Milk Fat Globule Membrane Proteomics: A 'Snapshot' of Mammary Epithelial Cell Biology

Christelle Cebo*

National Institute of Agronomy Research (INRA), UMR1313 Animal Genetics and Integrative Biology, FR-78350 Jouy-en-Josas, France

Summary

Lipids are released in milk as fat globules, which are droplets of apolar lipids surrounded by a complex membrane deriving from the mammary epithelial cell (MEC) and called the milk fat globule membrane (MFGM). The structure of the MFGM is highly complex and closely related to the mechanisms of milk fat globule secretion in the mammary epithelial cell. Indeed, MFGM is composed of two biological membranes, a phospholipid monolayer, deriving from the endoplasmic reticulum, and a phospholipid bilayer, which originates from the apical plasma membrane of the MEC, with variable amounts of cytoplasm trapped between. Biochemical techniques (*i.e.* sodium dodecyl sulphate-polyacrylamide gel electrophoresis followed by different staining procedures) have been employed historically to characterize major MFGM proteins, namely MUC-1, fatty acid synthase, xanthine oxidase, butyrophilin, lactadherin, and adipophilin. However, recent advances in the field of proteomics (mostly development of one-dimensional gel electrophoresis approach coupled with tandem mass spectrometry) have led to the identification of hundreds of proteins associated with the MFGM. Surprisingly, newly identified MFGM proteins were not only involved in lipid metabolic or exocytosis-related biological processes, but also in cell signalling, translation, or host defense-related mechanisms. Therefore, the milk fat globule should no longer be viewed as an inert structure only devoted to the delivery of lipids to the newborn, but rather as a dynamic and informative compartment which can contribute to the improvement of our comprehension of the mammary gland biology.

Key words: milk fat globule membrane, mass spectrometry-based proteomics, mammary gland biology, integrative biology

The milk fat globule: A highly complex structure delivering lipids to the neonate

Lipids are dispersed in the aqueous phase of milk as droplets of triacylglycerols surrounded by a complex membrane, called the milk fat globule membrane (MFGM). Milk lipid synthesis takes place in the mammary epithelial cell (MEC) by a step-to-step mechanism unique to secretory cells (1). Indeed, lipid droplet formation is thought to occur between the two leaflets of the endoplasmic reticulum (ER) leading to the budding of cytoplasmic lipid droplets (CLD), which are thus mainly composed of triacylglycerols surrounded by an ER-derived phospholipid monolayer. Then, CLD migrate, probably moving along the microtubule machinery, to the apical pole of the MEC where they are wrapped by the apical plasma

Abbreviations: 1D-LC-MS/MS: one-dimensional polyacrylamide gel electrophoresis coupled with liquid chromatography-tandem mass spectrometry; 2-DE: two-dimensional gel electrophoresis; 2D-DIGE: two-dimensional differential gel electrophoresis; ADRP: adipose differentiation-related protein, adipophilin, perilipin-2 (PLIN-2); BTN: butyrophilin subfamily 1 member A1; CLD: cytoplasmic lipid droplet; ER: endoplasmic reticulum; FAS: fatty acid synthase; LDH: lactadherin; MEC: mammary epithelial cell; MFGM: milk fat globule membrane; SDS-PAGE: sodium dodecyl sulphate-polyacrylamide gel electrophoresis; XO: xanthine oxidase
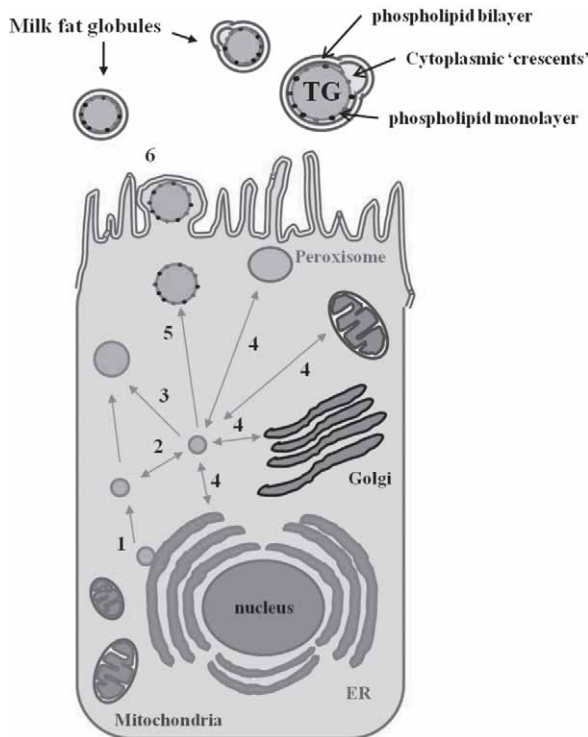
Fig. 1. Schematic view of the milk fat globule secretion from the mammary epithelial cell: 1. lipid droplet formation initiates in the endoplasmic reticulum (ER) through the accumulation of neutral lipids between the leaflets of the phospholipid bilayer; 2. lipid droplets can interact with each other through homotypic interactions; 3. lipid droplets can fuse with each other to form larger lipid droplets; 4. material and/or information can be exchanged during heterotypic interactions of lipid droplets with other cellular organelles including the ER, the Golgi complex, peroxisomes, or mitochondria; 5. lipid droplets move along the microtubule machinery to reach the apical pole of the mammary epithelial cell; 6. lipid droplets are finally released as fat globules in milk by being progressively enveloped by the apical plasma membrane. Consequently, most of MFGM-associated proteins belong to the ER and plasma membrane compartments. In addition, cytoplasmic inclusions, known as 'crescents', are often trapped between the outer membrane and the lipid core during the final secretion process. These crescents contain soluble, cytoplasmic proteins which can also be identified in MFGM isolates. ER=endoplasmic reticulum, TG=triacylglycerols

membrane to be released as fat globules in milk (Fig. 1). Hence, MFGM, whose composition is closely related to the mechanisms of milk fat globule secretion from the mammary cell, is composed of an inner phospholipid monolayer derived from the ER, surrounded by an outer phospholipid bilayer, originating from the plasma membrane during budding, with variable amounts of entrained cytoplasm in between (Fig. 1).

