-free protein quantification; this normalization would be somewhat biased by proteolysis. For example, LALBA and caseins, which are expected to be the highest represented proteins, did not achieve the highest score values, and this was likely due to their reduced size and the low number of expected tryptic peptides. LTF had the highest identification score, thus confirming its relatively high levels in human milk. Interestingly, tenascin-C, whose occurrence in milk and colostrum now well established [9,31,32] has been long-neglected, scored among the highest represented proteins in human milk. The specific role of this protein remains to be investigated.

Several entries were single-peptide identifications. However, the high accuracy of the ESI-LTQ-FT-ICR peptide mass measurements [29], the tests with a decoy database generator and the filtering operation that restricted acceptance only to those peptides occurring in three repeated analyses, minimized the false discovery rate at a threshold as low as 0.1%. Many of the entries identified through a single peptide are proteins that have already been described in human and/or bovine milk, which validates their inclusion in Table 1. Exemplificative single-peptide protein identifications supported by MASCOT and manually validated are shown in Supplementary Data.

3.2. Origin of human milk proteins

In addition to mammary-cell specific proteins, milk contains variable levels of blood-derived proteins. The identified proteins, primarily derived from the extracellular region, cytoplasm, intracellular organelles and inner or outer cell membranes, reflect their extracellular or intracellular origin. In agreement with previous findings, cytoplasm proteins also contributed to the MFGM proteome; this contribution is likely due to a tight association with the membrane proteins or a contamination of the MFGM preparations with cytoplasm proteins in the milk fat [10,33]. However, some proteins putatively involved in intracellular transport of solutes or lipid droplet proteins can have multiple locations and interact with several organelles or membrane compartments.

MFGM proteins are the most conserved across species, suggesting that the cellular anatomy of the secreting structures is maintained throughout evolution. It has been hypothesized that the complex ontogeny of the mammary gland as secretory organ, which dates back as far as 200 million years, may have occurred by co-opting existing structures and developmental pathways [1]. However, proteins with nutritional and immunological functions might be divergent to some extent.

3.3. Biological function of human milk proteins

The main recognized natural function of the most abundant milk proteins is supplying necessary nutrients to newborns. Caseins

Table 1
Shotgun proteomic identification of proteins in human milk serum (hMS) and MFGM fractions. Score values are reported along with the number of identified peptides (*), main cellular localization and supposed function. In the current experimental conditions, matches with score values > 30 were considered proof of identity or extensive homology ($p < 0.01$). A number of entries were single-peptide assignments. However, a great number of the entries were proteins already known to occur in milk, thus validating their identification. Symbols for cellular localization and function have been adapted from Ref. [18]: A, angiogenesis and blood functionality; B, bone maturation; CY, cytoplasm; D, cell differentiation; ER, endoplasmic reticulum; ES, extracellular space; G, cell growth/tissue development; I, defense/immunity; L, lipid droplet transport/secretion; MI, mitochondrial; N, synthesis and delivery of nutrients/metabolites, energy pathways; NU, nucleus; P, protein metabolism (protease/peptidase or inhibitor); PM, plasma membrane; PMa, plasma membrane associated; R, cellular recognition; S, signal transduction; T, cellular trafficking; U, unknown; V, nervous system maturation; Z, inflammation.

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
LTF	Lactotransferrin	P02788	ES	G, I, N	2917 (47)	4008 (75)	2367 (34)	2634 (47)
PIGR	Polymeric Ig-receptor	P01833	PM	I, L	1227 (21)	2469 (43)	571 (11)	852 (17)
TNC	Tenascin-C precursor	P24821	ES	D, G, V	2027 (40)	2188 (45)		
MRC1	Macrophage mannose receptor1 (CD206 antigen)	P22897	PM	I, Z	1487 (25)	1772 (35)	44 (1)	
C3	Complement C3	P01024	ES	I	1480 (26)	1286 (28)		
CO4B	Complement C4-B	P0C0L5	CY	I	1145 (18)	974 (14)	77 (1)	213 (2)
BSSL	Bile salt-stimulated lipase	P19835	ES	L, N	1069 (19)	1556 (26)	307 (5)	489 (8)
XDH	Xanthine dehydrogenase	P47989	CY, PMa	I, L	1013 (27)	1300 (29)	3256 (52)	3797 (74)
CO4A	Complement C4-A	P0C0L4	CY	I	975 (21)	834 (14)	186 (3)	217 (2)
MFGE8	Lactadherin	Q08431	PM, ES	I, L, R	721 (13)	1005 (23)	1225 (21)	2023 (36)
AACT	Alpha-1-antichymotrypsin	P01011	ES	I, N, P, Z	639 (12)	773 (14)	258 (3)	194 (3)
AAT	Alpha-1-antitrypsin	P01009	ES	I, N, P, Z	250 (4)	441 (6)		
BTN1A1	Butyrophilin subfamily 1 member A1	Q13410	PM	L	537 (13)	878 (17)	1288 (26)	2292 (40)
CSN1S1	Alpha-s1-casein	P47710	ES	B, I, N	361 (5)	575 (12)	51 (2)	93 (1)
CSN2	Beta-casein	P05814	ES	B, I, N	359 (6)	710 (18)	121 (3)	232 (6)
CSN3	Kappa-casein	P07498	ES	B, I, N	239 (5)	495 (12)	184 (5)	135 (4)
LALBA	Alpha-lactalbumin	P00709	ES	I, N	341 (7)	367 (7)	172 (3)	252 (4)
ALB	Serum Albumin	P02768	ES	L, N	533 (14)	358 (10)	257 (5)	183 (4)
CD14	Monocyte differentiation antigen CD14	P08571	PM	I, L	513 (7)	402 (8)	330 (4)	272 (4)
B4GALT1	Beta-1,4-galactosyltransferase 1	P15291	CY	N	495 (9)	354 (8)		
IGHA1	Ig alpha-1 chain C region	P01876	ES	I	875 (15)	1551 (31)	446 (7)	542 (9)
IGHM	Ig mu chain C	P01871	ES	I	1003 (15)	605 (13)		
IGHC1	Ig gamma-1 chain C	P01857	ES	I	489 (9)	296 (11)		
IGKC	Ig kappa chain C	P01834	ES	I	432 (6)	676 (10)	147 (4)	236 (6)
IGHA2	Ig alpha-2 chain C	P01877	ES	I	329 (4)	253 (7)		
IGJ	Ig J chain	P01591	ES	I	255 (4)	297 (5)		
IGLC1	Ig lambda chain C	P01842	ES	I	208 (5)	221 (6)	165 (2)	95 (2)
IGHC2	Ig gamma-2 chain C	P01859	ES	I	117 (3)	60 (2)		
IGHC3	Ig gamma-3 chain C	P01860	ES	I	43 (2)	51 (1)		
IGHC4	Ig gamma-4 chain C	P01861	ES	I	41 (1)			
	Ig kappa chain V-III region SIE	P01620	ES	I	241 (3)	202 (2)		
	Ig mu heavy chain disease protein	P04220	ES	I	65 (1)	76 (1)		
	Ig kappa chain V-I region Wes	P01611	ES	I		89 (1)		

