Zeus, Aesculapius, Amalthea and the proteome of goat milk

Vincenzo Cunsolo[*,a], Elisa Fasoli[b], Rosaria Saletti[a], Vera Muccilli[a], Serafina Gallina[a], Pier Giorgio Righetti[b], Salvatore Foti[a]

[a] Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy

[b] Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Via Mancinelli 7, Milano 20131, Italy

Keywords: goat milk proteome; low-abundance proteins; combinatorial peptide ligand libraries; milk allergy; orbitrap tribrid mass spectrometer

Corresponding Author:
Vincenzo Cunsolo, Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy.
Tel. 0039 095 7385029
Fax 0039 095 580138
E-mail: vcunsolo@unict.it

Abstract

The goat whey proteome has been explored in depth via capture with combinatorial peptide ligand libraries (CPLL) at three different pH values. A total of 452 unique species has been tabulated, a proteome discovery so far unmatched in any single other investigation of milk from any mammalian species. This massive discovery is probably related to: i) the extraordinary load of proteins onto the CPLL beads (i.e. two grams for each different pH capture) vs. barely 100 μL of beads; ii) the high resolution/high mass accuracy of mass spectral data; and iii) the use of two complementary tools, Mascot and PEAKS, each one contributing to a set of unique protein IDs. Due to the relative paucity of available protein annotations for goat, only 10% of the identified proteins belong to the capra, whereas 52% are specific of sheep and 37% are homologous to that of bovine milk. Gene Ontology classification of these proteins allowed to ascertain their molecular function, cellular location and the biological processes in which they are involved. Most of them play catalytic activity and protein binding, are mainly located in the extracellular region and are principally involved in metabolic processes, regulation of biological processes and response to stimulus. This work reports the largest description so far of the goat milk proteome, which has been compared with cow's milk proteome and would thus help to understand the importance of low-abundance proteins with respect to the unique biological properties of this nutrient.

1. Introduction

In the last few years extensive proteomic analysis has been applied to human, cow's and other mammalian milks, due to the importance of this biological fluid in nutrition. In fact milk is the most important food for young mammals and a common source of proteins and microelements for adult people as well. In addition, it is an important means for transferral of immunity to pathogens from the mother to the new-born, since it contains antimicrobial and immuno-modulatory proteins that are active in his digestive tract. In particular cow's milk (CM) is of great human nutritional and economic significance, although in the pre-proteomic days its repertoire of low-abundance proteins had not been much disclosed. So much so that in those days barely a couple of dozens of proteins had been fully characterized. An important step forward occurred in 2009, when D'Amato et al. [1] reported for the first time the presence of a grand total of 149 unique gene products. Soon after, research on the global composition of cow's whey took an exponential growth. Thus, in 2011 Hettinga et al. [2] reported as many as 192 proteins and Le et al. [3] listed as many as 293 unique species, by resorting to various pre-fractionation techniques. An interesting study has also been published by Lemay et al. [4] who, by using publicly available milk proteome data and mammary expressed sequence tags, reported identification of 197 milk protein genes and over 6,000 mammary genes in the bovine genome. Finally, the wealth of data present in the literature about bovine milk were summarized by D' Alessandro et al. [5], who compiled an exhaustive list of 573 non-redundant annotated protein entries. Most of the milk proteins were grouped under pathways, networks, or ontologies referring to nutrient transport, lipid metabolism, and objectification of the immune system response, respectively. Extensive reviews summarizing the state of the art have also recently appeared [6,7], the one by Agrawal et al. dedicated also to food security and safety issues.

Notwithstanding the large consumption and world-wide availability of CM, there are major problems on the use of this important source of food for infant nutrition: the growing occurrence of allergies. Between 5% and 15% of infants show symptoms suggesting adverse reactions to CM proteins [8], while estimates of the prevalence of cow's milk protein allergy (CMPA) vary from 2% to 7.5%. Thus, a search to find valid alternatives has a high priority. According to many clinical trials [9,10], the best natural candidate to this purpose appears to be donkey's milk (DM), which presents a strict resemblance to human milk (HM), with similar lactose and mineral contents, fatty acid and protein profiles. In particular, DM is the best candidate as a substitute of HM for clinical tolerability, palatability and nutritional adequacy for children affected by CMPA, furnishing additional physiological functions as well, such as antibacterial properties, digestive activity molecules and growth factors and hormones [11,12]. The relationship between hypo-allergenicity of DM and its proteome fraction has been recently explored by Cunsolo et al. In an in-depth proteomic investigation they have described 106 unique products in this biological fluid, among which 10% could be ascribed to the donkey, 70% were homologous to *Equus caballus* and only just about 3% could be attributed to bovine milk [13]. Indeed, milk proteins from donkey and *Bos taurus* share low-sequence similarity, as reported earlier by the same authors [14,15,16]. Particularly, sequence alignment of donkey's α_{s1}- and α_{s2}-CNs (caseins) with their bovine counterparts (i.e. two of the major cow's milk allergens) highlighted that the IgE-binding linear epitopes of both cow's α_s-CNs and the corresponding domains present in donkey's counterparts had remarkable differences in their amino acid sequences, which could be related to the already demonstrated low allergenic properties of DM.

Goats, as donkeys, have been widely used in ancient times for feeding babies, and also represent, in classical mythology, a famous example of breast-feeding animals. One might wonder about the curious title, but in Greek mythology the goat (*Capra hircus*)

has been regarded as a sacred animal, so much so that it is stated that Zeus (or Jupiter for the Romans), as a newborn in the island of Crete, had been nourished by the goat Amalthea, which, as a reward, got a place in the firmament as the constellation Capella. Also Aesculapius, the health God, had been nourished by a goat. Although these might be discarded as mere legends, in real life the goat has always been considered as a beneficial animal, a protector of human life. Thus, in ancient Greece, the rural population quite often adopted goats as foster-mothers for their newborns. It is a fact that a quite common first name in Greece, Egisto (Αἴγισθος in Greek or Aegisthus in Latin), literally means “nourished by a goat”.

Nowadays, the use of goat’s milk (GM) for feeding healthy babies or as a possible alternative to CM for allergic subjects is still debated [17,18,19]. GM is more easily digested than bovine milk also because of its higher content in essential fatty acids. Moreover, in diets administered by parents to children with atopic dermatitis, it seems to be less allergenic than CM and may be used as a dietary supplement in individuals with inflammatory and allergic conditions. On the other hand, it has been shown that many children who are allergic to CM are also sensitized to proteins of GM, probably reflecting the close phylogenetic relations between these animals and the high sequence identity of their homologue proteins [20]. In recent years, different studies have reported a comparison of the milk proteome profiles of some animal species, including goat, for identifying sources of hypoallergenic alternatives to bovine milk [21,22]. Moreover, by monitoring the characteristic peaks of the most abundant proteins, MALDI-TOF MS-based methods have been recently developed to detect fraudulent adulterations or unintended contaminations of other milks to DM and GM and therefore to assess the genuineness of these milks [23, 24]. However, a complete understanding of the composition and functions of GM proteins is hampered by incomplete knowledge of its milk proteome. Tay and Gam [25] have compared the proteomic data of bovine and

caprine milk, but indeed their report deals only with the characterization of the most abundant species. Similar data, on the major proteins in GM, have also been offered by Selvaggi et al. [26] and Almeda da Costa et al. [27], all of them dealing with at most a handful of high-abundance compounds. On the light of this scenario, we report here a most extensive investigation on the goat whey proteome by coupling the power of the CPLL technology, SDS-PAGE, high resolution nLC-nESI MS/MS and database searching, and describe its unique biological properties.

2. Materials and methods

2.1. Chemicals

ProteoMiner™ (combinatorial hexapeptide ligand library beads, CPLL), Laemmli buffer, 40% acrylamide/Bis solution, N,N,N',N'-tetramethylethylenediamine (TEMED), molecular mass standards and electrophoresis apparatus for one-dimensional electrophoresis were from Bio-Rad Laboratories, Inc., Hercules CA. β-mercaptoethanol, dithiothreitol (DTT), ammonium persulfate, 3-[3-cholamidopropyl dimethylammonio]-1-propanosulfonate (CHAPS), acetonitrile (ACN), trifluoroacetic acid (TFA), sodium dodecyl sulphate (SDS), iodoacetamide (IAA), formic acid (FA) and all other chemicals used all along the experimental work were current pure analytical grade products and purchased from Sigma-Aldrich Corporation, St Louis, MO. Water and acetonitrile (OPTIMA® LC/MS grade) for LC/MS analyses were purchased from Fisher Scientific, UK. Complete protease inhibitor cocktail tablets were from Roche Diagnostics (Basel, CH). Modified porcine trypsin was purchased from Promega (Madison, WI, USA).