Imaging techniques, namely labelling procedures followed by confocal analysis, have recently been used, thus improving our current view of milk fat globules, and a model for the structure of the human MFGM has been proposed (*2,3*). In this model, MFGM polar lipids are organized in two phases, a glycerolipid (phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and phosphatidylinositol)-rich, liquid-disordered phase where MFGM proteins are dispersed, and a liquid-ordered phase (featuring 'lipid rafts'), composed of sphingomyelin and cholesterol where glycoproteins are excluded. MFGM organization remains, however, largely speculative but appears as a highly complex structure delivering lipids as well as proteins to the offspring. Little is known about the rationale of such a complex organization and MFGM structure-to-function relationships need further investigation.

Major MFGM proteins: The molecular diversity of lactadherin through species

MFGM material can be resolved by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie staining into few protein bands corresponding to fatty acid synthase (FAS), xanthine oxidase (XO), butyrophilin (BTN), lactadherin (LDH; formerly known as PAS 6/7), and adipophilin (adipose differentiation-related protein (ADRP) or perilin-2). Mucins (which are large proteins containing more than 50 % O-glycans by mass) such as MUC-1 and MUC-15 (formerly known as PAS 3) are only visualized by SDS-PAGE followed by periodic-acid Schiff staining revealing only highly glycosylated proteins (Fig. 2).

Most of the studies published to date on MFGM proteins have been dedicated to the bovine species (*4*). However, the part of non-cow's milk production in the total milk production is highly variable across the world and interest in non-cattle milk is still growing (*5*). This highlights the need to better characterize milk from other species than cow. Accordingly, we have recently described major MFGM proteins from goat (*6*), horse (*7*) or camel milk (unpublished data), where data were scarce and we have highlighted prominent differences across species, especially for lactadherin, a major protein of the milk fat globule membrane (*8*). We have shown that lactadherin from goat's milk appears as a single polypeptide chain in 10 % SDS-PAGE, whereas two and four polypeptide chains are respectively identified for lactadherin from camel and horse milk. Although we identified post-transcriptional variants (*i.e.* splicing variants) for equine lactadherin, the molecular diversity of lactadherin across species is mostly explained by a differential glycosylation of the protein (*7–9*).

Major MFGM proteins have been extensively reviewed (*4*) and a review has recently been dedicated to proteomic analysis of MFGM components (*10*). We will therefore focus our discussion on recent advances in the field of MFGM proteomics, mostly the development of one-dimensional gel electrophoresis coupled with tandem mass spectrometry (1D-LC-MS/MS) together with MS-based quantitative approaches to analyze the MFGM proteome. A discussion will ensue regarding how proteomics have improved our knowledge on the biology of the mammary gland.

Methods of Analyzing the Protein Composition of the Milk Fat Globule Membrane

MFGM proteomics: A brief history

In eukaryotes, the proteome (*i.e.* the entire set of proteins expressed by a genome, cell, tissue or organism) is much larger than the genome (around 20 000 genes in the bovine species), in the sense that there are
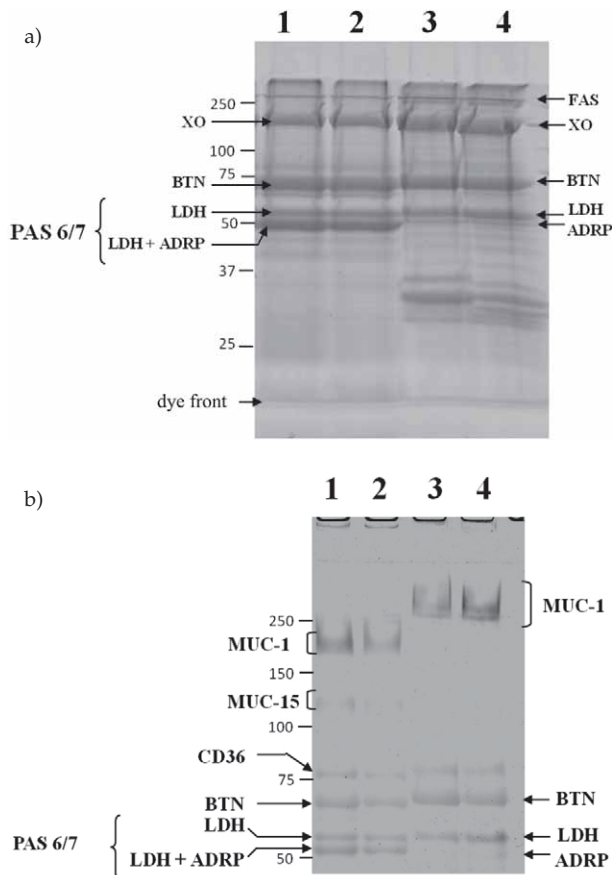
Fig. 2. Representative pattern of milk fat globule membrane (MFGM) proteins in SDS-PAGE: a) MFGM proteins of 25 μg from bovine (lanes 1–2) and caprine (lanes 3–4) milk were separated on 10 % SDS-PAGE and stained with Bio-Safe (Bio-Rad, Marnes-la-Coquette, France); b) MFGM proteins of 50 μg from bovine (lanes 1–2) or caprine (lanes 3–4) milk were loaded on 6 % SDS-PAGE and stained with periodic acid/Schiff reagent (PAS). Positions of major proteins of the milk fat globule membrane are indicated. FAS=fatty acid synthase, XO=xanthine oxidase, BTN=butyrophilin, LDH=lactadherin, ADRP=adipose-related differentiation protein (adipophilin). Positions of protein standards (kDa) are indicated to the left of the panel. Reprinted from Cebo *et al*. (*6*) with permission from Elsevier

more proteins than genes. Alternative splicing of genes and post-translational modifications (*e.g.* glycosylation or phosphorylation) of gene products greatly increase the number of proteins expressed by a cell at a given time. Biochemical techniques, namely SDS-PAGE followed by different staining procedures, have been initially employed to characterize MFGM proteins, mostly in the bovine species (*4*). The MFGM proteome must be viewed as a minute fraction of the mammary epithelial cell proteome, but it remains highly complex and cannot be adequately analyzed only by means of one-dimensional electrophoresis.

