Review

Applications of mass spectrometry techniques in the investigation of milk proteome

Vincenzo Cunsolo, Vera Muccilli, Rosaria Saletti and Salvatore Foti

Dipartimento di Scienze Chimiche, Università degli Studi di Catania, Viale A. Doria, 6, I-95125 Catania, Italy. E-mail: vcunsolo@unict.it

The introduction of “soft” desorption/ionization methods such as electrospray ionization and matrix-assisted laser desorption/ionization has determined a breakthrough in the application of mass spectrometry to the structural analysis of proteins. The contemporary advancement of bioinformatics, together with the possibility to combine these mass spectrometric methods with electrophoretic or chromatographic separation techniques has opened up the new field of proteome analysis and, more generally, has established these approaches as indispensable tools for protein and peptide analysis in complex mixtures, such as milk and milk-derived foods. Here, a necessarily not exhaustive series of current applications of mass spectrometry-based techniques for the characterization of milk proteins will be summarized. These include the characterization of milk protein polymorphism, determination of the structural modifications induced on milk proteins by industrial processes, investigation of milk adulterations and characterization of milk allergens.

Keywords: milk proteins, ESI-MS, MALDI-ToF-MS, tandem mass spectrometry, adulteration, milk proteins allergy, low-abundance proteins

Introduction

The development of the two “soft” techniques of desorption/ionization, electrospray ionization (ESI)\(^1\) and matrix-assisted laser desorption/ionization (MALDI),\(^2,3\) together with the fast-growing technological improvements in mass analyzers, tandem instruments and new fragmentation methods, has revolutionized the application of mass spectrometry to the structural analysis of proteins. The ability to provide accurate molecular mass determinations, together with high sensitivity, the unsurpassed feature for identifying post-translational modifications (PTMs) and the possibility of generating de novo amino acid sequence information, represent the unique features of these mass spectrometric methods. However, it should be noted that the study of proteins in general does not consist of the application of a single instrument, but requires complementary techniques and a series of tools, each of them best developed for a particular purpose. In this respect, the possibility to join mass spectrometry with electrophoretic or chromatographic separation techniques, together with the contemporary advancement of bioinformatics tools,\(^4\) has established these procedures as indispensable and irreplaceable tools to analyze proteins and peptide mixtures. These new analytical methodologies have been extensively applied to the investigation of proteins present in milk and derived products, due to their central role in nutrition.

The importance of milk in the human diet and in the world economy is well known and it is largely due to its unique nutritive quality, complexity and richness. Milk is the key element for infant nutrition as it represents the only complete source of all essential nutrients for newborns and infants, providing sugars, proteins, fats, vitamins and minerals for healthy...
Investigation of Milk Proteomes Using Mass Spectrometry Techniques

Investigation of Milk Proteomes Using Mass Spectrometry Techniques

Proteins represent one of the most important milk components from nutritional and physiological viewpoints. The high nutritional value of the milk protein fraction is related to the high content of essential amino acids. Moreover, milk proteins have native or latent biological functionality, being an important source of bioactive peptides. These peptides can be inactive as long as they are “hidden” in the primary structure of the proteins, but can become active upon hydrolysis by proteolytic enzymes during processing or gastrointestinal digestion. On the other hand, the wide use of cows’ milk in human diet has shown that a considerable percentage of subjects are allergic to its protein components. In particular, cows’ milk, used as substitute for breast feeding when mother’s milk is not available or advisable, represents the main source of allergens in infants. Therefore, a better knowledge of milk components, particularly proteins, is a priority for many researchers. Mass spectrometry-based techniques, which arguably represent the core tools in proteomic analyses, are widely used to characterize milk proteins not only in native fresh but also in processed milk.

Here, a necessarily not exhaustive series of current applications of mass spectrometry-based techniques for the characterization of milk proteins will be summarized. These include the characterization of milk protein polymorphism, determination of the structural modifications induced on milk proteins by industrial processes, investigation of milk adulterations and characterization of milk allergens. This overview will be preceded by a brief description of the classification and structure of milk proteins and of their molecular organization.