2.2. Goat milk treatment

The milk has been collected from one hundred animals of a strain called “Camosciata”, raised in a Sicilian farm. Immediately after collection the milk was added with one tablet

148 (per 100 mL) of the Roche cocktail of five protease inhibitors. Prior to CPLL treatment,
149 one liter of milk was first delipidated by centrifugation and then casein depleted by
150 precipitation at pH 4.6. After lyophilization, six grams of total proteins were dissolved in
151 150 mL of 30 mM phosphate buffer, pH 7.2. This solution was divided into three, fifty mL
152 fractions. One of them was loaded onto 100 µL of CPLL beads equilibrated into the
153 same pH 7.2, buffer, the other two fractions were titrated, respectively, to pH 2.2 and to
154 pH 9.0 and loaded onto the same CPLL bead volumes equilibrates at the two different
155 pH values [28]. The capture was performed batch-wise, in small flasks gently rocking on
156 a shacking platform for 2 hours. After that, the beads were rinsed twice with the
157 equilibration buffers, so as to remove any excess of non-adsorbed proteins (see the
158 scheme of Fig. 1). Desorption was implemented by washing the beads twice (each time
159 with 50 µL) with a boiling 4% SDS solution containing 20 mM DTT, 12.5 % (v/v)
160 glycerol, 0.005% (m/v) bromophenol blue, and 62.5 mM Tris-HCl (pH 6.8) [29].
161
162 ***2.3. SDS-PAGE analysis***
163 Ten µL of the above eluates (labelled as 2.2, 7.2 and 9.0) and of the non-treated
164 sample (CTRL) were loaded onto an SDS-PAGE gel, composed by a 4%
165 polyacrylamide stacking gel (125 mM Tris-HCl, pH 6.8, 0.1%, m/v, SDS) over a 12%
166 resolving polyacrylamide gel (in 375 mM Tris-HCl, pH 8.8, 0.1%, m/v, SDS buffer). The
167 cathodic compartment was filled with Tris-glycine buffer, pH 8.3, containing 0.1%, m/v,
168 SDS whereas the anode Tris buffer, at pH 8.8, was present. Electrophoresis was at 100
169 V until the dye front reached the bottom of the gel and 150 V until the end of the
170 separation. Staining and distaining were performed with Colloidal Coomassie Blue and
171 7% (v/v) acetic acid in water, respectively. After this, the four tracks labelled CTRL, 2.2,
172 7.2 and 9.0 in Fig. 2 were subdivided into 12 segments of the same dimensions along
173 the migration path, from anode to cathode and subjected to the standard procedure of

reduction and alkylation, followed by trypsin digestion [13]. The recovered peptides were first lyophilized and then resuspended in 30 µL of water/acetonitrile (98:2) added with 0.1% FA (formic acid).

2.4. Mass spectrometry analysis

Mass spectrometry data were acquired on a Thermo Fisher Scientific Orbitrap Fusion Tribrid (Q-OT-qIT) mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Liquid chromatography was carried out using a Thermo Scientific Dionex UltiMate 3000 RSLCnano system (Sunnyvale, CA). One microliter of the reconstituted samples was loaded onto an Acclaim® Nano Trap C18 Column (100 µm x 2 cm, 5 µm, 100Å). After washing the trapping column with solvent A (H_2O/ACN, 98/2 + 0.1% FA) at a flow rate of 7 uL/min for 3 min, the solution was switched from the trapping column onto a PepMap® RSLC C18 EASY-Spray column (75 µm x 50 cm, 2 µm, 100Å). Peptides were separated by elution at a flow rate of 0.3 uL/min and 40°C with a linear gradient of solvent B (ACN + 0.1% FA) in A from 1% to 15% in 10 min, followed by 15% to 30% in 65 min, 30% to 50% in 15 min, and 50% to 98% in another 10 min. We finished by holding 98% B for 5 minutes and re-equilibrating the column at 1% B for 20 minutes. Eluting peptide cations were converted to gas-phase ions by electrospray ionization using a source voltage of 1.75 kV and introduced into the mass spectrometer through a heated ion transfer tube (275 °C). Survey scans of peptide precursors from 200 to 1600 *m*/*z* were performed at 120K resolution (@ 200 *m*/*z*) with a $4 \cdot 10^5$ ion count target and a maximum injection time of 50 ms. Tandem MS was performed by isolation at 1.6 Th with the quadrupole, HCD fragmentation with normalized collision energy of 35, and rapid scan MS analysis in the ion trap. The MS^2 ion count target was set to 10^2 and the maximum injection time was 250 ms. Only those precursors with charge state 2–4 and an intensity above the threshold of $5 \cdot 10^3$ were sampled for MS^2. The dynamic exclusion

duration was set to 60 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 3 s cycles, meaning the instrument would continuously perform MS^2 events until the list of non-excluded precursors diminishes to zero or 3 s, whichever is shorter. MS/MS spectral quality was enhanced enabling the parallelizable time option (i.e. by using all parallelizable time during full scan detection for MS/MS precursor injection and detection).

Mass spectrometer 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. 3.0.63 software (Thermo Fisher Scientific).

2.5. Protein identification

LC/MS/MS data were processed by Proteome Discoverer v. 1.4.1.14 (Thermo Scientific). Data were searched against the "Ruminantia" UniProt database (SwissProt and trEMBL release March 2015, containing 91884 entries) using the Mascot algorithm (Matrix Science, London, UK, version 2.5.1). Full tryptic 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 phosphorylation of serine, threonine and tyrosine residues were included as variable modifications. The precursor mass tolerance threshold was 10 ppm and the max fragment mass error was set to 0.6 Da. Peptide spectral matches (PSM) were validated using Target Decoy PSM Validator node based on q-values at a 1% FDR. A protein was considered identified with: minimum of 2 peptides with a Mascot score >20 and Expect <0.05; proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of

parsimony. Proteome Discoverer also retrieved, when available, the annotation (biological processes, cellular components, and molecular functions) of the identified unique gene products.

Since the *Capra* has an incomplete protein database, LC/MS/MS data were also processed and analyzed by PEAKS *de novo* sequencing software (v. 7.0, Bioinformatics Solutions Inc., Waterloo, ON Canada). The amino acid sequences generated from each spectrum were searched against the "*Ruminantia*" UniProt database using the SPIDER algorithm, a dedicated search tool of PEAKS that is specially designed to detect peptide mutations and perform cross-species homology search [30].

3. Results

Figure 1 depicts the scheme of pre-fraction of whey proteins via the CPLL methodology, their separation by SDS-PAGE electrophoresis and characterization by MS analysis. During CPLL, note the massive overload in terms of proteins, amounting to a total of two grams for each pH value against a volume of barely 100 µL for the accepting beads. As shown below, this permitted a much higher visibility of the low- to very low-abundance species.

Figure 2 gives the SDS-PAGE profiling of the control, untreated sample (CTRL) versus the CPLL beads eluates from the three pH values (track 2.2, 7.2 and 9.0). It can be noted that, whereas in the control only a few, distinct bands are seen along the 10 to 250 kDa space, in all the three eluates many more components are visible, with Mr values covering the entire spectrum from ca. 10 kDa up to 250 kDa, with the pH 2.2 eluate being particularly enriched in proteins.

Figure 3 gives the Venn diagrams of the MS-identified proteins via various means. The upper panel displays the IDs as obtained by interrogating the databases either with the Mascot or with the PEAKS softwares. Only 206 unique products are common to the two

programs, Mascot contributing to an additional 98 species and PEAKS to another 148 unique gene products (for all IDs obtained by combining the results of two programs, see the supporting Table S1). The middle panel displays the relative contributions to the global discovery of the CTRL untreated whey vs. the combined eluates from the CPLL beads (for protein IDs by Proteome Discoverer *via* Mascot engine, see the supporting Table S2). It is important to highlight that the protein visibility in the untreated fluid is rather meagre (57 total species) vs. 449 unique gene products identified in the CPLL eluates, 54 proteins being common to both samples. Interestingly, the loss of proteins from the original sample is rather minute, barely 3 species, i.e. less than 1% of the species present, which means that the CPLL technique has been much improved over the years, considering that at the inception the loss of species in the initial sample could have been as much as 20% (see also the discussion). The lower panel instead displays the contribution of each capturing pH value. The diagram reveals that almost all of the identified proteins were obtained at pH 2.2 and 7.2, covering about 97% of the total. Moreover, the number of exclusive proteins (proteins revealed only at a particular pH) was relatively similar at pH 2.2 (113 out of 449) and pH 7.2 (103 out of 449), whereas the exclusive contribution of pH 9.0 was very low (14 out of 449). Finally, as expected, the level of redundancy (common proteins among the variouseluates) was significantly smaller between the two extreme pH 2.2 and 9.0 (only 10 common protein identifications) than between pH 7.2 and either of the other two conditions.

Figure 4 gives the taxonomic classification of the 452 proteins identified in whey, displayed in a pie chart. The proteins pertaining to the goat are only 10%, ovine proteins represent 52% of the total and those species homologous to bovine are as many as 37%. This “odd” distribution is due to the paucity of entries in the goat’s database (see also the discussion).

Figure 5 depicts the categorizing of the set of 452 unique gene products identified into: molecular function (5A), cellular components (5B) and biological processes (5C).

4. Discussion

There are some interesting aspects of the present report worth elaborating upon.