The development of two-dimensional approaches greatly improved the resolution of proteins from complex mixtures. This two-dimensional gel system combines isoelectric focusing (IEF; the proteins are separated according to their isoelectric point) in the first dimension to a discontinuous SDS-PAGE (the proteins are separated according to their molecular mass) in the second

dimension. This technique has been successfully applied to the separation of MFGM proteins in most species, including human (*11,12*), bovine (*13,14*), equine (*15*), and buffalo (*16*) species.

However, conventional 2-DE (*i.e.* isoelectric focusing followed by SDS-PAGE) is not appropriate for the separation of large and hydrophobic proteins as found in the MFGM. Because they interfere with the isoelectric focusing step of conventional 2-DE, ionic detergents such as SDS, which efficiently solubilize hydrophobic proteins, cannot be used to extract proteins from MFGM. Uncharged chaotropes (urea and thiourea) and non-ionic detergents are therefore used to solubilize samples for 2-DE. Improved extraction protocols have been proposed for resolving MFGM proteins by conventional 2-DE (*14, 17*). However, IEF-compatible lysis buffers do not extract hydrophobic proteins efficiently, and true-membrane proteins are consequently depleted during conventional 2-DE (*18*).

Due to the limitations of conventional 2-DE for the separation of highly hydrophobic proteins, alternative strategies have been proposed for separating membrane proteins. These strategies are based on the differential solubility of proteins in cationic detergents such as 16--benzyldimethyl-*n*-hexadecylammonium chloride (16--BAC) and in anionic detergents such as SDS (*19,20*). However, the resolving power of this non-conventional 2-DE approach (namely, 2D-16-BAC/SDS-PAGE) is considerably lower than the resolving power of conventional 2-DE.

A technological breakthrough occurred over the past decade with the development of one-dimensional gel electrophoresis approach coupled with tandem mass spectrometry (1D-LC-MS/MS). In this approach, proteins are first separated by SDS-PAGE, and then they are visualized by Coomassie Blue staining. Each lane is cut into a various number of slices (depending on the complexity of the sample), and each slice is digested with trypsin. Resulting peptides are subsequently analyzed by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). This approach, namely, pre--fractionation of samples by one-dimensional SDS-PAGE prior to LC-MS/MS analysis, is particularly well suited to the analysis of MFGM proteome. First, ionic detergents such as SDS possess a high capacity to solubilize hydrophobic proteins as found in the MFGM. Second, major discrepancies are observed in the MFGM with regard to protein abundance. For example, butyrophilin can constitute up to 40 % of the total protein content in the bovine MFGM (*4*). Prefractionation of samples by 1D electrophoresis therefore improves the detection of low abundance proteins associated with the MFGM.

This methodology has been first applied to the analysis of bovine MFGM proteome (*21*). By using this approach, Reinhardt and Lippolis (*21*) identified 120 proteins associated with the bovine MFGM. The method was then applied in the identification of MFGM-associated proteins in the human species (*22*). This approach has demonstrated its feasibility even for non fully-sequenced genomes such as *Ovis aries* (sheep) (*23*), since protein identification by MS can be efficiently performed, either by comparison with homologous proteins from other species or with translated expressed sequence tags (ESTs).

MFGM proteomics: The need to standardize sample extraction procedures

Whatever the method used for the analysis of MFGM proteome (2-DE or 1D-LC-MS/MS), a particular attention must be paid to sample preparation, from sampling to MFGM protein extraction. Multiple studies have been reported for MFGM proteome analysis, and most of these studies do not agree on the protocol for MFGM protein extraction, thus precluding inter-study comparison of the data. Indeed, several methods for the extraction of MFGM proteins have been employed to date. Roughly, four steps are required for MFGM protein extraction: (*i*) the separation of fat globules from whole milk (skimming), (*ii*) the washing of the cream (to avoid contamination by caseins or whey proteins), (*iii*) the release of membrane fragments by disruption of fat globules (*e.g.* freezing/thawing and homogenization), and (*iv*) the collection of MFGM pellet by ultracentrifugation (usually, 1 h at 100 000×*g* at low temperature followed by washing of the membrane pellet) (*21,24*). A direct extraction procedure of washed fat globules by chemical detergents (the most potent detergent being SDS) is also often employed (*6,12,15,25*). This latter protocol often requires delipidation of samples (*e.g.* protein precipitation by the chloroform/methanol extraction procedure and resolubilization of proteins) to avoid interfering with subsequent analyses (*17,25*).

However, authors do not agree on MFGM protein extraction procedures. Regarding the sampling, some authors use pooled milk samples (*12,15,21,22,24*), whereas individual aliquots are used for subsequent analyses in other studies (*6,7,23,25*). The question whether to pool the samples or not is an important issue which has been largely debated in other large-scale studies (*i.e.* transcriptomics), but relatively little in proteomics. Pooling has been proposed as a strategy to average biological variability, based on the assumption that measurements of the pooled signal correspond to the average of the individual signal measurements. However, the value of pooling is controversial, since pooling can be viewed as a loss of individual information (*26*).

Regarding the washing steps of fat globules, multiple procedures are employed. Some authors use ultrapure water (*13*), whereas others use saline buffers (*6,12, 15,21,22*). However, major discrepancies exist with regard to the composition of washing buffers, to the centrifugation speed, or to the operating temperature. These issues must be definitively addressed and standardized, since the choice of a washing buffer together with the number of washes of the cream influence the recovery as well as the purity of the MFGM fraction (*27*). Cooling of milk samples to 4 °C also induces changes in the distributions of MFGM proteins between the skimmed milk phase and the cream phase (*28*).