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
	Ig kappa chain V-II region MIL	P01616	ES	I		88 (1)		
	Ig kappa chain V-IV region	P01625	ES	I	219 (3)	89 (2)		
	Ig kappa chain V-I region GAL	P01599	ES	I	137 (1)			
	Ig kappa chain V-I region DEE	P01597	ES	I	132 (2)			
	Ig kappa chain V-I region CAR	P01596	ES	I	123 (1)	131 (1)		
	Ig kappa chain V-I region AG	P01593	ES	I	126 (1)			
	Ig kappa chain V-III region Ti	P01622	ES	I		164 (2)		
	Ig kappa chain V-III region GAL	P01781	ES	I	42 (1)			
	Ig kappa chain V-III region B6	P01619	ES	I		162 (2)		
	Ig kappa chain V-II region MIL	P01616	ES	I	86 (1)			
	Ig kappa chain V-I DEE	P01597	ES	I	132 (2)	117 (1)		
	Ig kappa chain V-III region HIC	P18136	ES	I	40 (1)			
	Ig kappa chain V-II region RPMI 6410	P06310	ES	I	75 (1)			
	Ig kappa chain V-I region BAN	P04430	ES	I	65 (1)	44 (1)		
	Ig kappa chain V-III region WOL	P01623	ES	I		206 (2)		
	Ig lambda chain V-III region	P80748	ES	I	177 (2)	158 (2)		
	Ig lamba chain V-III region	P01714	ES	I	91 (1)	96 (1)		
	Ig lamba chain V-I region HA	P01700	ES	I	59 (1)			
	Ig lambda chain V-VI region AR	P01721	ES	I	36 (1)			
	Ig lamba chain V-IV	P01717	ES	I	46 (1)	82 (2)		
	Ig heavy chain V-III region TEI	P01777	ES	I	119 (1)			
	Ig heavy chain V-III	P01766	ES	I	140 (1)	139 (3)		
	Ig heavy chain V-III region WAS	P01776	ES	I	120 (1)	Y		
	Ig heavy chain V-II region NEWM	P01825	ES	I	81 (1)	73 (1)		
	Ig heavy chain V-III region VH26	P01764	ES	I	51 (1)			
	Ig heavy chain V-III region CAM	P01768	ES	I	43 (1)			
CLU	Clusterin, apolipopretein J	P10909	ES	I, L, S	429 (7)	885 (13)	365 (7)	859 (15)
AZGP1	Zinc-alpha-2-glycoprotein	P25311	ES	L, S	428 (8)	261 (8)		
LGALS3BP	Galectin-3-binding protein	Q08380	PM	I, S	394 (6)	438 (8)		
ACTB	Beta-Actin, cytoplasmic 1	P60709	CY	L	328 (7)	94 (1)	421 (7)	237 (5)
ACTN1	Alpha-actin 1	P12814	CY	L	43 (1)	96 (1)		
ACTN2	Alpha-actin 2	P35609	CY	L		133 (2)		
ACTC1	Actin alpha-cardiac	P68032	CY	L	197 (3)			
ACTG1	Gamma-actin cytoplasmic2	P63261	CY	L	87 (2)		325 (7)	391 (6)
TF	Seroransferrin	P02787	ES	I, N	319 (8)	759 (10)		
CHRDL2	Chordin-like protein 2	Q6WN34	ES	B, D, G	314 (5)	190 (4)	131 (3)	48 (2)
CP	Ceruloplasmin	P00450	ES	N	266 (5)	182 (4)		

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
SPP1	Osteopontin (Bone sialoprotein 1)	P10451	ES	B, I, S	248 (3)	266 (6)	79 (1)	138 (2)
HEBP1	Heme-binding protein 1 (p22HBP)	Q9NRV9	CY	N	57 (1)		119 (2)	69 (1)
LY6D	Lymphocyte antigen Ly-6D precursor	Q14210	PM	I, R, S			117 (1)	
YES1	Proto-oncogene tyrosine-protein kinase Yes	P07947	PM	I, S			115 (1)	51 (1)
ENTPD3	Ectonucleoside triphosphate diphosphohydrolase 3 (CD39L)	O75355	PM	I			116 (2)	298 (6)
PADI2	Protein-arginine deiminase type-2	Q9Y2J8	CY	G, S			111 (1)	
BSG	Basignin (CD147)	P35613	PM	G, S			77 (3)	166 (3)
HSD17B7	3-Keto-steroid reductase	P56937	PM	L, N			115 (3)	59 (3)
ABCG2	ATP-binding cassette sub-family G member 2 (CD338)	Q9UNQ0	PM	N	219 (4)		771 (12)	746 (17)
ORM1	Alpha-1-acid glycoprotein 1 (orosomucoid-1)	P02763	ES	I, Z	216 (4)	243 (6)		
A1BG	Alpha-1B-glycoprotein	P04217	CY	I	77 (3)	101 (2)	398 (9)	404 (9)
ORM2	Alpha-1-acid glycoprotein (orosomucoid-2)	P19652	ES	I, Z	71 (1)			
MUC1	Mucin-1	P15941	PM	I, L	211 (5)	185 (3)	294 (5)	284 (7)
MUC5B	Mucin 5B	Q9HC84	ES	I, L		43 (1)		
EMR2	EGF-like module-containing mucin-like hormone receptor-like 2	Q9UHX3						48 (1)
AMY1A	Salivary alpha-amylase	P04745	ES	N	198 (5)	300 (6)		
CD36	Platelet glycoprotein 4 (CD36) trombospondin receptor	P16671	PM	I, L, Z	185 (3)	214 (5)	666 (11)	818 (14)
TLR2	Toll-like receptor 2 – CD282 antigen	O60603	PM	I, L	45 (1)		636 (14)	349 (10)
STOM	Erythrocyte band 7 integral membrane protein (stomatin)	P27105	PM	U			621 (9)	539 (12)
STXBP2	Syntaxin-binding protein 2	Q15833	PM	I, L, N			291 (5)	246 (5)
STX3	Syntaxin-3	Q13277	PM	I, L, N, S, T			166 (2)	55 (1)
CNP	2′,3′-Cyclic-nucleotide 3′-phosphodiesterase (CNPase)	P09543	PM	U	44 (1)		583 (10)	391 (9)
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1 (Biliary glycoprotein) CD66a	P13688	PM	I, R		45 (1)	183 (3)	185 (5)
DHRS1	Dehydrogenase/reductase SDR family member 1	Q96LJ7	PM	L			131 (3)	88 (1)
LPL	Lipoprotein lipase	P06858	PM, CY	L, N	177 (3)	152 (4)	151 (2)	178 (2)
QSOX1	Quiescin Q6 - Sulfhydryl oxidase 1	O00391	CY	G, L	172 (6)	77 (4)		
VAT1	Synaptic vesicle membrane prot. VAT-1	Q99536	CY	G, L, N	162 (2)	221 (4)	370 (8)	148 (5)
SNAP23	Synaptosomal-associated protein 23	O00161	PM	L, N, R			78 (1)	99 (1)
EEF1A1	Elongation factor 1-alpha 1	P68104	CY	N		199 (3)	287 (6)	237 (5)
EEF1A2	Elongation factor 1-alpha 2	Q05639	NU	N	75 (2)	89 (1)		
EEF2	Elongation factors 2	P13639	CY	N			160 (3)	59 (1)
THBS1	Thrombospondin-1	P07996	ES	D, G, R, S	56 (2)	134 (2)		
PLIN3	Mannose-6-phosphate receptor-binding protein 1 (perilipin-3)	O60664	PMa	L, N	148 (2)	83 (1)	1113 (16)	1098 (19)
HSPA5	78 kDa glucose-regulated protein	P11021	CY	I, L	143 (2)	61 (1)	219 (5)	255 (5)
FN1	Fibronectin	P02751	ES	D, G, R	51 (1)	175 (3)		
HP	Haptoglobin	P00738	ES	Z	140 (3)	85 (3)		