4.1. Focus on the methodology

The unique results here obtained in exploring in depth GM (a total of 452 unique gene products) seem to be due to the joining of important methodologies. The first one is the well-ingrained CPLL sample pre-treatment, which here has given an outstanding harvest, as the total discovery of 99% due to the CPLLs alone (Fig. 3, middle panel) demonstrates. This means an increment of eight-fold in the discovery, which we can mostly attribute to low- to very-low abundance proteins (LAPs), undetectable in the absence of CPLL treatment. The cause for such vast increment is two-fold: to start with, the capture performed at three different pH values, namely the standard one at pH 7.2, plus two additional processes, at acidic (pH 2.2) and alkaline (pH 9.0) pH values, which enlarges the harvesting ability of CPLLs. Indeed, as reported in the lower Venn diagram of Figure 3, showing the distribution of detected proteins in the three different CPLL eluates, the adopted strategy is indispensable to capture the greatest number of proteins from a complex sample. The eluate at the standard physiological pH 7.2 allowed the identification of 312 unique gene products, corresponding to 69.5% of the total identifications, whereas the remaining 137 protein species represent the contribute of the two additional acidic and alkaline eluates. It is important to note that the efficiency in harvesting proteins at different pH depends on a number of parameters and it is strictly related to the characteristics of protein species present in the starting mixture. In our case, by the comparison of proteins identified at three different pH values, it

appears that the best capture efficiency with CPLLs has occurred at pH 2.2 and pH 7.2, with a lower capture at pH 9.0, which contributed with only 14 (3.1%) exclusive proteins. But, most importantly, such a very large discovery seems to be due to the massive overloading of sample onto the bead library. It should be noted that we have applied a total of six-grams of milk proteins, divided into three lots of two-grams each for each capture with 100 µL of hexapeptide beads. This is an exaggerated overloading, hardly compatible with a standard chromatographic process, in which the column would rapidly go to saturation and simply adsorb the most abundance species, due to the law of mass action [31, 32]. On the contrary, due to the bio-affinity process regulating the CPLL action mechanism, each protein species should be captured in a very limited amount, the one saturating the specific ligand, a mechanism drastically cutting the level of high-abundance species while substantially enriching the rare species. To the point that the latter's visibility can be augmented by three orders of magnitude and, in the most favourable cases, up to four orders. It is small wonder that these rare species are now fully visible to MS analysis. A very similar result occurred when we analyzed the cytoplasmic proteome of human red blood cells [33]. There too we applied to the CPLL beads six-grams total proteins (here it should be considered that 98% of the total protein mass was indeed due to a single protein, hemoglobin, thus such massive overloading was a must). In turns, there too we could detect 1576 unique gene products whereas the best literature data published at that time listed barely 252 cytoplasmic proteins and this highlights the strong difference of CPLL sample pre-treatment vs. immune-depletion, as largely adopted for eliminating the most abundant species from human sera and in general biological liquids. According to Tu et al. [34], immuno-depletion does not seem to be the right choice for biomarker discovery or for bringing to the limelight rare species. Quoting them: "either top 7 or top 14 immuno-depletion resulted in a 25% increase in identified proteins compared to unfractionated plasma.

Although 23 low-abundance (<10 ng/mL) plasma proteins were detected, they accounted for only 5-6% of total protein identifications in immuno-depleted plasma. In both unfractionated and immuno-depleted plasma, the 50 most abundant plasma proteins accounted for 90% of cumulative spectral counts and precursor ion intensities, leaving little capacity to sample lower abundance proteins". Their conclusions: "Untargeted proteomic analyses using current LC-MS/MS platforms with immuno-depletion cannot be expected to discover low-abundance, disease-specific biomarkers in plasma". But it could be worse! According to Zhi et al. [35] immuno-depletion permits to see 10% less proteins than in control sera! It is thus seen that, finding a handful of LAPs (23 of them) on the vast ocean of plasma where, according to the Anderson's [36], more than one million species might be swimming around, is no big deal. That the removal of high-abundance proteins (HAPs) would not really solve the problem of scarce visibility of the low-abundance ones (LAPs) is also apparent here and in most papers dealing with the exploration of the proteomes of milk from different animal sources. In most cases, for improving the results, the analysis is not performed on whole milk, but rather on whey, i.e. on milk devoid of caseins, which by far are the most abundant species in this biological fluid (ratio casein to whey 80:20 in terms of protein mass, in cow milk). Yet, in whey, as in the present report, an analysis of the untreated control generated a total list of barely 57 IDs, a meagre discovery and a quite poor representation of the real goat milk proteome. This might also be due to the fact that, even after casein depletion, the situation in whey in terms of dynamic range is not much improved since here too only two proteins, α-lactalbumin and β-lactoglobulin still constitute 75% of the total protein mass.

The huge discovery has been also made possible thanks to the unique scan rate, very sensitive high resolution and high mass accuracy of the Orbitrap Fusion Tribrid mass spectrometer used to obtain mass spectral data. But there is another additional

methodology adopted that has also considerably expanded the number of identified species, as gleaned from the upper panel of Fig. 3, namely of having used, for identification of species, not only the Mascot program, but also the PEAKS *de novo* sequencing software, as developed by Zhang et al. [37]. In fact, it should be noted that the database "*Ruminantia*" presents only 2.9% of proteins belonging to *Capra*. Therefore, while characterization of proteins from *Capra* using the standard Mascot database search engine was accomplished, it also became necessary to generate *de novo* peptide sequences used for database search. Particularly, the use of the integrated tool SPIDER, which accounts for possible errors in *de novo* sequencing, and it has been designed to find potential sites of amino acid mutations in novel peptides that are homologous to peptides in a given protein database (cross-species database searching). On this respect, it can be appreciated that, with the second round of identifications via PEAKS, a considerably larger number of unique gene products has been identified, i.e. a total of 148 additional compounds, found solely via this program (206 IDs being in common between the two software's, and 98 specific of only Mascot) (see the supporting Tables 1S and 2S). By the comparison of the two supporting Tables, it can be noted that for many species common between the two approaches, the PEAKS allowed to increase the sequence coverage by the matching of novel peptides previously undetected via Mascot. Moreover, the SPIDER tool identified potential sites of amino acid mutations in 206 unique gene products, 82 of which exclusively found via this approach (see the supporting Table 1S and PEAKS report). It is important to note that the identification by *de novo* sequencing of many peptides with potential amino acid mutations is a consequence of the fact that 89% of the identified proteins belong to ovine or bovine species. On the other hand, also proteins pertaining to the goat were identified with potential sites of amino acid mutations. The presence of

a number of isoforms of the same protein may reflect the "bulk" composition of the milk here investigated, which has been collected from one hundred animals.

Another interesting result that can be appreciated from the middle Venn diagrams of Fig. 3 is highlighted too. It is seen here that the proteins present in the control, untreated sample but lost after CPLL capture are only three in number, i.e. less than 1% of the total protein population, a truly marginal loss. This was not so when the CPPL technique was first reported in 2005, when the losses lamented were up to 20% of the content of the initial sample, a severe limitation of CPLLs. The quite superior results here reported are due to continuous refinements of the methodology, as reported over the years. One of them has just been discussed, namely the capture implemented at three different pH values, instead of the conventional one at physiological pH values. To further improve the efficiency of such libraries, additional modifications to the initial protocols have been proposed. A succinylated variant has been described [33].28 A reduction of the ionic strength during the capture process contributed also to improved results as reported by Di Girolamo et al. [38]. Another important improvement has been focused on a more complete protein elution as described by Candiano et al. [29], in the presence of 4% SDS added with 3% DTE. Moreover, in attempting at driving the capture towards a more hydrophobic class of compounds, yet another variant has been introduced in the presence of high concentrations of lyotropic salts of the Hofmeister series (1 M ammonium sulphate) favouring hydrophobic interaction [39]. All of those refined protocols have now led to such unique performance of CPLL beads, thus rendering them a formidable tool for in depth exploration of any proteome.

4.2. Focus on biology

An interesting biological aspect of our data can be derived from Fig. 4, which gives the taxonomic classification of the discovered proteome of GM. Only about 10% of the

435 species have been found to belong to the goat, with an additional 52% being specific of
436 sheep milk. Yet, the proteins homologous to bovine milk still represent a sizable portion,
437 i.e. 37%. This distribution is largely due to the presence of specific entries in the
438 database "*Ruminantia*", where, out of a grand total of 91884 entries, 54983 (60%) are
439 specific of *Bos*, 27483 (30%) belong to the *Ovis* and only 2661 (2.9%) pertain to the
440 *Capra*, the remaining 7.1% (6757) belonging to other generic ruminants. Thus it has not
441 been possible to expand the number of proteins specific of the goat, due to paucity of
442 specific data.
443 Gene Ontology classification of the set of 452 unique gene products here identified
444 allowed to ascertain their molecular function, cellular location and the biological
445 processes in which they are involved (see Figure 5). Most of them play catalytic activity
446 and protein binding, are mainly located in the extracellular region and are principally
447 involved in metabolic processes, regulation of biological processes and response to
448 stimulus.
449 This list, which represents by far the most comprehensive description of GM proteome,
450 could serve as a starting point for a tentative comparison with CM proteins as reported
451 by D'Alessandro et al. [5], who compiled an exhaustive list of 573 non-redundant
452 annotated protein entries present in bovine milk. On this respect, first of all, each
453 component classified as "uncharacterized protein" has been submitted to a sequence
454 similarity search by BLAST (Basic Local Aligment Search Tool). By this approach, it has
455 been possible to classify the uncharacterized proteins by finding homologous proteins
456 present in databases, which generally share more than 85% sequence identity. Then,
457 all the 452 unique gene products have been grouped in 362 protein families, also taking
458 into account, when available, the corresponding gene code. Finally, they were
459 compared with proteins from bovine milk (see Table 1). The comparison reveals that
460 GM proteome shares about 45% (164 out of 362) of components with the corresponding

461 bovine milk proteome. On the contrary, the remaining 198 proteins (i.e. 55%) are
462 specific of GM, even if it is important to remind the old adage that "*absence of evidence*
463 *is not evidence of absence*". In fact, it is well known that milk protein composition is
464 strictly related to the stage of lactation, phenotype, degree of proteolysis by indigenous
465 milk enzymes and, last but not least, to the condition of the mammary gland. Thus,
466 proteins of blood serum origin (i.e. serotransferrin, BSA etc.) has been observed in the
467 whey from animals with clinical mastitis [40].
468 Among the 164 shared proteins, 7 belong to known bovine or caprine allergens, and
469 correspond to the major components of milk (i.e. caseins, alpha-lactalbumin, beta-
470 lactoglobulin and albumin), whereas two GM components are homologues to CM
471 allergens. In detail, our data allowed to identify: i) an uncharacterized protein from *Ovis*
472 *aries* (Acc. N. W5P1L7), sharing 93% of identity and 95% of similarity with the
473 prothrombin (Acc. N. P00735, *Bos d Thrombin*), a bovine allergen originating from
474 blood; and ii) the homologue caprine counterpart (Acc. N. Q29477) of bovine allergen
475 *Bos d LF* (i.e. lactotransferrin, Acc. N. P24627).
476 Table 1 also shows that proteome of GM presents several components which are
477 homologues to bovine milk glyco-proteins relatively resistant against proteolysis in the
478 gastrointestinal tract, and playing an important physiological role in the
479 defence/immunity mechanisms. Among these proteins, we found lactoperoxidase and
480 Milk fat globule-EGF factor 8/lactadherin. Lactoperoxidase is known may contribute to
481 the defense against both gram-positive and gram-negative pathogenic bacteria, and it
482 has been used by the dairy industry in developing countries for decades to preserve
483 microbial quality. Milk fat globule-EGF factor 8/lactadherin prevents symptomatic
484 rotavirus infection in breast-fed infants, and more in general, it seems positively interact
485 with damaged intestinal epithelium. Thus, lactadherin could have a potential role in the
486 prevention and treatment of intestinal injury in infants [41].