Multiple extraction methods of MFGM proteins have also been reported. For bovine MFGM proteome characterization, authors used a mechanical disruption of fat globules by several homogenization steps, followed by ultracentrifugation (100 000×*g* for 1 h). The membrane pellet is further washed and finally resuspended in 10 mM Tris-HCl buffer, 2 mM MgCl$_2$, pH=7.5, containing EDTA and protease inhibitors. An additional step is performed in order to extract intrinsic MFGM proteins (*21*). In a recent study, Hettinga *et al.* (*22*) used a different procedure to extract bovine MFGM proteins. After washing, the cream is mixed with an equal volume of ultrapure water, sonicated, and centrifuged to remove fat. The watery subnatant is assumed to contain MFGM proteins (*22*). The latter authors have identified most of the previously reported MFGM proteins (*21*), and they have reported the identification of almost 120 additional MFGM proteins. Although the authors suggest that the sensitivity of their identification method is higher compared with previous studies, the two extraction procedures of MFGM proteins are fully different (*21,22*). A delicate issue is now to define who of these two authors holds the true proteome of bovine MFGM. Direct extraction of MFGM proteins from fat globules by ionic detergents (mostly 2 % SDS (by mass per volume)) is also often employed (*6,12,25*). One can logically assume that this extraction protocol is more efficient with regard to sampling of proteins originating from the cytoplasm of the MEC, *i.e.* proteins trapped between the outer fat globule bilayer and the triacylglycerol core during the budding process.

An interesting study compared the yield, gross composition (including total lipids, protein, and carbohydrates), and enzymatic activities in the MFGM fraction prepared from three procedures: acidification, ultracentrifugation, and salting out of membrane fragments released by churning of the cream. Results showed that the composition of bovine MFGM is actually affected by the isolation method (*29*).

MS-based proteomics: Applications of quantification strategies to MFGM proteomics

The field of high-throughput proteomics has moved from simple protein identification to accurate quantification of protein abundances in a particular organism or tissue, at a particular time, or under particular conditions. Two main strategies have emerged for quantitative MS-based proteomics, including metabolic or chemical labelling of samples, and label-free quantitation, which is based on spectral counting or measurements of precursor ion intensity (*30,31*).

Chemical labelling can be performed either at the protein (*e.g.* isotope-coded affinity tag method or ICAT) or at the peptide level. At the peptide level, the most popular method is probably the isobaric tags for relative and absolute quantitation (iTRAQ) method. The iTRAQ method is based on the covalent labelling of the N-terminus and side chain amines of peptides from protein digestions with tags of various masses. These samples are then pooled and usually fractionated by LC-MS/MS. A database search is performed after that using the fragmentation data to identify the labelled peptides and hence the corresponding proteins. However, these methods are time-consuming and they are frequently using expensive labelling kits. A fast and inexpensive labelling method has been described recently (*31*). Peptides arising from trypsin digestion of samples (*e.g.* control and treated samples) are labelled with distinct stable isotope dimethyl labels. The labelled samples are mixed together and subsequently analyzed by LC-MS. The mass difference of the dimethyl labels is used to compare the peptide a-

bundance in the different samples (*31*). Regardless of the technique, each labelling strategy provides only relative quantification of a given protein between samples, thus precluding inter-study comparisons.

A method for the absolute quantification of proteins has nonetheless been described. This method relies on the use of a synthetic internal standard peptide that is introduced at a known concentration during sample preparation. This peptide mimics a peptide produced during proteolysis of a target protein, except that it is enriched in stable isotopes. The mass difference allows the mass spectrometer to discriminate between these two species. The ratio between the amounts of the isotope-labelled internal standard peptide and the unlabelled peptide (*i.e.* arising from the trypsin digestion of the targeted protein) permits the quantification of the desired protein (*32*). However, this quantification method cannot be applied to large-scale proteomic studies.

Besides these labelling methods, label-free proteomics has gained great popularity during the past few years. Label-free approaches for quantitative MS-based proteomics can be divided into two main strategies: spectral counting, which is based on counting the number of peptides assigned to a protein in an MS/MS experiment, and integrated measurements of chromatographic data (ion intensities and peak area data). Combining the methods of label-free quantitation seems to be necessary to obtain reliable results and to gain sensitivity. Label-free proteomics requires normalization steps of raw data for accurate quantitation and rigorous statistical assessment of data. A major concern with MS-based quantitative proteomics (either in a labelling or a label-free approach) arises when peptide sequences can be matched to more than one protein during MS/MS identification. Working only with unique peptides is difficult to consider. Limitations additionally arise from the resolution of LC signal, since some peptides may have overlapping retention times and are therefore co-eluted. However, it is generally admitted that MS-based quantitative proteomics is more accurate than gel-based quantitation methods (*33*). It has also been reported recently that expression data generated using label-free proteomics are in good accordance with expression data arising from ELISA (enzyme-linked immuno sorbent assay) experiments (*34*).

Quantification of MFGM proteins has been performed either by using labelling techniques (*24,35,36*) or by label-free proteomics (*23,25,37,38*). Applications of quantitative MS-based MFGM proteomics in the lactation biology area, ranging from the characterization of developmental changes throughout lactation (*24,25*) to the identification of the biological response of mammary gland to bacterial infection (*i.e.* mastitis (*37*)) will be discussed further in detail.

Limitations of MS-based proteomics

Recent advances in MS-based proteomics have largely contributed to making significant strides in the field of biology. However, despite a constant upgrading of mass spectrometers, the reliability of MS-based proteomics is low. A major concern in proteomics is the high complexity of biological samples, especially in shotgun proteomic strategies. Actually, a single MS/MS spectrum can match many proteins, either those belonging to different protein families but presenting a high sequence homology, or those arising from alternatively spliced genes (*39*). From this point of view, benefits of 2-DE separation of proteins, which results in reducing the complexity of samples subjected to MS identification, are indisputable.

To precisely address the question of reliability of MS-based proteomics, the Human Proteome Organization (HUPO) has recently performed a test sample study (*40*). A test sample made up of twenty human proteins of high purity and at equimolar ratios was distributed to twenty-seven labs (both academic labs and vendors) worldwide. Each protein of the test sample that contained one or more unique tryptic peptides of 1250 Da was chosen to test for ion selection and peptide sampling in mass spectrometers. Even when working on this low-complexity sample, only seven labs identified the twenty proteins correctly, and only one lab identified all tryptic peptides of 1250 Da. Common problems encountered by participating labs were identified as environmental contaminations, database matching and curation of protein identifications (*40*).