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
HPR	Haptoglobin-related protein	P00739	ES	Z		69 (1)		
ACSL4	Long-chain-fatty-acid–CoA ligase 4	O60488	PM, ER	L, N	136 (2)		1179 (20)	1050 (20)
ACSL3	Long-chain-fatty-acid–CoA ligase 3	O95573	PM, ER	L, N		68 (1)	1043 (17)	735 (17)
AMY2B	Alpha-amylase 2B	P19961	ES	N	131 (2)			
AHSG	Alpha-2-HS-glycoprotein (Fetuin A)	P02765	CY	B, D, M, Z	128 (2)	117 (3)		
NPC2	Epididymal secretory protein E1	P61916	ES	L	89 (1)	143 (2)		
C4BPA	C4b-binding protein alpha chain	P04003	ES	I	77 (3)	163 (3)	43 (1)	
ENO3	Beta-enolase	P13929	CY	A, G, I, N	81 (1)			
HAPLN3	Hyaluronan and proteoglycan link protein	Q96S86	ES	U	51 (1)			
FOLH1	Glutamate carboxypeptidase 2	Q04609	PM	N			237 (6)	282 (5)
MARCKS	Myristoylated alanine-rich C-kinase substrate (MARCKS)	P29966	CY, PMa	L	50 (1)	65 (2)	88 (1)	93 (2)
B2 M	Beta-2-microglobulin	P61769	CY	I	50 (1)			
TC2 N	Membrane targeting tandem C2 domain-containing protein 1	Q8N9U0	NU	L, N			228 (6)	221 (4)
GAPDHS	Glyceraldehyde 3-phopshate dehydrogenase testis specific	O14556	CY	G, N	41 (1)			
GGT1	Gamma-glutamyltranspeptidase 1	P19440	PM	N, Z			379 (6)	644 (10)
GNB1	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit b-1 (trasducin b-1)	P62873	PM	L, S, T			480 (7)	238 (6)
GNB2	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit b-2 (trasducin b-2)	P62879	CY	L, S, T			314 (9)	443 (5)
GNAQ	Guanine nucleotide-binding protein G(q) subunit alpha	P50148	PM	L, S, T			300 (6)	293 (8)
GNAS	Guanine nucleotide-binding protein G(s) subunit alpha	P63092	PM	L, N, S, T	38 (1)		163 (3)	171 (5)
GNAI2	Guanine nucleotide-binding protein G(i), alpha-2 subunit	P04899	PM	L, N, S, T	51 (1)		273 (5)	167 (2)
GNAI3	Guanine nucleotide-binding G(k) subunit alpha	P08754	PM	L, S			69 (1)	270 (6)
GNA13	Guanine nucleotide-binding protein alpha-13 subunit	Q14344	PM	L, S			141 (3)	72 (2)
GNA11	Guanine nucleotide-binding subunit alpha-11	P29992	PM	L, S			66 (2)	80 (2)
GNB4	Guanine nucleotide-binding protein subunit beta 4 (transducin 4B)	Q9HAV0	PM	S, T			133 (2)	169 (2)
AUP1	Ancient ubiquitous protein 1	Q9Y679	ER, CY	S, N, T				55 (1)
PITRM1	Presequence Protease	Q5JRX3	MI	P	38 (1)			
PGK2	Phosphoglycerate kinase, testis specific	P07205	CY	N	37 (1)			
SEC16A	Protein transport protein SEC16A	O15027	ER	N, T	37 (1)			
LRRFIP2	Leucine-rich repeat flightless-interacting protein 2	Q9Y608	U	S	37 (1)			
MYOF	Myoferlin	Q9NZM1	PM	G, L, S			40 (2)	119 (2)
RHOC	Rho-related GTP-binding protein RhoC	P08134	PM	L, S			72 (5)	
RHOG	Rho-related GTP-binding protein RhoG	P84095	PM	S, T			48 (2)	59 (1)
SAR1A	GTP-binding protein SAR1a	Q9NR31	ER	N			46 (1)	46 (1)
ARL4A	ADP-ribosylation factor-like	P40617	NU	S			49 (1)	43 (1)
PPIA	Peptidyl-prolyl cis-trans isomerase A (Cyclophilin A)	P62937	CY	N			232 (6)	215 (6)