Milk proteins

Milk proteins can be grouped into three classes according to their different solubilities, which reflect their different structures and functional roles. These three groups of proteins, known as caseins, whey proteins and milk fat globule membrane (MFGM) proteins, can be separated as follows (Figure 1): (i) centrifugation of whole milk separates a fat layer containing the MFGM protein and a skimmed milk fraction; (ii) ultracentrifugation or acidification at pH 4.2–4.6 of the skimmed milk fraction separates caseins as an insoluble fraction and whey proteins as a soluble fraction.

The most abundant class of proteins in bovine milk is constituted by caseins (CNs), which are a group of proteins coded by four tightly-linked autosomal genes (CSN1S1, CSN1S2, CSN2 and CSN3). They are classically sub-divided according to the homology of their primary structures into four families: \( \alpha_S1 \), \( \alpha_S2 \), \( \beta \)- and \( \kappa \)-CN. Almost all caseins in milk are organized as macromolecular aggregates of proteins and minerals, known as the casein micelles. The structure of the micelles is not yet well recognized, but several studies have demonstrated a predominant surface location for \( \kappa \)-CN, which probably plays a fundamental role in stabilizing the micelle structure. The amount of CNs shows remarkable species variation; indeed, while they account for 80% (w/w) of all bovine milk proteins, in human mature milk, caseins represent only 35% and 50% of total protein content, respectively. Moreover, alignment of the amino acid sequences also shows considerable differences across species; as expected, the highest level of amino acid identities are observed amongst the ruminants (Table 1), but the identity of the primary structures rapidly decrease with the genetic distance between these and other species.

---

Figure 1. Analytical workflow for the separation of the three classes of milk proteins: MFGM proteins, caseins and whey proteins.
other species (Figure 2). Several studies have shown that caseins are quite small molecules with molecular masses of 18–25 kDa, with a great heterogeneity generated by post-translational processing, alternative splicing of the gene product or genetic polymorphisms, that affect the primary structure features and the quantity of each protein. All caseins are phosphoproteins, showing different levels of phosphorylation, which occurs at serine or threonine residues located in the Ser/Thr-Xxx-Glu/Asp/pSer motif (where Xxx is any amino acid residue and pSer is a phosphoseryl residue). In contrast, κ-cNs are the only group of caseins that also contain carbohydrate moieties. The carbohydrate groups are attached to the κ-CN via O-glycosidic bonds to serine and threonine residues present in the C-terminal region of the protein. The amino acid sequences of αs2- and κ-CN also contain cysteine residues, which are normally linked by intra- and inter-molecular disulphide bonds, whereas in the primary structures of αs1- and β-CN cysteine residues are absent. Caseins lack stable secondary and tertiary structures and, therefore, are easily susceptible to proteolysis. In this respect, the β-CN appears to be the most susceptible to the action of the endogenous milk protease plasmin. Plasmin cleaves at specific sites of β-CN producing a series of complementary C-terminal and N-terminal polypeptide fragments known as γ-caseins and proteose peptone (PP) components, respectively. These minor components are soluble in acidic condition and,

<table>
<thead>
<tr>
<th>Protein</th>
<th>Cow vs goat</th>
<th>Cow vs ewe</th>
<th>Cow vs human</th>
<th>Cow vs mare</th>
<th>Cow vs donkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>αs1-CN</td>
<td>86</td>
<td>83</td>
<td>27</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>αs2-CN</td>
<td>87</td>
<td>88</td>
<td>NA</td>
<td>57</td>
<td>60</td>
</tr>
<tr>
<td>β-CN</td>
<td>91</td>
<td>92</td>
<td>47</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>κ-CN</td>
<td>83</td>
<td>84</td>
<td>51</td>
<td>53</td>
<td>54</td>
</tr>
<tr>
<td>α-LA</td>
<td>95</td>
<td>94</td>
<td>76</td>
<td>73</td>
<td>72</td>
</tr>
<tr>
<td>β-LG</td>
<td>96</td>
<td>95</td>
<td>NA</td>
<td>52</td>
<td>53</td>
</tr>
<tr>
<td>SA</td>
<td>NA</td>
<td>91</td>
<td>75</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>LF</td>
<td>90</td>
<td>91</td>
<td>69</td>
<td>72</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 1. Sequence comparison by clustalW2 tool (http://www.clustal.org/) of major milk proteins from different mammalian species. Protein entries were retrieved from nrNCBI database.