Indeed, proteomics generates huge amounts of MS/MS spectra, and the manual sequencing of corresponding peptides is not feasible. Consequently, search algorithms are used to match the MS/MS spectra of peptides with the protein sequence in databases. However, the recovery of results produced by different search engines is fairly low, especially when dealing with complex protein mixtures of variable abundance. For example, when using the same MS/MS dataset, the overlap between Sequest and Mascot (two popular search engines) represents only 51 % of the Mascot matches and 39 % of the Sequest matches (*41*). A 'reversed database' (*i.e.* a database in which the sequences have been scrambled to produce only false positive identifications) can be used to estimate the false discovery rate (FDR). However, reversed databases do not evaluate the false negative error rate. Another frequently encountered problem is the incomplete status of search databases. Thus, using different databases for MS/MS spectra matching may lead to irreproducibility of protein identification results.

MFGM Proteomics: A Powerful Tool to Improve Our Understanding of Mammary Epithelial Cell Biology

Challenging views of milk lipid secretion

Although lipid droplets can be synthesized by other cell types (*e.g.* adipocytes, hepatocytes, *etc.*), lipids are secreted into milk by a mechanism unique to the MEC. Triacylglycerol synthesis is thought to occur between the two leaflets of the ER, thus resulting in the release of cytoplasmic lipid droplets (CLD) coated with a phospholipid monolayer. However, based on the observations arising from freeze-fracture experiments, some authors suggested that lipid droplet biosynthesis takes place in specialized cup-shaped regions of the ER, and that adipophilin appears to function in transferring lipids from the ER to the lipid droplet surface (*42*). The same authors proposed that the final budding of CLD from the MEC and their release as fat globules in milk is con-

trolled by butyrophilin homotypic interactions rather than xanthine oxidase-butyrophilin-adipophilin interactions, as previously hypothesized (*1,43,44*). The recent finding that adipophilin C-terminal domain is able to bind to plasma membrane by itself provides new insights in molecular steps involved in milk lipid secretion (*45,46*).

Despite all these challenging views, one can agree on the fact that lipid droplets are released into the cytoplasm of the MEC as a core of triacylglycerols surrounded by a monolayer originating from the ER, and that these cytoplasmic lipid droplets are released as fat globules in milk by being progressively enveloped by the apical plasma membrane of the MEC, with variable amounts of cytoplasm entrained between the two biological membranes. The milk fat globule membrane (MFGM), the membrane surrounding lipids in milk, can therefore be viewed as a 'sampling' of the mammary epithelial cell. In addition, recent studies have suggested that lipid droplets are not inert structures involved in the delivery of lipids to the newborn. Instead, it has been demonstrated that lipid droplets can interact with each other (homotypic interactions; Fig. 1) or with other organelles of the MEC, including the Golgi apparatus, peroxisomes, or mitochondria (heterotypic interactions (*47,48*); Fig. 1). It has also been demonstrated that lipid droplets can serve as transient storage depot for proteins, allowing their proper delivery to their ultimate destination within the cell through the microtubule machinery in a controlled manner over time (*49*). Hence, the MFGM reflects dynamic changes occurring within the MEC and can therefore provide a 'snapshot' of mammary gland biology under particular conditions.

The recent technological breakthroughs of MS-based proteomics together with the unique secretion process of lipids occurring within the MEC emphasize the MFGM proteomics as a powerful tool to unravel mammary gland biology. In this context, significant improvements in the comprehension of lactation biology have been achieved over the past five years. MFGM proteomic studies can be classified into three main research areas: (*i*) the comprehension of lipid secretory pathways within the MEC, and the alterations of these lipid secretory mechanisms under (*ii*) physiological conditions (*e.g.* lactation stage), or (*iii*) pathological conditions (*e.g.* mastitis).

MFGM proteomics as a powerful tool to identify partners of lipid secretion in mammals

Identification of proteins in the MFGM may help to identify genuine partners of lipid droplet secretion in the MEC and their release as fat globules in milk. First evidence for a link between ER and lipid secretory mechanisms was reported in a study based on 2-DE proteomics and comparing mammary and liver CLD proteomes on one hand, and MFGM composition on the other hand (*50*). By this method, authors identified 15 proteins specifically found both in MFGM and in mammary CLD fractions, thus suggesting a potential role of these proteins in initial steps of lipid secretion in the MEC (*50*). A 1D-LC-MS/MS approach in the bovine species led to the identification of 120 MFGM-associated proteins. Among these proteins, more than 50 % were associated with a membrane/protein trafficking, cell signalling, or fat/

transport metabolism function (*21*). One can reasonably hypothesize that these proteins are involved in lipid secretory mechanisms in the MEC, from the initial budding in the ER to the secretion as fat globules in milk. Several studies identifying MFGM-associated proteins have already been reported (*22,23*), and we are currently working on MFGM proteins from other species (unpublished data). Taken together, these studies will open new roads in the understanding of lipid secretion pathways in mammary epithelial cells.

MFGM proteomics to unravel mammary epithelium cell biology

Variations in the MFGM composition may be linked to dynamic changes occurring within the mammary epithelial cell. MS-based quantitative proteomics is used to investigate variations in MFGM protein abundances, for example throughout lactation. This has already been completed in the bovine (*24*) and human (*25*) species. In the cow, Reinhardt and Lippolis (*24*) used a shotgun proteomic approach coupled with iTRAQ labelling of peptides to quantify protein changes in the MFGM from colostrum and mature milk (*24*). Interestingly, they identified 26 up-regulated proteins and 19 down-regulated proteins in the mature MFGM compared with colostrum MFGM. Proteins associated with lipid synthesis (*e.g.* acyl-CoA synthetase, cell death-inducing DFFA-like effector A (CIDE-A), fatty acid-binding protein (FABP), *etc.*) and secretion (remarkably, the tripartite complex of proteins xanthine oxidase, butyrophilin, and adipophilin) were most expressed in mature MFGM, whereas apolipoproteins or immune-related proteins (*e.g.* lactoferrin, Ig γ-1 chain C region, *etc.*) were more expressed in colostrum MFGM. In the human MFGM, Liao *et al.* (*25*) used a label-free MS-based proteomic approach to characterize protein changes in a 12-month lactation period. They also reported quantitative variations regarding the MFGM protein composition that may be related to dynamic changes in the mammary gland throughout lactation (*25*). Despite the fact that these studies have major discrepancies with regard to the MFGM sample extraction protocol or MS-based quantitative analyses, they pointed out differential expression of MFGM proteins throughout lactation that may be related to the widely recognized nutritional value of MFGM (*51*).