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
ATP2B2	Plasma membrane calcium-transporting ATPase 2	Q01814	PM	N, T	35 (1)		377 (6)	152 (6)
MAL2	Protein MAL2	Q969L2	PM	L, T			67 (1)	
CACNA2D1	Voltage-dependent calcium channel subunit alpha-2/delta-1	P54289	PM	T			61 (2)	133 (5)
ORAI1	Calcium release-activated calcium channel	Q96D31	PM	I, T			58 (1)	
SLC16A1	Monocarboxylate transporter 1	P53985	PM	L, N, T			54 (1)	57 (1)
SLC7A2	Low-affinity cationic amino acid transporter 2	P52569	PM	N, T			53 (1)	91 (2)
BST2	Bone marrow stromal antigen 2 (CD317)	Q10589	PM	I, T			53 (1)	51 (1)
SLC15A2	Oligopeptide transporter, kidney isoform	Q16348	PM	N, T	56 (1)		50 (1)	277 (7)
SCARB1	Scavenger receptor class B CD36 antigen-like 1	Q8WTV0	PM	I, L, R			47 (1)	145 (5)
EZR	Ezrin (cytovillin)	P15311	CY, PMa	R, S			46 (1)	
TMEM30A	Cell cycle control protein 50A	Q9NV96	PM	D, G, S			45 (1)	
ATRN	Attractin	O75882	PM	I, G, V, Z			40 (1)	
DPYSL3	Dihydropyrimidinase-related protein 3	Q14195	CY	G, V, S				58 (1)
WDR61	WD repeat protein 61	Q9GZS3	U	U			53 (1)	
EIF5A	Eukaryotic translation initiation factor 5A-1	P63241	CY	D, I, S			48 (1)	
ARF1	ADP-ribosylation factor 1	P84077	ER, Golgi	N, T			81 (4)	
ARF4	ADP-ribosylation factor 4	P18085	ER	N, T			52 (1)	
RAB1A	Ras-related protein Rab-1A	P62820	ER, PM	S, T	98 (2)		319 (5)	307 (5)
RAB1A	Ras-related protein Rab-1B	Q9H0U4	ER, PM	S, T			363 (5)	120 (4)
RAB7A	Ras-related protein Rab-7	P51149	ER, Golgi	L, N, T	40 (1)	45 (1)	430 (8)	379 (6)
RAP1B	Ras-related protein Rap-1B	P61224	PM	L, S	42 (1)		194 (3)	101 (2)
RALB	Ras-related protein Ral-B	P11234	PM	G, I, L, T	65 (1)		182 (3)	99 (3)
RAB8A	Ras-related protein Rab-8A	P61006	PM	T			62 (1)	
RAB2A	Ras-related protein Rab-2A	P61019	ER	L, N			191 (5)	155 (3)
RAB2B	Ras-related protein Rab-2B	Q8WUD1	PM	L, N	39 (1)			
RAP1A	Ras-related protein Rap-1A	P62834	PM	S	46 (1)		51 (2)	66 (2)
RRAS	Ras-related protein R-Ras	P10301	PM	L, R			50 (1)	
RAB10	Ras-related protein Rab-10	P61026	PM	S, T		159 (3)	41 (1)	209 (3)
RAB3D	Ras-related protein Rab-3D	O95716	PM	L, N			39 (1)	
RAB25	Ras-related protein Rab-25	P57735	PM	L, R				117 (1)
RAB5 C	Ras-related protein Rab-5 C	P51148	PM	L, N, T				49 (1)
RRAS2	Ras-related protein R-Ras2	P62070	PM	S				44 (1)
RAB35	Ras-related protein Rab-35	Q15286	PM	S, T			75 (2)	61 (1)
RAB5A	Ras-related protein Rab-5A	P20339	PM	L, T				43 (1)
RAB18	Ras-related protein Rab-18	Q9NP72	PM	L, T	62 (1)	40 (1)	408 (7)	
RALA	Ras-related protein Ral-A	P11233	PM	D, G, L, N, S, T	76 (1)	95 (1)	87 (1)	
RAC1	Ras-related C3 botulinum toxin substrate 1	P63000	PM	G, N			121 (3)	112 (1)

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
GLIPR2	Golgi-associated plant pathogenesis-related 1	Q9H4G4	Golgi	U	95 (1)	152 (2)	262 (4)	183 (5)
ABHD5	Abhydrolase domain-containing protein 5	Q8WTS1	CY	D, L, N			182 (2)	183 (2)
CIDEA	Cell death activator CIDE-A	O60543	PM	S	37 (1)		261 (5)	181 (3)
CD97	CD97 antigen	P48960	PM	I, R, S				65 (1)
BSCL2	Seipin	Q96G97	ER,PMa	U				65 (2)
LRG1	Leucine-rich alpha-2-glycoprotein	P02750	ES	U	117 (4)	251 (6)		
CFI	Complement factor I (C3B/C4B inactivator)	P05156	ES	I	117 (1)	140 (2)		
KRT1	Keratin, type II cytoskeletal 1	P04264	CY	L	276 (6)	111 (4)	874 (12)	795 (14)
KRT76	Keratin, type II cytoskeletal 2	Q01546	CY	L				63 (1)
KRT9	Keratin, type I cytoskeletal 9	P35527	CY	L	119 (4)	431 (8)	752 (11)	631 (15)
KRT10	Keratin, type I cytoskeletal 10	P13645	CY	L	113 (1)	78 (1)	630 (11)	292 (6)
KRT6A	Keratin, type II cytoskeletal 6A	P02538	CY	L			51 (1)	61 (1)
KRT2	Keratin, type II cytoskeletal 2 epidermal	P35908	CY	L		101 (2)	260 (5)	197 (6)
MARKSL1	MARKS related protein	P49006	PM	S			218 (2)	201 (4)
FANS	Fatty Acid syntase	P49327	PMa	L, N	45 (5)	46 (2)	1541 (35)	766 (23)
FABP3	Fatty acid-binding protein, heart	P05413	CY	L, N, T	111 (4)	193 (3)	257 (3)	377 (5)
FABP7	Fatty acid-binding protein, brain	O15540	CY	N, T			47 (2)	67 (1)
CRABP1	Cellular retinoic acid-binding protein 1	P29762	CY	N			81 (1)	100 (1)
CRABP2	Cellular retinoic acid-binding protein 2	P29373	CY, NU	N			211 (4)	237 (4)
PLIN2	Adipophilin (perilipin 2)	Q99541	PM	L	105 (9)	369 (12)	1180 (17)	1817 (30)
RHOA	Transforming protein RhoA	P61586	PM	S			269 (4)	143 (2)
TTR	Transthyretin	P02766	ES	G, N, V	88 (2)			
SDCBP	Syntenin-1 (Syndecan-Binding protein 1)	O00560	ER, PM	I, L, S, T	87 (2)	60 (1)		
FOLR1	Folate receptor alpha	P15328	PM	L, N	83 (3)	49 (3)	392 (6)	220 (7)
ADCY3	Adenylate cyclase type 3	O60266	PM	N, S	57 (1)		188 (4)	134 (4)
HLA-G	HLA class I histocompatibility antigen, alpha chain G	P17693	PM	I	60 (1)			
HLA-B46	HLA class I histocompatibility antigen, B-46 alpha chain	P30484	PM	I			42 (2)	156 (2)
HLA-B53	HLA class I histocompatibility antigen, B-53 alpha chain	P30491	PM	I				209 (3)
HLA-B13	HLA class I histocompatibility antigen, B-13 alpha chain	P30461	PM	I				142 (2)
HLA-DRB1	HLA Class II histocompatibility antigen, DRB-1 beta	P04229	PM	I		112 (2)	88 (3)	197 (8)
HLA-B	HLA class I histocompatibility antigen, B-41 alpha chain	P30479	PM	I			160 (3)	
HLA-DRB1-11	HLA class II histocompatibility antigen, DRB1-11 beta	P20039	PM	I		91 (2)	418 (6)	552 (9)
HLA-DRA1	HLA class II histocompatibility antigen, DR alpha chain	P01903	PM	I			316 (5)	370 (6)
HLA-DRB1-8	HLA class II histocompatibility antigen DRB1-8 beta	Q30134	PM	I			74 (1)	
HLA-DPA1	HLA class II histocompatibility antigen	P20036	PM	I				48 (1)
SERPING1	Plasma protease C1 inhibitor	P05155	ES	I, N, P	57 (1)			
HSPA8	Heat shock cognate 71 kDa	P11142	CY	I, L, Z	79 (1)		199 (2)	
HSPA1L	Heat shock 70 kDa protein 1L	P34931	CY	N		74 (1)		