The largest group of goat proteins that, up to the present, haven't been found in bovine milk, comprises components that: i) are precursor of peptides involved in potential hypotensive effects (i.e. angiotensinogen); ii) may mediate inflammation and tumor progression (i.e matrix metalloproteinase); iii) are involved in the normal mammary gland development (i.e. dystroglycan); iv) contribute to the immunological protection of the fetus (factor H); or v) may be classified as defense factors (complement component C2 and C6). Moreover, the identification of proteins belong to or related to complement system (vitronectin, factor H, fibulin-1, peroxiredoxin 2) confirms the anti-inflammatory properties of goat milk and its function in the regulation of immuno response for the maintenance of immune homeostasis [19]. The group of proteins only found in GM also includes the haemoglobin subunit beta from *ovis* (Acc. N. P02075), which corresponds to the homologue bovine, a blood-derived protein identified in red meat as partly muscle-specific and heat-resistant allergen (*Bos d HG,* Acc. N. P02070). Primary structure comparison of these two proteins reveals that they share 93% of identity and 95% of similarity, showing only ten amino acid point mutations, and it should be hypothesized that also the homologue caprine may represent a potential minor allergen.

5. Conclusions

Despite many efforts, up to the present a complete knowledge of the composition and functions of milk proteins has been hampered by partial characterization of the milk proteomes. With respect to goat, not much was known about the repertoire of low-abundance proteins of the milk of this farm animal. In the current study, by coupling the power of the Combinatorial Peptide Ligand Library (CPLL) technology, SDS-PAGE separation, high resolution nLC-nESI MS/MS and database searching, we have been able to identify 452 unique gene products in GM. The unique results here reported are probably related to the well-ingrained CPLL sample pre-treatment, implemented at three

different pH values instead of the conventional one at physiological pH values, together with the extraordinary loading of whey proteins onto the CPLL beads. A considerable enlargement of the number of identified species relies on the acquisition of data in very sensitive high resolution/high mass accuracy and is also due to the complementary approaches used for database searching, namely the Mascot engine and the PEAKS *de novo* sequencing software. Indeed, taking into account the paucity of goat's entries, the largest proportion (89%) of our identification was obtained by cross-species database searching, obtaining 52% of proteins from *Ovis* and 37% from *Bos*, the two species presenting the closest phylogenetic relationships with goat. A preliminary comparison of the qualitative composition of goat and cow milk proteome reveals that they share about 45% of their protein fraction, including major allergens but also proteins with potential beneficial properties. The other 55% seems to be exclusive of goat milk and comprises an heterogeneous group of proteins most of all deriving adventitiously from blood, somatic or epidermal cells, rather than the mammary gland. The origin and role of these minor proteins, with diverse functions and occurring in goat milk, need to be deeply explored but remain absolutely interesting.

Supplementary data

Table S1: List of the 452 unique gene products identified by the use of the two complementary tools, Mascot and PEAKS. When available, gene ontology annotation is also reported. Table S2: List of IDs, and peptides matched, by the use of Proteome Discoverer *via* Mascot engine. Report PEAKS: Results obtained by the use of PEAKS software, including results statistics, list of identified proteins, peptide matching and potential amino acids mutation found by SPIDER tool.

Acknowledgments

This work was supported by a grant from PO FERS 2007/13 4.1.2.A, project "Piattaforma regionale di ricerca translazionale per la salute".

Legends

Figure 1

Scheme of the pre-fractionation protocol of goat whey with CPLLs. The capture is performed at three pH values, namely 2.2, 7.2 and 9.0. Note the massive protein load for each capture, two grams for each pH value against barely 100 µL bead volume.

Figure 2

SDS-PAGE of the control (CTRL) sample vs. the eluates of CPLL capture at the three different pH values (tracks 2.2, 7.2 and 9.0). Each lane has been cut into twelve segments, the proteins digested and sent to MS analysis. Mr: molecular mass ladder. Staining with micellar Coomassie blue.

Figure 3

Venn diagrams giving the statistics of the identified proteins. Upper panel: assessment of protein IDs as obtained by the Mascot and Peaks software's, respectively. The IDs common to the two programs are 206 unique gene products. Middle panel: Contribution to the total identifications as obtained in the control, untreated whey (CTRL) and in the three combined treatments with CPLL beads. This last capture incremented the global discovery by 800%. Lower panel: Contribution to the total identifications as obtained in the eluates from the pH 2.2, 7.2 and 9.0 CPLL captures. Outside the circles are the total numbers of proteins at a given pH (indicated inside parenthesis). Inside the diagram the exclusive proteins from the three eluates are reported. At the center are common identifications from the various eluates. All other numbers are common identifications between two close fractions. Inside the diagram the percent values are also reported.

Figure 4

Pie chart of the taxonomic classification of the global 452 unique gene products identified. The proteins specific to the goat are 9%, those of ovine origin are 52% and those homologous to *Bos taurus* are 37%.

Figure 5

Pie charts of (A) molecular functions; (B) Cellular components and (C) biological processes pertaining to the 452 unique gene products described in goat's whey.

578

579 **Table 1** – Protein Family identified in Goat Milk: gene code, description of protein family and presence of the
580 corresponding homologue protein family in cow's milk (CM) are reported. Asterisks indicate bovine proteins
581 known resisting against gastrointestinal digestion. Circles indicate proteins known as goat milk allergens. Crosses
582 indicate proteins known as cow milk allergens.

N.	Gene	Description of Protein Family	Found in CM
1	A1BG	Alpha-1B-glycoprotein	YES
2	A2M	Alpha-2-macroglobulin	YES
3	ABCG2	ATP-binding cassette sub-family G member 2	YES
4	ACO1	Aconitate hydratase	YES
5	ACTB	Beta-actin	YES
6	ADFP	Perilipin	YES
7	ADI1	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	-
8	AEBP1	Uncharacterized protein (Fragment) (Acc. N. W5Q3W1) similar to Adipocyte enhancer-binding protein 1	-
9	AFM	Uncharacterized protein (Acc.N. W5PXI6) similar to Afamin	-
10	AGRN	Uncharacterized protein (Fragment) (Acc. N. F1MSI2) similar to Agrin	-
11	AGT	Angiotensinogen (Fragment)	-
12	AHSG	Alpha-2-HS-glycoprotein	YES
13	AKR1A1	Uncharacterized protein (Acc. N. W5NUN8) similar to Alcohol dehydrogenase (NADP+)	YES
14	ALB	Albumin ° ×	YES
15	ALDH1A1	Retinal dehydrogenase 1	-
16	ALDOC	Fructose-bisphosphate aldolase (Fragment)	-
17	AMBP	Uncharacterized protein (Acc. N. W5P812) similar to Alpha-1-microglobulin, AMBP, trypstatin	YES
18	APOA1	Uncharacterized protein (Acc. N. W5NX51) similar to Apolipoprotein A-1	YES
19	APOA1BP	Uncharacterized protein (Fragment) (Acc. N. W5P816) similar to NAD(P)H-hydrate epimerase	-
20	APOE	Apolipoprotein E	YES
21	ARF3	ADP-ribosylation factor 3	-
22	ARSG	ARSG protein	-
23	art5	NAD(P)(+)--arginine ADP-ribosyltransferase (Fragment)	-
24	ASL	Argininosuccinate lyase	-
25	ATP6AP1	V-type proton ATPase subunit S1	YES
26	ATP6AP2	Uncharacterized protein (Acc. N. W5NT93) similar to Renin receptor	YES
27	AZGP1	Zinc-alpha-2-glycoprotein	YES
28	AZU1	Uncharacterized protein (Acc. N. W5PIP4) similar to Azurocidin	-
29	B4GALNT	Uncharacterized protein (Fragment) (Acc. N. W5PRD6) similar to N-acetyl-beta-glucosaminyl-glycoprotein 4-beta-N-	-
30	B4GALT1	Beta-1,4-galactosyltransferase 1	YES
31	BGN	Biglycan	-
32	BTD	Uncharacterized protein (Acc. N. W5Q0V2) similar to Biotinidase	YES
33	BTN1A1	Butyrophilin subfamily 1 member A1	YES
34	C1R	Uncharacterized protein (Acc. N. W5P336) similar to Complement C1r subcomponent	-
35	C2	Complement component C2 (Fragment)	-
36	C3	Complement component 3	YES
37	C4A	Uncharacterized protein (Acc. N. E1BH06) similar to Complement component 4	YES
38	C5	Uncharacterized protein (Acc. N. W5P6F4) similar to Complement component 5	YES
39	C6	Complement component C6	-
40	C7	Uncharacterized protein (Acc. N. W5PH81) similar to Complement component 7	YES
41	C8A	Uncharacterized protein (Acc. N. W5PDR5) similar to Complement component C8 alpha chain	-
42	C8B	Uncharacterized protein (Fragment) (Acc. N. W5PE53) similar to Complement component 8, beta chain	YES
43	C8G	C8G protein	-
44	C9	Complement component C9	YES
45	CALM2	Calmodulin 2 (Fragment)	-
46	CANT1	Uncharacterized protein (Acc. N. E1BGL5) similar to Calcium-activated nucleotidase 1	YES
47	CATHL1	Preprocathelicidin antimicrobial peptide (Fragment)	YES
48	CATHL2	Cathelicidin-2	YES
49	CD109	Uncharacterized protein (Fragment) (Acc. N. W5P8E9) similar to CD109 antigen	YES
50	CD14	Monocyte differentiation antigen CD14	YES
51	CD36	CD36 molecule	YES
52	CFD	Uncharacterized protein (Acc. N. W5PJ66) similar to Complement factor D	YES
53	CFI	Uncharacterized protein (Acc. N. W5P5I3) similar to Complement factor I	YES
54	CHI3L1	Chitinase-3-like protein 1	YES
55	CHKB	Uncharacterized protein (Acc.N. G3X782) similar to Choline/ethanolamine kinase	-
56	CLEC3B	Tetranectin	YES
57	CLU	Clusterin	YES
58	CP	Ceruloplasmin	YES
59	CPB2	Uncharacterized protein (Acc. N. W5PD62) similar to Carboxypeptidase B2	YES
60	CPQ	Carboxypeptidase Q	-