A recent study has reported that MFGM proteomics can also be used to decipher the biological response of mammary cells to bacterial infection (*37*). MFGM samples were collected from lactating ewes spontaneously infected by *Mycoplasma agalactiae*, an important pathogen of small ruminants, affecting several tissues including the mammary gland, and causing significant losses in the sheep and goat milk industries. MFGM samples from *M. agalactiae*-positive or -negative ewes were extracted and subjected to either 2D-DIGE or label-free MS-based proteomics. Authors identified 68 statistically significant differentially expressed MFGM proteins upon *M. agalactiae* infection. Proteins involved in inflammation and immune defense (*e.g.* cathelicidin-1, calgranulin B, *etc.*) or oxidative stress (such as the mitochondrial superoxide dismutase) were over-expressed during a natural *M. agalactiae* infection, whereas the expression of genuine MFGM pro-

teins devoted to the lipid synthesis and secretion (*e.g.* xanthine oxidase, perilipin-2, fatty acid binding protein, *etc.*) decreased (*37*). Taken together, these data provide new insights in the biological response of the mammary gland to bacterial infection and point out MS-based proteomics as a valuable tool to characterize early markers of mastitis in lactating animals.

Future Challenges: Integrating 'Omics' Data

Large-scale datasets arising from 'omics' experiments (*i.e.* transcriptomics, proteomics, or even metabolomics) are accumulating. However, our ability to integrate them meaningfully is poor. To make sense to the community, protein or gene lists must be converted into biology.

To address the need for consistent descriptions of gene products in different databases, the Gene Ontology (GO) project has developed three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner (*52*). Molecular function describes activities, such as catalytic or binding activities, that occur at the molecular level. Broad functional terms are catalytic activity, transporter activity, or binding. Narrower functional terms are GTPase activity (GO:0003924) or carbohydrate binding (GO:0030246). A biological process is a series of events accomplished by one or more ordered assemblies of molecular functions, *e.g.* lipid biosynthetic process (GO:0008610). The cellular component ontology describes locations at the levels of subcellular structures and macromolecular complexes. Examples of cellular components are apical plasma membrane (GO: 0016324), or SNARE complex (GO:0031201).

Besides GO annotation of protein or gene lists, enrichment analysis is a meaningful strategy that increases the likelihood for researchers to identify biological processes which are under investigation. The principle of enrichment analysis is that if a biological process is abnormal in a given study, the co-functioning genes (or proteins) should have a higher potential (enriched) to be selected. To decide the degree of enrichment, a certain background must be set up in order to perform the comparison. For example, if 10 % of the user's genes are kinases *vs.* 1 % of the genes in the human genome (this is the gene population background), kinases are enriched in the user's study, and therefore may play important roles in that study.

A range of free web-based browsers helps scientists to browse the ontologies (full list available at *http://www. geneontology.org/GO.tools.shtml*). DAVID (The Database for Annotation, Visualization and Integrated Discovery (*53*)) is one of the most popular software solutions dedicated to the functional annotation and enrichment analysis of large datasets. By the year 2010, DAVID tools had been cited in over 2000 publications. A paper published in *Nature Protocols* describes step-by-step procedure to use DAVID (*54*).

Biological interpretation of large-scale datasets also requires visualizing them in protein-protein interaction networks. Within a cell, protein interactions range from pairwise (*e.g.* ligand-receptor binding during signal transduction) to the formation of macromolecular complexes involving many proteins. These protein networks can be visualized using dedicated software, either free or under commercial license. Cytoscape (*55*) is an open source software platform for visualizing complex networks and integrating these with any type of attribute data. A lot of freely available plug-ins can be used, for example for inferring new networks or for functional enrichment of networks (*56*). Ingenuity Pathway Analysis (IPA) is data analysis software from Ingenuity® Systems (*57*). It analyzes data from a variety of experimental platforms and provides accurate biological insight into the interactions between genes or proteins. The advantage of IPA software compared to others is probably that modelled relationships are manually reviewed for accuracy.

Future challenges to a better understanding of biological systems (*e.g.* lipid secretion pathways in the mammary gland) will be to simultaneously integrate 'omics' data from different experiments on a single sample, for example transcriptomics and proteomics on the milk fat globule. Indeed, it has recently been shown that the milk fat globule contains high-quality messenger RNA that can be used in transcriptomic studies, either in the human (*58,59*) or non-human species (*7*). Thus, the milk fat globule can now be viewed as a sampling of the mammary gland that can be used to improve our integrated view of mammary gland biology. Several bioinformatic packages devoted to the integrative analysis of multiple datasets from different technology platforms are already available on the web (*e.g.* IntegrOmics (*60*)).

Concluding Remarks

If traditional biochemical approaches have been primarily used to characterize major MFGM proteins, technological breakthroughs which occurred over the past decade (most notably, the development of 1D-LC-MS/MS) have led to the identification of hundreds of MFGM-associated proteins. In addition, the fact that the MFGM most likely reflects the mammary epithelial cell content and the emergence of quantitative proteomics may help to decipher biological mechanisms occurring in the mammary gland with regard to physiological (*e.g.* lactation stage) or pathological (*e.g.* mastitis) conditions. However, there are a wide range of procedures for the preparation of MFGM material, from milk sampling to MFGM protein extraction. Standardization of methods when working on MFGM is thus mandatory to increase the recovery of proteomic studies and to improve our integrated view of mammary gland biology.

References

1. I.H. Mather, T.W. Keenan, Origin and secretion of milk lipids, *J. Mammary Gland Biol. Neoplasia, 3* (1998) 259–273.

2. C. Lopez, M.N. Madec, R. Jimenez-Flores, Lipid rafts in the bovine milk fat globule membrane revealed by the lateral segregation of phospholipids and heterogeneous distribution of glycoproteins, *Food Chem. 120* (2010) 22–33.

3. C. Lopez, O. Ménard, Human milk fat globules: Polar lipid composition and *in situ* structural investigations revealing the heterogeneous distribution of proteins and the lateral segregation of sphingomyelin in the biological membrane, *Colloids Surf. B, 83* (2011) 29–41.