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
ASM3B	Acid sphingomyelinase-like phosphodiesterase 3b	Q92485	CY	L, N			207 (5)	86 (6)
VTN	Vitronectin	P04004	ES	I, R	39 (2)	59 (1)		
C4BPA	C4b-binding protein alpha chain	P04003	ES	I		163 (3)		
CA6	Carbonic anhydrase 6	P23280	ES	D, G	67 (2)	115 (2)	310 (6)	196 (4)
PSAP	Prosaposin A	P07602	ES	L, N, V	45 (1)	90 (2)		
TIMP1	Metalloproteinase inhibitor 1	P01033	ES	D, G, N, P, Z	38 (1)	87 (2)		
APOH	Beta-2-glycoprotein 1 (apolipoprotein H)	P02749	ES	L, Z	48 (1)	85 (2)	110 (4)	327 (8)
APOD1	Apolipoprotein D1	P05090	ES	L, N, V				87 (1)
TMBIM1	Transmembrane BAX inhibitor motif-containing 1	Q969X1	PM	S			109 (2)	133 (3)
PODXL	Podocalyxin-like protein 1	O00592	PM	Z	46 (1)			
CD9	CD9 antigen – tetraspanin	P21926	PM	I, Z	43 (1)		94 (2)	57 (3)
SERINC2	Serine incorporator 2	Q96SA4	PM	U			94 (1)	107 (4)
GPX4	Phospholipid hydroperoxide glutathione peroxidase	P36969	CY	G, Z			90 (3)	40 (2)
UBB	Ubiquitin	P62988	CY	P		68 (1)	90 (1)	67 (1)
ZDHHC20	Probable palmitoyltransferase ZDHHC20	Q5W0Z9	PM	L, N			84 (1)	53 (1)
ANXA1	Annexin A1	P04083	CY, PMa	L, N, T			138 (2)	71 (5)
ANXA2	Annexin A2	P07355	CY, PMa	L, N, T			83 (1)	
GGT5	Gamma-glutamyltransferase 5	P36269	PM	D, G, N, T			83 (2)	51 (1)
P4HB	Protein disulfide-isomerase	P07237	CY	I, L, P			80 (3)	99 (1)
CD105	Endoglin–CD105 antigen	P17813	PM	A, R			80 (2)	72 (1)
LYN	Tyrosine-protein kinase Lyn	P07948	PM	G, S			76 (1)	60 (2)
CST3	Cystatin-C	P01034	ES	P		88 (1)		
CD81	CD81 antigen (tetraspanin-28)	P60033	PM	G, I, S			141 (2)	254 (3)
CALR	Calreticulin	P27797	ES	L, N			76 (1)	
LSS	Lanosterol synthase	P48449	ER, PMa	L, N	43 (1)	51 (1)	527 (12)	463 (8)
NSDHL	Sterol-4-alpha-carboxylate 3-dehydrogenase	Q15738	PM	L, N		51 (1)	430 (6)	199 (5)
KNG1	Kininogen-1	P01042	ES	A		48 (1)		
SLC6A14	Sodium- and chloride-dependent neutral and basic amino acid transporter B	Q9UN76	PM	T			332 (5)	400 (9)
SELENBP1	Selenium-binding protein 1	Q13228	CY	N			203 (4)	45 (3)
SLC5A1	Sodium/glucose cotransporter 1	P13866	PM	N, T	78 (1)	85 (1)	203 (3)	192 (4)
SLC1A6	Excitatory amino acid transporter 4	P48664	PM	N, T		39 (1)		
SLC34A2	Sodium-dependent phosphate transport protein 2B	O95436	PM	N, T	70 (1)		134 (4)	132 (2)
HPX	Hemopexin	P02790	ES	N	51 (1)			
SLC13A2	Solute carrier family 13 member 2	Q13183	PM	T			164 (2)	42 (3)
SOSTDC1	Sclerostin domain-containing protein	Q6X4U4	ES	U	54 (1)	66 (1)		
NPTN	Neuroplastin	Q9Y639	PM	S, R		54 (1)	41 (1)	92 (3)
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	P04406	CY	N, T	37 (1)	59 (1)		41 (1)

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
SYTL1	Synaptotagmin-like protein 1	Q8IYJ3	PM	L, T				57 (1)
MDN1	Midasin	Q9NU22	NU	N, T	41 (1)	38 (1)		
HSD17B4	Peroxisomal multifunctional enzyme type 2	P51659	CY	L	40 (1)			
VCP	Transitional endoplasmic reticulum ATPase	P55072	CY, NU	N, T		59 (1)		
UGP2	UTP–glucose-1-phosphate uridylyltransferase 2	Q16851	CY	N		82 (2)	42 (1)	44 (1)
UGP1	UTP–glucose-1-phosphate uridylyltransferase 1	Q07131	CY	N			38 (1)	
ACE	Angiotensin-converting enzyme, somatic isoform	P12821	ES	A	40 (1)		156 (3)	173 (4)
CFB	Complement factor B precursor	P00751	ES	I	66 (1)	68 (1)		
DHRSX	Dehydrogenase/reductase SDR family member on chromosome X	Q8N5I4	PMa	L, N, S			53 (1)	
CHI3L2	Chitinase-3-like protein1	P36222	ES	I	64 (1)	103 (2)		
SLC3A2	4F2 cell-surface antigen heavy chain (CD98)	P08195	PM	G, L, S, T			46 (1)	193 (3)
YWHAZ	Protein kinase C inhibitor protein 1	P63104	CY	S			44 (1)	53 (1)
TXN	Thioredoxin	P10599	ES	N, I, S			44 (1)	71 (1)
LYZ	Lysozyme C	P61626	ES	I			42 (1)	
PGK1	Phosphoglycerate kinase 1	P00558	CY	N			37 (2)	65 (1)
YWHAE	14-3-3 protein epsilon	P62258	CY	S			43 (1)	
SLC46A1	Proton coupled folate transporter	Q96NT5	PM	N, T				52 (1)
IGF1	Insulin-like growth factor IA (somatomedin C)	P01343	CY	D, G, S			40 (1)	
SEC14L1	SEC14-like protein 1	Q92503	Golgi, PM	U			43 (1)	
FAM38A	Protein FAM38A	Q92508	PM, ER	G, R				42 (1)
MYADM	Myeloid-associated differentiation marker	Q96S97	PM	D			42 (2)	121 (2)
CD47	Leukocyte surface antigen CD47	Q08722	PM	I, R, T				44 (3)
CD59	CD59 glycoprotein precursor	P13987	PM	I, L	55 (1)	44 (1)	38 (2)	407 (7)
PGDH	6-phosphogluconate dehydrogenase	P52209	CY	N				44 (1)
C20ORF117	Protein KIAA0889	O94964	U	U			39 (1)	
NCSTN	Nicastrin	Q92542	PM	P, S, T			38 (1)	42 (1)
PLG	Plasminogen	P00747	ES	I, N, P	39 (1)		42 (1)	103 (2)
CDC42	Cell division control protein 42 homolog	P60953	PM	D, S	53 (1)		44 (1)	50 (1)
ACACA	Acetyl-CoA carboxylase 1	Q13085	CY	L, N			38 (1)	
NDRG1	Protein NDRG1	Q92597	Cy, PM	D, G			37 (1)	
CDKN3	Cyclin-dependent kinase inhibitor 3	Q16667	CY	D, G			37 (1)	
KCNS1	Potassium voltage-gated channel subfamily S member 1	Q96KK3	PM	N, R,T			36 (1)	
SLC1A5	Neutral amino acid transporter B(0)	Q15758	PM	I, R, T			36 (1)	41 (1)
CFL1	Cofilin-1	P23528	CY, NU	A, L			35 (1)	
NUP93	Nucleoporin Nup93	Q8N1F7	NU	N, T		43 (1)		
NUP160	Nucleoporin Nup160	Q12769	NU	N, T			39 (1)	
ITPR2	Inositol 1,4,5-trisphosphate receptor type 2	Q14571	ER	S, R, T		43 (1)		