Figure 2. Schematic representation of the phylogenetic branching tree diagram of cow and other mammalian species.
Investigation of milk protein polymorphism

The pioneering work reporting the use of mass spectrometry to analyze milk proteins was published in 1990 by Beavis and Chait, which demonstrated the capability of MALDI-MS to investigate the complex protein mixtures of two crude biological fluids such as a commercial bovine milk and human breast milk. Subsequently, numerous mass spectrometry studies have been carried out on milk proteins. They have been mainly dedicated to the investigation of milk molecular composition, including protein identification by molecular mass determination or sequence characterization by enzymatic mass mapping. Even if the first applications of ESI-MS were aimed at verifying the identity of milk proteins whose amino acid sequences were known, numerous different genetic variants, splicing variants, changes in the level of phosphorylation or glycosylation and localization of post-translational modifications were also detected. As a result, a number of genetic variants of both caseins and whey proteins from milk of various species are known today; they are classified by a logical nomenclature system which includes the protein family (i.e., $\alpha$-casein, $\beta$-casein, $\alpha$-lactalbumin etc.) followed by a progressive Roman letter (A, B, C, etc.) indicating the genetic variant, with a superscript (for example, $\beta$-CN A$^1$, $\beta$-CN A$^{25}$) if needed.

Increasing interest for an exhaustive characterization of casein fraction has occurred because it has been shown that changes in milk composition, caused by casein heterogeneity, deeply influence both the nutritional and the technological properties of milk. The extensive heterogeneity determines that over 150 protein spots can be readily detected following two-dimensional electrophoresis of whole bovine milk. One of the most remarkable features of this heterogeneity is due to post-translational modifications, such as the phosphorylation at serine, threonine and tyrosine residues. Investigation of the phosphorylation degree of caseins and localization of the modified sites can easily be performed by MS-approaches (Figure 3). In particular, since phosphorylation increases the molecular mass of 80 Da for each phosphate group, information about the phosphorylation level can be obtained by measuring (for example, by ESI-MS) the molecular mass of caseins before and after enzymatic dephosphorylation with alkaline phosphatase. However, unambiguous identification of the phosphorylation sites requires tandem mass spectrometry (MS/MS) analyses. Indeed, although in the fortuitous case of a mono-phosphorylated peptide carrying only one potential site of phosphorylation, the assignment of the modified residue is straightforward, in general, the presence of several potential sites of phosphorylation makes the MS/MS analysis of phosphopeptides an indispensable tool for the identification of the modified residue.