61	CREG1	Uncharacterized protein (Acc. N. W5PRG8) similar to Cellular repressor of E1A-stimulated genes 1	YES
62	Crisp3	Cysteine-rich secretory protein 3	YES
63	CRK	Uncharacterized protein (Acc. N. E1BQ32) similar to Adapter molecule crk	-
64	CRP	Uncharacterized protein (Acc. N. W5PD71) similar to C-reactive protein	-
65	CSN1S1	Alpha-S1-casein ° [x]	YES
66	CSN1S2	Alpha-S2-casein ° [x]	YES
67	CSN2	Beta-casein ° [x]	YES
68	CSN3	Kappa-casein * ° [x]	YES
69	CST3	Uncharacterized protein (Acc. W5P887) similar to Cystatin C	YES
70	CTSA	Lysosomal protective protein	-
71	CUTA	Uncharacterized protein (Fragment) (Acc. N. W5PJA0) similar to Protein CutA	-
72	DAG1	Dystroglycan	-
73	DDR1	Discoidin domain receptor family, member 1	YES
74	DKK3	DKK3 protein	-
75	ECM1	Uncharacterized protein (Acc. N. W5QI29) similar to E matrix protein 1	YES
76	ENO1	Uncharacterized protein (Acc. W5PIG7) similar to Alpha Enolase	YES
77	ERP29	ER resident protein 29	-
78	EXT2	Uncharacterized protein (Acc. N. W5NU84) similar to Exostosin-2	-
79	EZR	Uncharacterized protein (Fragment) (Acc. N. W5P2V0) similar to Ezrin, villin 2	YES
80	F2	Uncharacterized protein (Acc. N. W5P1L7) similar to C factor II, prothrombin x	YES
81	FAM20A	Uncharacterized protein (Fragment) (Acc. N. W5Q021) similar to Protein FAM20A	-
82	FAM20C	Uncharacterized protein (Fragment) (Acc. N. F1MXQ3) similar to Dentin matrix protein 4	-
83	FCGBP	Uncharacterized protein (Acc. N. W5P8R7) similar to IgGFc-binding protein	-
84	FCGR2C	Fc-gamma-RII-D	-
85	FETUB	Uncharacterized protein (Acc. N. W5QH54) similar to Fetuin-B	-
86	FGG	Uncharacterized protein (Acc. N. W5Q5A6) similar to Fibrinogen gamma chain	YES
87	Fh	Factor H (Fragment)	-
88	FUCA1	Tissue alpha-L-fucosidase	YES
89	FUT11	Uncharacterized protein (Acc. N. W5PE31) similar to Alpha-(1,3)-fucosyltransferase 11	-
90	GAA	Lysosomal alpha-glucosidase	-
91	GALM	Aldose 1-epimerase	YES
92	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	YES
93	GC	Uncharacterized protein (Fragment) (Acc. N. W5PTH1) similar to Vitamin D-binding protein	YES
94	GDI1	Uncharacterized protein (Fragment) (Acc. N. W5P3E8) similar to Rab GDP dissociation inhibitor alpha	YES
95	GFRA	Uncharacterized protein (Acc. N. W5PLU8) similar to GDNF family receptor alpha-2	-
96	GGT6	Gamma-glutamyltransferase 6	-
97	GLA	Uncharacterized protein (Acc. N. W5NU86) similar to Alpha-galactosidase A	-
98	GLG1	Uncharacterized protein (Acc. N. E1BDY3) similar to Golgi apparatus protein 1	-
99	GLUL	Glutamine synthetase	-
100	GLYCAM1	Glycosylation-dependent cell adhesion molecule 1	YES
101	GNPTG	N-acetylglucosamine-1-phosphotransferase subunit gamma	-
102	GOT1	Aspartate aminotransferase	-
103	GPI	Glucose-6-phosphate isomerase	YES
104	GSN	Gelsolin isoform b	YES
105	HBB	Hemoglobin subunit beta x	-
106	H-FABP	Heart fatty acid-binding protein	-
107	HHIPL2	Uncharacterized protein (Acc. N. W5PV41) similar to HHIP-like protein 2	-
108	HPN	HPN protein	YES
109	HPX	Uncharacterized protein (Acc. N. W5QAB1) similar to Hemopexin	YES
110	HSP90AB1	HSP90AB1 protein (Fragment)	YES
111	HSPA13	Uncharacterized protein (Acc. N. W5Q411) similar to Heat shock 70kDa protein 13	YES
112	HSPA1A	Uncharacterized protein (Acc. N. W5PG95) similar to Heat shock 70kDa protein 1A	YES
113	HSPA5	Heat shock 70kDa protein 5 isoform 1	YES
114	HSPA8	Heat shock cognate 71 kDa protein	YES
115	HSPG2	Uncharacterized protein (Fragment) (Acc. N. F1MER7)	-
116	HYI	Uncharacterized protein (Fragment) (Acc. N. W5NQH5) similar to Hydroxypyruvate isomerase	-
117	HYOU1	Uncharacterized protein (Acc. N. W5PSK1) similar to Hypoxia up-regulated protein 1	-
118	IAH1	Uncharacterized protein (Acc. N. W5Q0L0) similar to Isoamyl acetate-hydrolyzing esterase 1 homolog	-
119	IDH1	Isocitrate dehydrogenase 1 (Fragment)	YES
120	IDS	Uncharacterized protein (Acc. N. F1N2D5) similar to Iduronate 2-sulfatase	-
121	IGHE	Uncharacterized protein (Acc. N. W5PGT9) similar to Ig epsilon chain C region	-
122	IGHM	Uncharacterized protein (Fragment) (Acc. N. G5E5T5) similar to Immunoglobulin mu chain C region *	YES
123	IGJ	Uncharacterized protein (Fragment) (Acc. N. W5PPQ8) similar to Immunoglobulin J chain *	YES
124	IGL @	Immunoglobulin light chain lambda gene cluster *	YES
125	IST1	IST1 homolog	-
126	ITFG1	Uncharacterized protein (Acc. N. W5Q3P7) similar to Integrin alpha FG-GAP repeat containing 1	-
127	ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	YES
128	ITIH2	Uncharacterized protein (Acc. N. W5PW21) similar to Inter-alpha-trypsin inhibitor heavy chain H2	YES
129	ITIH3	Uncharacterized protein (Acc. N. W5NSH2) Inter-alpha-trypsin inhibitor heavy chain H3	-
130	ITIH4	Uncharacterized protein (Acc. N. W5NRG7) similar to Inter-alpha-trypsin inhibitor heavy chain H4	YES
131	KRAS	Uncharacterized protein (Acc. N. E1BMX0) similar to GTPase Kras	YES