Table 1 (*Continued*)

Symbol	Identified protein	UNIPROT accession number	Subcellular location	Role	hMS score (*)	Deglycosylated hMS score (*)	hMFGM score (*)	Deglycosylated hMFGM score (*)
ACSL1	Long-chain-fatty-acid–CoA ligase 1	P33121	PMa	L, N			34 (1)	42 (1)
TCEB1	Transcription elongation factor B	Q15369	NU	N, P		46 (1)		
EEA1	Early endosome antigen 1	Q15075	CY	L, N, T				41 (1)
SSR1	Translocon-associated protein alpha chain	P43307	ER, PMa	P, R, S				41 (1)
PRSS8	Prostasin	Q16651	PM	A, P, S, Z				40 (1)
ITGB4	Integrin beta4 (CD104 antigen)	P16144	PM	D, R		42 (1)		
SEMA6A	Semaphorin 6A	Q9H2E6	PM	D, V, T		41 (1)		
TLL2	Tolloid-like protein 2	Q9Y6L7	ES	D, G, P				40 (1)
SYNE2	Nesprin-2	Q8WXH0	PM	D, V				40 (1)
BMP1	Bone morphogenetic protein 1	P13497	ES	B, G, P				39 (1)
CLDN4	Claudin-4	O14493	PM	I, R, T				37 (1)

are also involved in delivering microelements, such as metal ions, through casein phosphopeptides. Through gastrointestinal digestion, caseins, LTF and other proteins are known to release bioactive peptides, which may exert a number of physiological functions [28].

Among the dominant molecular entities of the human milk proteome, LTF, PIGR, IgA, IgG, κ-casein, and BSSL are glycoproteins. LALBA, the most abundant protein in mature human milk, is minimally glycosylated (<1%). It is thought that glycoproteins are involved in conferring passive immunity to the newborn. On the other hand, many of the identified proteins have, as yet, undetermined function. Despite the strict evolutionary selection of the lactating machinery, it is possible that some identified proteins may not have a specific function, e.g., those coming from the blood. Nevertheless, for the purposes of discussion, proteins were categorized into functional groups based on their primary task(s) according to the UniProt database (Fig. 2).

Several hydrolytic proteins mediate the digestion of macronutrients (e.g., BSSL, LPL, AMY1A, AMY2B and proteases). GGT1 and GGT5 transfer the glutamyl moiety of glutathione to several

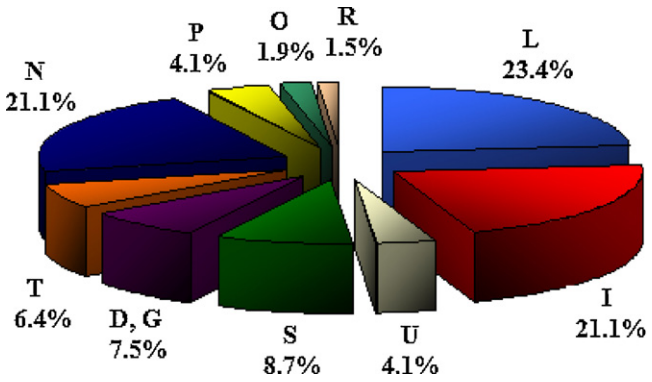

Fig. 2. Functional categorization of human milk proteins. The primary supposed function, according to the UniProt database, was assigned. Isoforms of immunoglobulins and keratins were excluded by computation. Symbols for primary protein function: D, G, cell differentiation, cell growth, tissue development; I, defense/immunity; L, lipid droplet transport/secretion; N, synthesis and delivery of nutrients/metabolites, energy pathways; O, others (angiogenesis, inflammation, nervous system maturation); P, protein metabolism (protease/peptidase or inhibitor); R, cellular recognition; S, signal transduction; T, cellular trafficking; U, unknown.

acceptors modulating redox homeostasis. Some protease inhibitors (AAT, AACT, SERPING1, TIMP1, CST3) balance the action of proteases, which may prevent extensive degradation before they reach the gastrointestinal tract.

A series of proteins (21.1%) are delegated to the synthesis/transport and delivery of non-lipid nutrients/metabolites. Among these were generic (ALB) or specific protein carriers, those able to bind metal ions (CAL3, SELENBP1), vitamins (CRABP1, CRABP2, FOLR1, TTR), fatty acids (ACSL1, ACSL3, ACSL4, FANS, FABP3, FABP7, APOH, APOD1, SLC16A1, ZDHHC20), heme (HPX), and steroids (LSS, NSDHL). Selected proteins, such as LTF and TF, have been thought to exert a bacteriostatic action sequestering nutrients that are vital for bacteria. The energy/metabolism-related enzymes, such as those of glycolysis, can be categorized in this group.

Through the binding/transport of nutrients, many proteins cooperate to promote cellular proliferation/differentiation and cell growth/differentiation (7.5%). Among these, protein components with hormone-like activity are specifically involved in this role (CA6, CDC42, IGF1, NDRG1, RALA, TMEM30A, THBS1). CHRDL2, SPP1, AHSG, BMP1 induce the growth/differentiation of bone-related cells. Proteins with a specialized signaling function (8.7%) in extended metabolic pathways [18] have been separately categorized.