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

5. B. Faye, G. Konuspayeva, The sustainability challenge to the dairy sector – The growing importance of non-cattle milk production worldwide, *Int. Dairy J. 24* (2012) 50–56.

6. C. Cebo, H. Caillat, F. Bouvier, P. Martin, Major proteins of the goat milk fat globule membrane, *J. Dairy Sci. 93* (2010) 868–876.

7. C. Cebo, E. Rebours, C. Henry, S. Makhzami, P. Cosette, P. Martin, Identification of major milk fat globule membrane proteins from pony mare's milk highlights the molecular diversity of lactadherin across species, *J. Dairy Sci. 95* (2012) 1085–1098.

8. C. Cebo, P. Martin, Inter-species comparison of milk fat globule membrane proteins highlights the molecular diversity of lactadherin, *Int. Dairy J. 24* (2012) 70–77.

9. J. Hvarregaard, M.H. Andersen Lars Berglund, J.T. Rasmussen, T.E. Petersen, Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat globules, *Eur. J. Biochem. 240* (1996) 628–636.

10. M. Cavaletto, M.G. Giuffrida, A. Conti: Milk Fat Globule Membrane Components – A Proteomic Approach. In: *Bioactive Components of Milk*, Springer, New York, NY, USA (2008) pp. 129–142.

11. J. Charlwood, S. Hanrahan, R. Tyldesley, J. Langridge, M. Dwek, P. Camilleri, Use of proteomic methodology for the characterization of human milk fat globular membrane proteins, *Anal. Biochem. 301* (2002) 314–324.

12. D. Fortunato, M.G. Giuffrida, M. Cavaletto, L.P. Garoffo, G. Dellavalle, L. Napolitano *et al.*, Structural proteome of human colostral fat globule membrane proteins, *Proteomics, 3* (2003) 897–905.

13. B.Y. Fong, C.S. Norris, A.K.H. MacGibbon, Protein and lipid composition of bovine milk-fat-globule membrane, *Int. Dairy J. 17* (2007) 275–288.

14. L. Bianchi, M. Puglia, C. Landi, S. Matteoni, D. Perini, A. Armini *et al.*, Solubilization methods and reference 2-DE map of cow milk fat globules, *J. Proteomics, 72* (2009) 853–864.

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

16. C. D'Ambrosio, S. Arena, A.M. Salzano, G. Renzone, L. Ledda, A. Scaloni, A proteomic characterization of water buffalo milk fractions describing PTM of major species and the identification of minor components involved in nutrient delivery and defense against pathogens, *Proteomics, 8* (2008) 3657–3666.

17. S. Quaranta, M.G. Giuffrida, M. Cavaletto, C. Giunta, J. Godovac-Zimmermann, B. Canas *et al.*, Human proteome enhancement: High-recovery method and improved two-dimensional map of colostral fat globule membrane proteins, *Electrophoresis, 22* (2001) 1810–1818.

18. R.J. Braun, N. Kinkl, M. Beer, M. Ueffing, Two-dimensional electrophoresis of membrane proteins, *Anal. Bioanal. Chem. 389* (2007) 1033–1045.

19. J. Hartinger, K. Stenius, D. Högemann, R. Jahn, 16-BAC/SDS-PAGE: A two dimensional gel electrophoresis system suitable for the separation of integral membrane proteins, *Anal. Biochem. 240* (1996) 126–133.

20. R.P. Zahedi, C. Meisinger, A. Sickmann, Two-dimensional benzyldimethyl-*n*-hexadecylammonium chloride/SDS-PAGE for membrane proteomics, *Proteomics, 5* (2005) 3581–3588.

21. T.A. Reinhardt, J. Lippolis, Bovine milk fat globule membrane proteome, *J. Dairy Res. 73* (2006) 406–416.

22. K. Hettinga, H. van Valenberg, S. de Vries, S. Boeren, T. van Hooijdonk, J. van Arendonk, J. Vervoort, The host defense proteome of human and bovine milk, *PLoS One, 6* (2011) e19433.

23. S. Pisanu, S. Ghisaura, D. Pagnozzi, G. Biosa, A. Tanca, T. Roggio *et al.*, The sheep milk fat globule membrane proteome, *J. Proteomics, 74* (2011) 350–358.

24. T.A. Reinhardt, J.D. Lippolis, Developmental changes in the milk fat globule membrane proteome during the transition from colostrum to milk, *J. Dairy Sci. 91* (2008) 2307–2318.

25. Y. Liao, R. Alvarado, B. Phinney, B. Lönnerdal, Proteomic characterization of human milk fat globule membrane proteins during a 12 month lactation period, *J. Proteome Res. 10* (2011) 3530–3541.

26. N.A. Karp, K.S. Lilley, Investigating sample pooling strategies for DIGE experiments to address biological variability, *Proteomics, 9* (2009) 388–397.

27. T.T. Le, J. Van Camp, R. Rombaut, F. van Leeckwyck, K. Dewettinck, Effect of washing conditions on the recovery of milk fat globule membrane proteins during the isolation of milk fat globule membrane from milk, *J. Dairy Sci. 92* (2009) 3592–3603.

28. J.A. Dickow, L.B. Larsen, M. Hammershøj, L. Wiking, Cooling causes changes in the distribution of lipoprotein lipase and milk fat globule membrane proteins between the skim milk and cream phase, *J. Dairy Sci. 94* (2011) 646–656.

29. C. Kanno, D.H. Kim, A simple procedure for the preparation of bovine milk fat globule membrane and a comparison of its composition, enzymatic activities, and electrophoretic properties with those prepared by other methods, *Agric. Biol. Chem. 54* (1990) 2845–2854.

30. B.B. Gao, L. Stuart, E.P. Feener, Label-free quantitative analysis of one-dimensional PAGE LC/MS/MS proteome: Application on angiotensin II-stimulated smooth muscle cells secretome, *Mol. Cell. Proteomics, 7* (2008) 2399–2409.

31. P.J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A.J.R. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, *Nat. Protoc. 4* (2009) 484–494.

32. D.S. Kirkpatrick, S.A. Gerber, S.P. Gygi, The absolute quantification strategy: A general procedure for the quantification of proteins and post-translational modifications, *Methods, 35* (2005) 265–273.