132	KRT1	Uncharacterized protein (Fragment) (Acc. N. W5Q611) Keratin, type II cytoskeletal 1	-
133	KRT10	Keratin 10 (Epidermolytic hyperkeratosis; keratosis palmaris et plantaris)	YES
134	KRT14	Keratin type I cytoskeletal 14	-
135	KRT2	Uncharacterized protein (Acc. N. E1B991) similar to Keratin, type II cytoskeletal 2 epidermal	-
136	KRT24	Uncharacterized protein (Fragment) (Acc. N. F1MFW9) similar to Keratin, type I cytoskeletal 24	-
137	KRT3	Uncharacterized protein (Fragment) (Acc. N. G3MXL3) similar to Keratin, type II cytoskeletal 3	-
138	KRT5	Uncharacterized protein (Acc. N. W5Q687) similar to Keratin, type II cytoskeletal 5	-
139	LALBA	Alpha-lactalbumin ° x	YES
140	LAMA4	Uncharacterized protein (Acc. N. W5PIM5) similar to Laminin subunit alpha-4	-
141	LAMB1	Uncharacterized protein (Acc. N. W5PBD8) similar to Laminin subunit beta-1	-
142	LAMC1	Uncharacterized protein (Acc. N. W5QD08) similar to Laminin subunit gamma-1	-
143	LAP3	CYT aminopeptidase	YES
144	LBP	Uncharacterized protein (Acc. N. W5QBW5) similar to Lipopolysaccharide-binding protein	YES
145	LCAT	Lecithin-cholesterol acyltransferase	-
146	LCN2	Uncharacterized protein (Fragment) (Acc. N. W5PMH6) similar to Neutrophil gelatinase-associated lipocalin, lipocalin 2, NGAL	YES
147	LCP1	Uncharacterized protein (Fragment) (Acc. N. W5PDD0) similar to Lymphocyte cytosolic protein 1, L-plastin	YES
148	LDHB	L-lactate dehydrogenase	YES
149	LDLR	Uncharacterized protein (Acc. W5Q887) similar to Low-density lipoprotein receptor	-
150	LGALS3	Galectin	-
151	LGB	Beta-lactoglobulin * ° x	YES
152	LGMN	Uncharacterized protein (Acc. N. W5PUH5) similar to Legumain	-
153	LMAN2	LMAN2 protein	-
154	LOC100125610	Uncharacterized protein (Acc. N. W5PPS4) similar to Elongation factor 1-alpha	YES
155	LOC101110099	Uncharacterized protein (Acc. N. W5PZD0) similar to Secretoglobin family 1D member	YES
156	LOC101111083	Uncharacterized protein (Fragment) (Acc. N. W5NXP3) similar to Serpin A3-5	YES
157	LOC101112066	Uncharacterized protein (Fragment) (Acc. N. W5P3R3) similar to Plasminogen	YES
158	LOC101113086	Amine oxidase	-
159	LOC101116799	Uncharacterized protein (Acc. N. W5PHS2) similar Odorant-binding protein-like	YES
160	LOC101117146	Uncharacterized protein (Acc. N. W5PHP7) similar to Serpin A3-7	YES
161	LOC101120775	Uncharacterized protein (Acc. W5P1W2) similar to Folate receptor alpha	YES
162	LOC101122014	Uncharacterized protein (Acc. N. W5QIU6) similar to C-type lectin domain family 2 member H	-
163	LOC788112	LOC788112 protein	YES
164	LPO	Lactoperoxidase *	YES
165	LRG1	Leucine-rich alpha-2-glycoprotein 1	YES
166	LRRC15	Uncharacterized protein (Acc. N. W5NVT0) similar to Leucine-rich repeat-containing protein 15	-
167	LTA4H	Leukotriene A(4) hydrolase	-
168	LTF	Lactotransferrin x	YES
169	m6p/igf2r	Mannose-6-phosphate/insulin-like growth factor II receptor (Fragment)	-
170	M91_00876	Calumenin (Fragment)	-
171	M91_01352	Dickkopf-related protein 3 (Fragment)	-
172	M91_03148	Rho-related GTP-binding protein RhoC (Fragment)	-
173	M91_03372	Alpha-aminoadipic semialdehyde dehydrogenase	-
174	M91_04257	Ras-related protein Rab-14 (Fragment)	-
175	M91_04353	Rho GDP-dissociation inhibitor 2	-
176	M91_04808	ERO1-like protein alpha (Fragment)	-
177	M91_05027	Iduronate 2-sulfatase (Fragment)	-
178	M91_05160	Charged multivesicular body protein 4b	-
179	M91_05756	Dipeptidyl peptidase 4	-
180	M91_07836	Ig gamma-1 chain C region (Fragment)	YES
181	M91_07998	Annexin	YES
182	M91_08147	Ig gamma-3 chain C region (Fragment)	YES
183	M91_08149	Ig alpha-1 chain C region (Fragment)	YES
184	M91_08924	Plasma glutamate carboxypeptidase (Fragment)	-
185	M91_10626	Keratin, type II cytoskeletal 7	-
186	M91_10641	Keratin, type II cytoskeletal 6A	YES
187	M91_11805	Acyl-protein thioesterase 1 (Fragment)	-
188	M91_11915	Inhibitor of carbonic anhydrase (Fragment)	-
189	M91_12334	Fibulin-1 (Fragment)	-
190	M91_12826	GDH/6PGL endoplasmic bifunctional protein (Fragment)	-
191	M91_12888	Transcobalamin-2 (Fragment)	-
192	M91_13458	Uncharacterized protein (Acc. N. L8HT95) similar to Ig lambda chain V-IV region Bau	-
193	M91_13695	15 kDa protein A	-
194	M91_14042	Glutathione S-transferase P (Fragment)	-
195	M91_14230	Phospholipase D3	-
196	M91_14602	Purine nucleoside phosphorylase (Fragment)	YES
197	M91_14679	14-3-3 protein epsilon (Fragment)	YES
198	M91_14691	Pigment epithelium-derived factor	-
199	M91_15862	Metalloproteinase inhibitor	-
200	M91_16070	Calcium and integrin-binding protein 1	YES
201	M91_16468	Ras-related protein Rap-2c	-
202	M91_17207	Renin receptor (Fragment)	-

203	M91_17340	Rho GDP-dissociation inhibitor 1	YES
204	M91_17650	Triosephosphate isomerase	-
205	M91_17790	Macrophage-capping protein (Fragment)	-
206	M91_18159	Polyubiquitin-C (Fragment)	-
207	M91_18249	Phosphatidylcholine-sterol acyltransferase	-
208	M91_18800	Brain-specific serine protease 4 (Fragment)	-
209	M91_19261	UDP-glucuronic acid decarboxylase 1 (Fragment)	-
210	M91_19493	Sulfatase-modifying factor 2 (Fragment)	-
211	M91_19736	Protein canopy-like protein 2 (Fragment)	-
212	M91_19748	ATP synthase subunit beta (Fragment)	-
213	M91_20802	Ectonucleoside triphosphate diphosphohydrolase 6	-
214	M91_20862	Ephrin-A1 (Fragment)	-
215	MAN2A2	Alpha-mannosidase	YES
216	MANBA	Beta-mannosidase	-
217	map28	MAP28 protein	-
218	MASP1	Uncharacterized protein (Acc. N. F1MVS9) similar to Mannan-binding lectin serine peptidase 1	YES
219	MASP2	Uncharacterized protein (Acc. N. W5P229) similar to Mannan-binding lectin serine protease 2	-
220	MDH1	Malate dehydrogenase CYPic	YES
221	MFI2	Uncharacterized protein (Acc. N. W5QGM9) similar to Melanotransferrin	-
222	MIF	Macrophage migration inhibitory factor	-
223	MMP2	Matrix metallopeptidase 2	-
224	MPO	MPO protein	YES
225	MST1	Uncharacterized protein (Fragment) (Acc. N. W5PUG1) similar to Hepatocyte growth factor-like protein	-
226	mstn	Myostatin	-
227	MUC15	Mucin 15 CS associated	YES
228	MUC20	Uncharacterized protein (Acc. N. W5QGD9) similar to Mucin-20	-
229	N/A	Adenosylhomocysteinase (Fragment)	-
230	N/A	Angiopoietin-like protein 4 (Fragment)	-
231	N/A	Beta-2-microglobulin	YES
232	N/A	Beta-hexosaminidase	YES
233	N/A	Cathepsin B	YES
234	N/A	Cathepsin F	-
235	N/A	Fructose-bisphosphate aldolase (Fragment)	YES
236	N/A	GDI2	-
237	N/A	Lipoprotein lipase	YES
238	N/A	Matrix metalloproteinase-9	-
239	N/A	MHC class I heavy chain (Fragment)	-
240	N/A	Milk fat globule EGF factor 8 protein *	YES
241	N/A	MMP-2 protein (Fragment)	-
242	N/A	Monocyte differentiation antigen CD14	-
243	N/A	N-acetylgalactosaminidase alpha	-
244	N/A	Peptidyl-prolyl cis-trans isomerase (Fragment)	YES
245	N/A	Prokineticin-1 (Fragment)	-
246	N/A	Putative uncharacterized protein	-
247	N/A	Ras-related protein RAB11A (Fragment)	YES
248	N/A	Retinol-binding protein 4	YES
249	N/A	Serum amyloid A protein	YES
250	N/A	Thrombospondin-1 (Fragment)	-
251	N/A	Uncharacterized protein (Acc. N. F1MH40) similar to IGK protein	YES
252	N/A	Uncharacterized protein (Acc. N. W5P8F9)	-
253	N/A	Uncharacterized protein (Acc. N. W5P9V5) similar to Polymeric immunoglobulin receptor	YES
254	N/A	Uncharacterized protein (Acc. N. W5QHZ5) similar to Ig kappa chain C region	-
255	N/A	Uncharacterized protein (Fragment) (Acc. N. W5PDQ0) similar to Mucin	YES
256	N/A	Uncharacterized protein (Fragment) (Acc. N. W5PG90) similar to Protease, serine, 8	YES
257	N/A	Uncharacterized protein (Fragment) (Acc. N. W5QE23)	-
258	N/A	Uncharacterized protein (Fragment) (Acc. N. W5QHZ8) similar to Ig kappa chain V-I region HK102	-
259	N/A	Very low density lipoprotein receptor VLDL-R2	-
260	N/A	Vitronectin (Fragment)	-
261	NAGLU	NAGLU protein	-
262	NID1	Uncharacterized protein (Fragment) (Acc. N. W5P094) similar to NID1 protein	-
263	NME2	Nucleoside diphosphate kinase B	YES
264	NUCB1	Uncharacterized protein (Acc. N. W5PS94) similar to Nucleobindin 1	YES
265	NUCB2	Uncharacterized protein (Acc. N. W5P1C2) similar to Nucleobindin 2	YES
266	OGFOD3	Uncharacterized protein (Acc. N. W5Q560) similar to PKHD domain-containing transmembrane protein	-
267	OS9	Uncharacterized protein (Acc. N. W5P6E0) similar to Osteosarcoma amplified protein 9, endoplasmic reticulum lectin	YES
268	P4HB	Protein disulfide-isomerase	YES
269	PABPC1	Polyadenylate-binding protein 1	-
270	PCMT1	Protein-L-isoaspartate O-methyltransferase (Fragment)	-
271	PCOLCE	Uncharacterized protein (Acc. N. W5Q517) similar to Procollagen C-endopeptidase enhancer	-
272	PDIA6	PDIA6 protein (Fragment)	-
273	PEBP1	Phosphatidylethanolamine-binding protein 1	YES