Interestingly, the most consistent group of proteins (23.4%) is part of the complex machinery that delivers lipids to the breast-fed newborn. Recent proteomic analyses indicate that, although they have been considered passive storage depots, lipid droplets are, rather, discrete organelles present in most cell types, both prokaryotic and eukaryotic, where they are involved in multiple functions [34]. Lipid droplet proteins participate in lipid synthesis and, as trans-membrane proteins, mediate the distribution of neutral lipids and phospholipids to different membrane-bound organelles. The primary organelles of the secretory pathway are the endoplasmic reticulum (ER) and the Golgi complex. Membrane trafficking between these major elements of the secretory pathway is largely mediated by discontinuous carriers. Small guanosine triphosphatases (GTP-ases) modulate vesicle formation, motility, targeting, and docking. The Rab proteins are low molecular weight GTP-binding proteins that form the largest branch of the Ras superfamily of GTP-ases. The surprisingly large number of GTP-binding proteins, RAB isoforms, RAP1A and RAP1B, membrane proteins

such as STX3, which has a SNARE-like domain, ANXA1, ANXA2, VAT1, SEC16A and SEC14L1 found in human milk is explained by their involvement in milk fat synthesis and secretion, which are among the main tasks of the mammary gland cells. RAB proteins, together with the so-called PAT proteins that include PLIN2 (adipophilin) and PLIN3, are thought to regulate the transient interaction of specific membrane systems, with the aim of delivering lipids between membrane compartments [35]. Homologues of RAB proteins and GTP-ases and their relevance in milk are progressively emerging in proteomic investigations of bovine MFGMs [10,24,26]. NPC2 participates in the regulation of lipid composition, while SNAP23 is an essential component of the high-affinity receptor for the general membrane fusion machinery and is an important regulator of vesicular transport, docking and fusion. It has been suggested that the protein system that regulates the lipid metabolism in milk might have an indirect major protective function [36]. Lipid droplets may also include other cytoskeleton proteins, such as actin isoforms, keratins and MARCKS, with structuring and propelling functions [37]. However, keratins identified in human milk likely arise from the flaking of mammary epidermal tissue. ARF protein isoforms are also involved in membrane targeting during recognition and fusion steps. EEA1 binds RAB5 and mediates the homotypic fusion of endosomes [38]. The majority of the lipid droplet-associated proteins are found in the MFGM fraction because of the fusion of lipid droplets with the cell outer membrane during excretion.

Proteins that regulate intra- and extra-cellular trafficking (6.4%) are specifically indicated in Table 1. The major protein ATP2B2, required for the relevant secretion of calcium in the milk and already described in bovine mammary tissue and MFGMs [10,24], as well as VCP, CACNA2D1 and ORAI1 should belong to this group. No phosphate transport was thought to occur across the apical membrane of the mammary secretory cell until a sodium-dependent phosphate transporter was identified in bovine MFGMs [10,24]. Therefore, the occurrence of SLC34A2 in human milk is not surprising. An expected sodium-dependent glucose transporter (SLC5A1), as well as amino acid (SLC1A6, SLCA14, SLC13A2, SLC1A5) and oligopeptide (SLC15A2) transporters, was found. Proteomic data indicate an apical localization of the transporters, associated with the plasma membrane. A further small set of cell membrane proteins (1.5%) seems to be specifically involved in mutual cell recognition.

A group of 56 (21.1%) identified proteins have specific immunocompetent functions and act as modulators that compensate for developmental delays in the immune system in early infancy. Secretory IGA (sIGA) is known to be a major component in conferring passive immunity to the breastfed infant. Several antibody chains have been identified (IGHM, IGHC1, IGKC, IGHA2, IGJ, and IGLC1) that are, along with numerous isoforms, likely related to individual variability. The majority of IGs are known to be excreted via a selective receptor-mediated intracellular route. These IGs may be blood-derived or produced by mammary plasma cells. The translocation of IgA across the mammary epithelial cells is assisted by PIGR, which is expressed on the mucosal epithelium as one of the main human milk proteins.

Both the adaptive and the innate immune system operate to confer passive immunity to the newborn. Immune cells are effectors of the immunomodulatory activity of milk. Thus, there are complement-related proteins that are derived from neutrophils, macrophages and lymphocytes, as well as proteins involved in antigen processing and presentation of peptide/lipopolysaccharide by MHC complexes of Toll-like receptors (TLR2). CD14 seems to have a direct involvement in pathogen responses; it functions as a co-receptor for the detection/recognition of bacterial lipopolysaccharides, enhances host cell activation, and likely imprints the neonatal immune system. CD14 and Toll-like receptors are primarily expressed by antigen-presenting cells. Thus, their occurrence, particularly at the level of the MFGMs, suggests that mammary cells may have a direct role in detecting infections. Indeed, the mammary epithelium itself plays an active role in host defense by synthesizing innate immune factors, such as LTF, and releasing peptides with antibacterial activity. Apolipoproteins also have been shown to be involved in anti-inflammatory functions and in the balance of the immune response [39]. Several modulators of cell differentiation may have a chemotactic role, contributing to induce the response of the innate immune system. For instance, in addition to the role played in vesicular trafficking, SDCBP in human milk is an IgA-inducing factor for naive B cells [40]. However, a detailed discussion on the immunological mechanisms promoted by milk is beyond the purposes of this article. The list of the immune components of bovine and human milk and colostrum has been recently updated, and the presumed interactomic pathways have been reviewed elsewhere [17,23,41].

4. Conclusion

The present investigation represents one of the most comprehensive inventories of the human milk proteome, thus contributing to its completion. The protein annotation in human milk has continuously progressed over the last decade, coinciding with impressive analytical advances.

Although upgraded MS instruments have declared limits of detection reaching the low femtomole or attomole range, the actual MS sensitivity is determined by the complexity of the sample and by the limited dynamic range of the instrument. Thus, only partially overlapping protein subsets have been defined for human milk, while a plethora of minor proteins are described exclusively in single reports. For instance, very low-abundance proteins, such as cytokines, are rarely harvested in discovery-driven proteomic analyses and have been detected only using targeted proteomic approaches [42]. Our analysis did not make exceptions, as many minimally abundant yet well-established polypeptides such as phosphatases (human isoform), lactoperoxidase, kallikrein, interleukins, protease inhibitors, hormones (prolactin) and growth factors commonly escape identification. This information implies that the more than 300 currently identified gene products could be only a surface view of the entire human milk proteome. The partial inconsistencies in the reported human milk proteome subsets should be primarily ascribed to the pitfalls of the analytical strategies and to a laboratory-dependent identification bias, rather than to a stochastic sampling of the proteome or to actual individual variability in milk samples [43].

A recent estimate has established that a minor percentage of peptides (not larger than 16%) can be targeted for MS/MS in an ordinary shotgun analysis of complex proteomes [44]. Most likely the introduction of a previous off-line chromatographic dimension can be helpful in further enlarging the human milk proteome coverage.

For this purpose, it is important to keep in mind that whatever the method for descriptive analysis of complex proteomes, no current protocol yet provides a comprehensive snapshot of complex biological systems in a single run. For this reason, specifically designed strategies are required to target post-translationally modified proteins in human milk [31,45].

Acknowledgements

MS experiments have been performed at the National Center for Glycomics and Glycoproteomics of Indiana University in Bloomington, USA. Permanence of G.P. in the laboratory directed by Prof. Milos V. Novotny was supported by the “short-term mobility program” of National Research Council of Italy (CNR). The

authors gratefully acknowledge the American Journal Experts for manuscript revision.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2012.01.014.