33. K.A. Neilson, N.A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. Assadourian *et al.*, Less label, more free: Approaches in label-free quantitative mass spectrometry, *Proteomics, 11* (2011) 535–553.

34. J.L. Boehmer, J.A. DeGrasse, M.A. McFarland, E.A. Tall, K.J. Shefcheck, J. L. Ward *et al.*, The proteomic advantage: Label-free quantification of proteins expressed in bovine milk during experimentally induced coliform mastitis, *Vet. Immunol. Immunopathol. 138* (2010) 252–266.

35. J. Lu, S. Boeren, S.C. de Vries, H.J.F. van Valenberg, J. Vervoort, K. Hettinga, Filter-aided sample preparation with dimethyl labeling to identify and quantify milk fat globule membrane proteins, *J. Proteomics, 75* (2011) 34–43.

36. B.Y. Fong, C.S. Norris, Quantification of milk fat globule membrane proteins using selected reaction monitoring mass spectrometry, *J. Agric. Food Chem. 57* (2009) 6021–6028.

37. M.F. Addis, S. Pisanu, S. Ghisaura, D. Pagnozzi, G. Marogna, A. Tanca *et al.*, Proteomics and pathway analyses of the milk fat globule in sheep naturally infected by *Mycoplasma agalactiae* provide indications of the *in vivo* response of the mammary epithelium to bacterial infection, *Infect. Immun. 79* (2011) 3833–3845.

38. M. Affolter, L. Grass, F. Vanrobaeys, B. Casado, M. Kussmann, Qualitative and quantitative profiling of the bovine milk fat globule membrane proteome, *J. Proteomics, 73* (2010) 1079–1088.

39. A.I. Nesvizhskii, R. Aebersold, Interpretation of shotgun proteomic data, *Mol. Cell. Proteomics, 4* (2005) 1419–1440.

40. A.W. Bell, E.W. Deutsch, C.E. Au, R.E. Kearney, R. Beavis, S. Sechi *et al.*, A HUPO test sample study reveals common problems in mass spectrometry-based proteomics, *Nat. Methods, 6* (2009) 423–430.

41. K. Boutilier, M. Ross, A. Podtelejnikov, C. Orsi, R. Taylor, P. Taylor, D. Figeys, Comparison of different search engines using validated MS/MS datasets, *Anal. Chim. Acta, 534* (2005) 11–20.

42. H. Robenek, O. Hofnagel, I. Buers, M.J. Robenek, D. Troyer, N.J. Severs, Adipophilin-enriched domains in the ER membrane are sites of lipid droplet biogenesis, *J. Cell Sci. 119* (2006) 4215–4224.

43. H. Robenek, I. Buers, O. Hofnagel, M.J. Robenek, D. Troyer, N.J. Severs, Compartmentalization of proteins in lipid droplet biogenesis, *Biochim. Biophys. Acta, 1791* (2009) 408–418.

44. H. Robenek, O. Hofnagel, I. Buers, S. Lorkowski, M. Schnoor, M.J. Robenek *et al.*, Butyrophilin controls milk fat globule secretion, *PNAS, 103* (2006) 10385–10390.

45. B.M. Chong, P. Reigan, K.D. Mayle-Combs, D.J. Orlicky, J.L. McManaman, Determinants of adipophilin function in milk lipid formation and secretion, *Trends Endocrin. Metab. 22* (2011) 211–217.

46. B.M. Chong, T.D. Russell, J. Schaack, D.J. Orlicky, P. Reigan, M. Ladinsky, J.L. McManaman, The adipophilin C terminus is a self-folding membrane-binding domain that is important for milk lipid secretion, *J. Biol. Chem. 286* (2011) 23254–23265.

47. S. Murphy, S. Martin, R.G. Parton, Lipid droplet-organelle interactions; sharing the fats, *Biochim. Biophys. Acta, 1791* (2009) 441–447.

48. M. Beller, K. Thiel, P.J. Thul, H. Jäckle, Lipid droplets: A dynamic organelle moves into focus, *FEBS Lett. 584* (2010) 2176–2182.

49. S. Cermelli, Y. Guo, S.P. Gross, M.A. Welte, The lipid-droplet proteome reveals that droplets are a protein-storage depot, *Curr. Biol. 16* (2006) 1783–1795.

50. C.C. Wu, K.E. Howell, M.C. Neville, J.R. Yates, J.L. MacManaman, Proteomics reveal a link between the endoplasmic reticulum and lipid secretory mechanism in mammary epithelial cells, *Electrophoresis, 21* (2000) 3470–3482.

51. K. Dewettinck, R. Rombaut, N. Thienpont, T.T. Le, K. Messens, J. Van Camp, Nutritional and technological aspects of milk fat globule membrane material, *Int. Dairy J. 18* (2008) 436–457.

52. The Gene Ontology Consortium, Gene ontology: Tool for the unification of biology, *Nat. Genet. 25* (2000) 25–29.

53. The Database for Annotation, Visualization and Integrated Discovery (DAVID), SAIC-Frederic, Inc., Frederic, MD, USA *(http://david.abcc.ncifcrf.gov/home.jsp)*.

54. D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc. 4* (2009) 44–57.

55. Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization *(http://www.cytoscape.org)*.

56. M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Cytoscape 2.8: New features for data integration and network visualization, *Bioinformatics, 27* (2011) 431–432.

57. Ingenuity® Systems, Redwood City, CA, USA *(http://www.ingenuity.com/products/pathways_analysis.html)*.

58. P.D. Maningat, P. Sen, M. Rijnkels, A.L. Sunehag, D.L. Hadsell, M. Bray, M.W. Haymond, Gene expression in the human mammary epithelium during lactation: The milk fat globule transcriptome, *Physiol. Genomics, 37* (2009) 12–22.

59. P.D. Maningat, P. Sen, A.L. Sunehag, D.L. Hadsell, M.W. Haymond, Regulation of gene expression in human mammary epithelium: Effect of breast pumping, *J. Endocrinol. 195* (2007) 503–511.

60. K.A. Lê Cao, I. González, S. Déjean, IntegrOmics: An R package to unravel relationships between two omics datasets, *Bioinformatics, 25* (2009) 2855–2856.