274	PEPD	Uncharacterized protein (Acc. N. W5P2V3) similar to Xaa-Pro dipeptidase	-
275	PFN1	Profilin-1	YES
276	PGK1	Phosphoglycerate kinase	-
277	PGLS	6-phosphogluconolactonase	YES
278	PGLYRP1	Peptidoglycan-recognition protein	YES
279	PGM1	Uncharacterized protein (Fragment) (Acc. N. W5PJB6) similar to Phosphoglucomutase-1	-
280	PLBD2	Uncharacterized protein (Acc. N. W5PCE0) similar to Putative phospholipase B-like 2	-
281	PLOD1	Procollagen-lysine 2-oxoglutarate 5-dioxygenase 1	-
282	PLTP	Phospholipid transfer protein	-
283	pofut2	Protein O-fucosyltransferase 2b	-
284	PPIA	Peptidyl-prolyl cis-trans isomerase	YES
285	PPIB	Peptidyl-prolyl cis-trans isomerase B	-
286	PPP1R7	Uncharacterized protein (Fragment) (Acc. N. W5Q9D0) similar to Protein phosphatase 1 regulatory subunit 7	-
287	PRCP	Uncharacterized protein (Acc. N. W5P726) similar to Lysosomal Pro-X carboxypeptidase	-
288	PRDX2	Peroxiredoxin 2	-
289	PRDX4	Uncharacterized protein (Fragment) (Acc. N. W5Q532) similar to Peroxiredoxin-4	-
290	PRDX6	Peroxiredoxin-6	-
291	PROM2	Uncharacterized protein (Fragment) (Acc. N. W5PXF5) similar to Prominin-2	-
292	PrP	Major prion protein	-
293	PRSS22	Uncharacterized protein (Acc. N. W5NPP5) similar to Protease, serine, 22-like	-
294	PRSS27	Uncharacterized protein (Acc. N. W5QBI0) similar to Serine protease 27-like protein	-
295	PSAP	Uncharacterized protein (Acc. N. W5PAJ2) similar to Prosaposin, co-beta glucosidase	YES
296	PSMB1	Proteasome subunit beta type	YES
297	PTGDS	Prostaglandin-H2 D-isomerase	YES
298	PTX3	Uncharacterized protein (Acc. N. W5NXM6) similar to Pentraxin-related protein PTX3	YES
299	PYCARD	PYD and CARD domain	-
300	QSOX1	Sulfhydryl oxidase	YES
301	RAB11B	Ras-related protein Rab-11B	YES
302	RAB18	Ras-related protein Rab-18	YES
303	RAB1A	Uncharacterized protein (Fragment) (Acc. N. W5QFH5) similar to Ras-related protein Rab1a	YES
304	RAB1B	Ras-related protein Rab-1B (Fragment)	YES
305	RAB2A	Ras-related protein Rab-2A	YES
306	RAB5A	Uncharacterized protein (Acc. N. W5Q2D9) similar to Ras-related protein Rab5a	YES
307	RAB5B	Uncharacterized protein (Fragment) (Acc. N. W5PLV3) similar to Ras-related protein Rab5b	YES
308	RAB5C	Ras-related protein Rab-5C	YES
309	RAB6B	Ras-related protein Rab-6B	-
310	RAB7A	Ras-related protein Rab-7a	YES
311	RAC3	RAC3 protein	-
312	RAP1B	Ras-related protein Rap-1b	YES
313	RAP2B	RAP2B protein	-
314	RBP1	Retinol binding protein 1, cellular	-
315	RCN2	Reticulocalbin 2 EF-hand calcium binding domain	-
316	RET	Uncharacterized protein (Fragment) (Acc. N. W5P1A8) similar to Proto-oncogene tyrosine-protein kinase receptor Ret	-
317	RGMB	Uncharacterized protein (Fragment) (Acc. N. F1MFY9) similar to RGM domain family, member B	-
318	RHOA	Transforming protein RhoA	YES
319	RNASE4	Uncharacterized protein (Acc. N. W5PTS4) similar to Ribonuclease 4, RNase family member 4	YES
320	RPS27A	Uncharacterized protein (Fragment) (Acc. N. W5NX91) similar to Ubiquitin-40S ribosomal protein S27a	-
321	RPSA	40S ribosomal protein SA	-
322	SAA1	Serum amyloid A protein	YES
323	SDCBP	Syntenin	-
324	SDF4	45 kDa calcium-binding protein	YES
325	SDS	Uncharacterized protein (Fragment) (Acc. N. W5PBW1) similar to L-serine dehydratase/L-threonine deaminase	-
326	SELENBP1	Uncharacterized protein (Acc. N. W5QIK8) similar to Selenium-binding protein 1	YES
327	SEMA7A	Uncharacterized protein (Fragment) (Acc. N. W5P0W4) similar to Semaphorin-7A	-
328	SERPINA1	Alpha-1-antitrypsin transcript variant 1 *	YES
329	SERPINA3-1	Serpin A3-1 *	YES
330	SERPINB1	Serpin peptidase inhibitor clade B ovalbumin member 1	-
331	SERPINC1	Antithrombin-III	YES
332	SERPIND1	SERPIND1 protein	YES
333	SERPINF1	SERPINF1	-
334	SERPING1	Uncharacterized protein (Acc. N. W5PJZ2) similar to Serpin peptidase inhibitor clade G member 1	YES
335	SHBG	Uncharacterized protein (Acc. N. W5Q0R1) similar to Sex hormone-binding globulin	-
336	SIAE	Uncharacterized protein (Acc. N. W5PLH2) similar to Sialic acid acetylesterase	-
337	SIL1	Uncharacterized protein (Acc. N. W5Q5W1) similar to Nucleotide exchange factor SIL1	YES
338	SLC3A2	Uncharacterized protein (Acc. N. W5Q8K4) similar to Solute carrier family 3 member 2	YES
339	SMPD1	Sphingomyelin phosphodiesterase	-
340	SOD3	Superoxide dismutase [Cu-Zn]	-
341	SPINT1	Uncharacterized protein (Fragment) (Acc. N. W5QGE5) similar to Serine peptidase inhibitor, Kunitz type 1	-
342	ST14	Uncharacterized protein (Fragment) (Acc. N. W5PZP2) similar to Suppressor of tumorigenicity 14 protein homolog	-
343	ST6GALNAC2	Uncharacterized protein (Acc. N. W5PA89) similar to Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2	-
344	ST8SIA1	Alpha-N-acetylneuraminide alpha-2,8-sialyltransferase	-

345	STC1	Uncharacterized protein (Acc. N. W5PJZ9) similar to Stanniocalcin 1	YES
346	TAGLN2	Transgelin	-
347	TF	Uncharacterized protein (Acc. N. W5PF65) similar to Transferrin	YES
348	TGFBI	Uncharacterized protein (Fragment) (Acc. N. W5Q0F3) similar to Transforming growth factor-beta-induced protein ig-h3	-
349	TGFBR3	Uncharacterized protein (Acc. N. E1B9H5) similar to Transforming growth factor beta receptor type 3	-
350	THBS1	Thrombospondin-1	-
351	TINAGL1	Uncharacterized protein (Fragment) (Acc. N. W5NTG6) similar to Tubulointerstitial nephritis antigen-like 1	-
352	TKT	Transketolase	YES
353	TN-X	Tenascin-X	-
354	TPP1	Uncharacterized protein (Acc. N. W5Q689) similar to Tripeptidyl peptidase I	YES
355	TTR	Transthyretin	YES
356	VAT1	Uncharacterized protein (Acc. N. F1MUP9) similar to Synaptic vesicle membrane protein VAT-1 homolog	YES
357	VH	VH region (Fragment)	-
358	VWA1	Uncharacterized protein (Fragment) (Acc. N. W5NZR6) similar to von Willebrand factor A domain-containing protein 1	-
359	XDH	Uncharacterized protein (Acc. N. W5PMT0) similar to Xanthine dehydrogenase/oxidase	YES
360	XXYLT1	Uncharacterized protein (Acc. N. W5QGQ7) similar to C1H3ORF21 protein	-
361	XYLT2	Uncharacterized protein (Acc. N. W5P3L6) similar to Xylosyltransferase 2	-
362	YKT6	Synaptobrevin homolog YKT6	YES

583

584

585

REFERENCES

[1] D'Amato A, Bachi A, Fasoli E, Boschetti E, Peltre G, Senechal H, Righetti PG. In-depth exploration of cow's whey proteome via combinatorial peptide ligand libraries. J Proteome Res 2009;8:3925−36.