References

[1] D.G. Lemay, D.J. Lynn, W.F. Martin, M.C. Neville, et al., Genome Biol. 10 (2009) R43.
[2] A. Mangé, V. Bellet, E. Tuaillon, P. van de Perre, J. Solassol, J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci. 876 (2008) 252.
[3] K. Murakami, M. Lagarde, Y. Yuki, Electrophoresis 19 (1998) 2521.
[4] D. Fortunato, M.G. Giuffrida, M. Cavaletto, L.P. Garoffo, G. Dellavalle, L. Napolitano, C. Giunta, C. Fabris, E. Bertino, A. Coscia, A. Conti, Proteomics 3 (2003) 897.
[5] M. Cavaletto, M.G. Giuffrida, A. Conti, Clin. Chim. Acta 347 (2004) 41.
[6] A. Conti, M.G. Giuffrida, M. Cavaletto, Proteomics of Human Body Fluids, Humana Press, Totowa, NJ, 2007.
[7] M. Cavaletto, M.G. Giuffrida, A. Conti, Adv. Exp. Med. Biol. 606 (2008) 129.
[8] M. Yamada, K. Murakami, J.C. Wallingford, Y. Yuki, Electrophoresis 23 (2002) 1153.
[9] D.J. Palmer, V.C. Kelly, A.M. Smit, S. Kuy, C.G. Knight, G.J. Cooper, Proteomics 6 (2006) 2208.
[10] T.A. Reinhardt, J.D. Lippolis, J. Dairy Res. 73 (2006) 406.
[11] G. Smolenski, S. Haines, F.Y. Kwan, J. Bond, V. Farr, S.R. Davis, K. Stelwagen, T.T. Wheeler, J. Proteome Res. 6 (2007) 207.
[12] C. D'Ambrosio, S. Arena, A.M. Salzano, G. Renzone, L. Ledda, A. Scaloni, Proteomics 8 (2008) 3657.
[13] E.M. Keidel, D. Ribitsch, F. Lottspeich, Proteomics 10 (2010) 2089.
[14] K. Mann, M. Mann, Proteome Sci. 9 (2011) 1.
[15] B.R. Fonslow, P.C. Carvalho, K. Academia, S. Freeby, T. Xu, A. Nakorchevsky, A. Paulus, J.R. Yates III, J. Proteome Res. 10 (2011) 3690.
[16] A. D'Amato, A. Bachi, E. Fasoli, E. Boschetti, G. Peltre, H. Sénéchal, P.G. Righetti, J. Proteome Res. 8 (2009) 3925.
[17] Y. Liao, R. Alvarado, B. Phinney, B. Lönnerdal, J. Proteome Res. 10 (2011) 3530.
[18] A. D'Alessandro, A. Scaloni, L. Zolla, J. Proteome Res. 9 (2010) 3339.
[19] L. Bianchi, M. Puglia, C. Landi, S. Matteoni, D. Perini, A. Armini, M. Verani, C. Trombetta, P. Soldani, P. Roncada, G. Greppi, V. Pallini, L. Bini, J. Proteomics 72 (2009) 853.
[20] C.C. Wu, J.R. Yates III, Nat. Biotechnol. 21 (2003) 262.
[21] M.R. Roe, T.J. Griffin, Proteomics 6 (2006) 4678.
[22] Y. Shen, N. Tolić, R. Zhao, L. Pasa-Tolić, L. Li, S.J. Berger, R. Harkewicz, G.A. Anderson, M.E. Belov, R.D. Smith, Anal. Chem. 73 (2001) 3011.
[23] K. Hettinga, H. van Valenberg, S. de Vries, S. Boeren, T. van Hooijdonk, J. van Arendonk, J. Vervoort, PLoS One 6 (2011) e19433.
[24] T.A. Reinhardt, J.D. Lippolis, J. Dairy Sci. 91 (2008) 2307.
[25] A. Le, L.D. Barton, J.T. Sanders, Q. Zhang, J. Proteome Res. 10 (2011) 692.
[26] C. Vanderghem, C. Blecker, S. Danthine, C. Deroanne, E. Haubruge, F. Guillonneau, Int. Dairy J. 18 (2008) 885.
[27] M.W. Duncan, R. Aebersold, R.M. Caprioli, Nat. Biotechnol. 28 (2010) 659.
[28] P. Ferranti, M.V. Traisci, G. Picariello, A. Nasi, V. Boschi, M. Siervo, C. Falconi, L. Chianese, F. Addeo, J. Dairy Res. 71 (2004) 74.
[29] A.I. Nesvizhskii, R. Aebersold, Mol. Cell Proteomics 4 (2005) 1419.
[30] Y. Liao, R. Alvarado, B. Phinney, B. Lönnerdal, J. Proteome Res. 10 (2011) 5409.
[31] G. Picariello, P. Ferranti, G. Mamone, P. Roepstorff, F. Addeo, Proteomics 8 (2008) 3833.
[32] J.W. Froehlich, E.D. Dodds, M. Barboza, E.L. McJimpsey, R.R. Seipert, J. Francis, H.J. An, S. Freeman, J.B. German, C.B. Lebrilla, J. Agric. Food Chem. 58 (2010) 6440.
[33] C.C. Wu, K.E. Howell, M.C. Neville, J.R. Yates III, J.L. McManaman, Electrophoresis 21 (2000) 3470.
[34] R.V. Farese Jr., T.C. Walther, Cell 139 (2009) 855.
[35] P. Liu, R. Bartz, J.K. Zehmer, Y.S. Ying, M. Zhu, G. Serrero, R.G. Anderson, Biochim. Biophys. Acta 1773 (2007) 784.
[36] J.K. Zehmer, Y. Huang, G. Peng, J. Pu, R.G. Anderson, P. Liu, Proteomics 9 (2009) 914.
[37] M. Hamosh, J.A. Peterson, T.R. Henderson, C.D. Scallan, R. Kiwan, R.L. Ceriani, M. Armand, N.R. Mehta, P. Hamosh, Semin. Perinatol. 23 (1999) 242.
[38] S.Y. Dejgaard, A. Murshid, A. Erman, O. Kizilay, D. Verbich, R. Lodge, K. Dejgaard, T.B. Ly-Hartig, R. Pepperkok, J.C. Simpson, J.F. Presley, J. Cell Sci. 121 (2008) 2768.
[39] M. Danielsen, M.C. Codrea, K.L. Ingvartsen, N.C. Friggens, E. Bendixen, C.M. Røntved, Proteomics 10 (2010) 2240.
[40] M.M. Sira, T. Yoshida, M. Takeuchi, Y. Kashiwayama, T. Futatani, H. Kanegane, A. Sasahara, Y. Ito, M. Mizuguchi, T. Imanaka, T.A. Miyawaki, Int. Immunol. 21 (2009) 1013.
[41] A. D'Alessandro, L. Zolla, A. Scaloni, Mol. Biosyst. 7 (2011) 579.
[42] T. Nilsson, M. Mann, R. Aebersold, J.R. Yates III, A. Bairoch, J.J. Bergeron, Nat. Methods 7 (2010) 681.
[43] A.W. Bell, E.W. Deutsch, C.E. Au, R.E. Kearney, R. Beavis, S. Sechi, T. Nilsson, J.J. Bergeron, Nat. Methods 6 (2009) 423.
[44] A. Michalski, J. Cox, M. Mann, J. Proteome Res. 10 (2011) 1785.
[45] A.G. Poth, H.C. Deeth, P.F. Alewood, J.W. Holland, J. Proteome Res. 7 (2008) 5017.