[2] Hettinga K, van Valenberg H, de Vries S, Boeren S, van Hooijdonk T, van Arendonk J, Vervoort J. The host defense proteome of human and bovine milk. PLoS One 2011;6: e19433.

[3] Le A, Barton LD, Sanders JT, Zhang Q. Exploration of bovine milk proteome in colostral and mature whey using an ion-exchange approach. J Proteome Res 2011;10:692−704.

[4] Lemay DG, Lynn DJ, Martin WF, Neville MC, Casey TM, Rincon G, Kriventseva EV, Barris WC, Hinrichs AS, Molenaar AJ, Pollard KS, Maqbool NJ, Singh K, Murney R, Zdobnov EM, Tellam RL, Medrano JF, German JB, Rijnkels M. The bovine lactation genome: insights into the evolution of mammalian milk. Genome Biol 2009;10:R43.

[5] D'Alessandro A, Zolla L, Scaloni A. The bovine milk proteome: cherishing, nourishing and fostering molecular complexity. An interactomics and functional overview. Mol Biosyst 2011;7:579−97.

[6] Roncada P, Piras C, Soggiu A, Turk R, Urbani A, Bonizzi L. Farm animal milk proteomics. J Proteomics 2012;75:4259-74.

[7] Agrawal GK, Sarkar A, Righetti PG, Pedreschi R, Carpentier S, Wang T, Barkla BJ, Kohli A, Ndimba BK, Bykova NV, Rampitsch C, Zolla L, Rafudeen MS, Cramer R, Bindschedler LV, Tsakirpaloglou N, Ndimba RJ, Farrant JM, Renaut J, Job D, Kikuchi S, Rakwal R. A decade of plant proteomics and mass spectrometry: translation of technical

advancements to food security and safety issues. Mass Spectr Reviews 2013;32:335-65.

[8] Vandenplas Y, Koletzko S, Isolauri E, Hill D, Oranje AP, Brueton M, Staiano A, Dupont C. Guidelines for the diagnosis and management of cow's milk protein allergy in infants. Arch Dis Child 2007;92:902-8.

[9] Bertino E, Gastaldi D, Monti G, Baro C, Fortunato D, Perono Garoffo L, Coscia A, Fabris C, Mussap M, Conti C. Detailed proteomic analysis on DM: insight into its hypoallergenicity. Front Biosci 2010;1:526-36.

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

[11] Salimei E, Fantuz F. Equid milk for human consumption. Int Dairy J 2012;24:130-42.

[12] Monti G, Bertino E, Muratore MC, Coscia A, Cresi F, Silvestro L, Fabris C, Fortunato D, Giuffrida MG, Conti A. Efficacy of donkey's milk in treating highly problematic cow's milk allergic children: An *in vivo* and *in vitro* study. Pediatr Allergy Immunol 2007;18: 258-64.

[13] Cunsolo V, Muccilli V, Fasoli E, Saletti R, Righetti PG, Foti S. Poppea's bath liquor: the secret proteome of she-donkey's milk. J Proteomics 2011;74:2083-99.

[14] Cunsolo V, Cairone E, Fontanini D, Criscione A, Muccilli V, Saletti R, Foti S. Sequence determination of αs1-casein isoforms from donkey by mass spectrometric methods. J Mass Spectrom 2009;44:1742–53.

[15] Saletti R, Muccilli V, Cunsolo V, Fontanini D, Capocchi A, Foti S. MS-based characterization of *αs2*-casein isoforms in donkey's milk. J Mass Spectrom 2012;47: 1150-9.

[16] Cunsolo V, Muccilli V, Saletti R, Foti S. Review: Applications of mass spectrometry techniques in the investigation of milk proteome. Eur J Mass Spectrom 2011;17:305-20.

[17] Caira S, Pizzano R, Picariello G, Pinto G, Cuollo M, Chianese L, Addeo F. Allergenicity of milk proteins. In: Hurley WL, editor. Milk protein. InTech http://www.intechopen.com/books/milk-protein; 2012, p.173-214.

[18] Ah-Leung S, Bernard H, Bidat E, Paty E, Rancé F, Scheinmann P, Wal JM. Allergy to goat and sheep milk without allergy to cow's milk. Allergy 2006;61:1358-65.

[19] Jirillo F, Magrone T. Anti-inflammatory and anti-allergic properties of donkey's and goat's milk. Endocr Metab Immune Disord Drug Targets 2014;14:27-37.

[20] Järvinen KM, Chatchatee P. Mammalian milk allergy: clinical suspicion, cross-reactivities and diagnosis. Curr Opin Allergy Clin Immunol 2009;9:251-8.

[21] Hinz K, O'Connor PM, Huppertz T, Ross RP, Kelly AL. Comparison of the principal proteins in bovine, caprine, buffalo, equine and camel milk. J Dairy Res 2012;79:185-91.

[22] Yang Y, Bu D, Zhao X, Sun P, Wang J, Zhou L. Proteomic analysis of cow, yak, buffalo, goat and camel milk whey proteins: quantitative differential expression patterns. J Proteome Res 2013;12:1660-7.

[23] Cunsolo V, Muccilli V, Saletti R, Foti S. MALDI-TOF mass spectrometry for the monitoring of she-donkey's milk contamination or adulteration. J Mass Spectrom 2013;48:148-153.

[24] Di Girolamo F, Masotti A, Salvatori G, Scapaticci M, Muraca M, Putignani L. A sensitive and effective proteomic approach to identify she-donkey's and goat's milk adulterations by MALDI-TOF MS fingerprinting. J Mol Sci 2014;15:13697-13719.

[25] Tay EP, Gam LH. Proteomics of human and the domestic bovine and caprine milk. Asian Pacific J Mol Biol Biotechnol 2011;19:45-53.

[26] Selvaggi M, Laudadio V, Dario C, Tufarelli V. Major proteins in goat milk: and updated overview on genetic variability. Mol Biol Rep 2014;41:1035-48.

[27] Almeida da Costa WK, Souza EL, Beltrão-Filho EM, Vasconcelos GK, Santi-Gadelha T, de Almeida Gadelha CA, Franco OL, de Cássia Ramos do Egypto Queiroga R, Magnani M. Comparative protein composition analysis of goat milk produced by the Alpine and Saanen breeds in northeastern Brazil and related antibacterial activities. PLoS One 2014; 9:e93361.

[28] Fasoli E, Farinazzo A, Sun CJ, Kravchuk AV, Guerrier L, Fortis F, Boschetti E, Righetti PG. Interaction among proteins and peptide libraries in proteome analysis: pH involvement for a larger capture of species. J Proteomics 2010;73:733-42.

[29] Candiano G, Dimuccio V, Bruschi M, Santucci L, Gusmano R, Boschetti E, Righetti, PG, Ghiggeri GM. Combinatorial peptide ligand libraries for urine proteome analysis: investigation of different elution systems. Electrophoresis 2009;30:2405-11.

[30] Han Y, Ma B, Zhang K. SPIDER: Software for Protein Identification from Sequence Tags Containing De Novo Sequencing Error. J Bioinform Comput Biol 2005;3:697-716.

[31] Guiochon G. The limits of the separation power of unidimensional column liquid chromatography. J Chromatogr A, 2006;1126:6-49.

[32] Forssén P, Samuelsson J, Fornstedt T. Relative importance of column and adsorption parameters on the productivity in preparative liquid chromatography II: Investigation of separation systems with competitive Langmuir adsorption isotherms. J Chromatogr A, 2014;1347:72–79.

[33] Roux-Dalvai F, Gonzalez de Peredo A, Simó C, Guerrier L, Bouyssié D, Zanella A, Citterio A, Burlet-Schiltz O, Boschetti E, Righetti PG, Monsarrat B. Extensive analysis of the cytoplasmic proteome of human erythrocytes using the Peptide ligand library technology and advanced mass spectrometry. Mol Cell Proteomics 2008;7:2254-69.

[34] Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ, Liebler DC. Depletion of abundant plasma proteins and limitations of plasma proteomics. J Proteome Res 2010;9:4982-91.

[35] Zhi W, Purohit S, Carey C, Wang M, She JX. Proteomic technologies for the discovery of type 1 diabetes biomarkers. Diabetes Sci Technol 2010;4:993-1002.

[36] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002;1:845-67.

[37] Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie G, Ma B. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. Mol Cell Proteomics 2012;11:M111.010587.

[38] Di Girolamo F, Boschetti E, Chung MC, Guadagni F, Righetti PG. "Proteomineering" or not? The debate on biomarker discovery in sera continues. J Proteomics 2011;74:589-94.

[39] Santucci L, Candiano G, Petretto A, Lavarello C, Bruschi M, Ghiggeri GM, Citterio A, Righetti PG. Combinatorial ligand libraries as a two-dimensional method for proteome analysis. J Chromatogr A 2013;1297:106-12.

[40] Hogarth CJ, Fitzpatrick JL, Nolan AM, Young FJ, Pitt A, Eckersall PD. Differential protein composition of bovine whey: a comparison of whey from healthy animals and from those with clinical mastitis. Proteomics 2004;4:2094-100.

[41] Bu HF, Zuo XL, Wang X, Ensslin MA, Koti V, Hsueh W, Raymond AS, Shur BD, Tan XD. Milk fat globule-EGF factor 8/lactadherin plays a crucial role in maintenance and repair of murine intestinal epithelium. J Clin Invest 2007;117:3673–83.