However, detection and characterization of the phosphopeptides is not trouble-free, as phosphorylation is often sub-stoichiometric. Hence, in a typical tryptic digest of a phosphorylated protein, the phosphopeptides are typically present in lower abundance than the unphosphorylated ones. In addition, in a typical positive ion mode acquisition of a MALDI mass spectrum, phosphopeptides show lower ionization efficiency with respect to their unphosphorylated counterparts. As a consequence, analysis of phosphopeptides often requires a preliminary step of enrichment. The most widely used method for selectively enriching phosphopeptides from a mixture is immobilized metal ion affinity chromatography (IMAC). Using this technique, phosphopeptides are enriched, or even purified, because they are selectively bound to immobilized metal ions via their phosphate moiety. Subsequently, they are eluted and characterized by mass spectrometry. An example of this strategy is provided by the primary structure characterization and identification of phosphorylation sites of an unknown variant of goat $\beta$-casein. Determination of the phosphorylation level was obtained by reversed-phase high-performance liquid chromatography (RP-HPLC)/ESI-MS analysis of both phosphorylated and dephosphorylated casein fractions. The results demonstrated that the unknown variant was present in two forms, corresponding to the species carrying five and six phosphate groups. Identification of the phosphorylated residues was achieved by tryptic digestion of the phosphorylated protein and MS analyses of the tryptic mixture. However, because the MALDI-MS analysis in positive ion mode of the tryptic digest did not show recognizable signals corresponding to possible phosphopeptides, in order to detect these fragments, a selective enrichment by IMAC was needed. MALDI-MS of the phosphopeptide-enriched fraction (Figure 4) showed only peaks attributable to mono- and multi-phosphorylated peptides and gave preliminary
information about the phosphorylation sites. In particular, the most intense signal at m/z 2061.7 was attributable to the expected tryptic fragment FQSEEQQQtEDELQDK (Mr 1980.8 Da) carrying a phosphate group. Indeed, an accompanying peak at m/z 1981.8, 80 Da lower, was also present in the spectrum and interpreted as having originated by loss of a phosphoric group of the ion at m/z 2061.7. Due to the presence of two potential phosphorylation sites (Ser and Thr residues), MS/MS analysis was necessary in order to determine the site of modification. The MS/MS spectrum (Figure 5) of the doubly charged molecular ion of a phosphorylated fragment with Mr 2060.7 Da showed, in addition to the base peak at m/z 982.7, corresponding to the loss of phosphoric acid from the molecular ion, a series of y and b ions consistent with the expected tryptic fragment carrying the modification at the serine residue.

Gel electrophoresis combined with MS analyses represents a complementary method of choice for identifying phosphorylated caseins and other post-translational modifications. This methodology, applied on ovine milk, ensured identification of more than thirty phosphorylated caseins (\(a_s -\), \(a_{s2} -\), and \(b -\)) and four \(\kappa -\)casein components, including non-allelic, differentially phosphorylated and glycosilated forms.\(^{29}\) Analogously, analysis of O-glycosylation site occupancy in bovine kappa-casein glycoforms was performed by coupling 2-DE separation, in-gel digestion and MS/MS characterization.\(^{30}\) These results showed that glycanes were not randomly distributed among the five potential glycosylation sites in kappa-casein. Rather, the observed hierarchy of site occupation between glycoforms argues for an ordered addition of glycans to the protein.

The extensive polymorphism of the casein fraction is also related to the presence at all the encoding loci of at least a “null” allele associated with the absence of the respective casein in milk. These “null” alleles may express truncated caseins. Indeed, the identification and characterization of a truncated goat \(\beta -\)casein associated with a null \(\beta -\)casein allele (CSN2\(^{25}\)) was reported.\(^{31}\) This protein, present as a minor component in the casein fraction of an individual goat milk sample, was characterized by the classical peptide mass mapping approach. MS-data of the tryptic digest demonstrated unequivocally that the protein sequence corresponded to the trait 1-166 of a mature \(\beta -\)CN variant A generated, as predicted at DNA level, by a premature stop codon. The absence of \(\beta -\)CN has direct consequences in cheese-making, as milks lacking this protein show a longer rennet coagulation time to normal milks and their curd firmness is much poorer.\(^{32}\)

Analogously, LC/ESI-MS was used to study the differences between the primary structures of two common genetic variants, A and C, and the less frequent variant D, of ovine \(a_{s1} -\)casein.\(^{33}\) This work revealed that these differences were simple silent substitutions, which, however, alter the degree of phosphorylation. As a consequence of these structural differences, milks
Figure 4. MALDI mass spectrum of the phosphorylated tryptic fragments of the β-CN variant D. The phosphorylated fragments were isolated and enriched by IMAC.

Figure 5. MS/MS spectrum of the doubly charged phosphorylated fragments with M, 2060.7 Da (see MALDI-MS in Figure 4).
containing the less common variant D have a tendency to coagulate and, therefore, result in a poor cheese yield.

However, it should be highlighted that, even if the most abundant proteins deeply influence both nutritional and technological properties of milk, its proteome is a more complicated system because of the existence of numerous low-abundant or trace components, which needed to be explored in order to have an exhaustive picture of the milk protein composition. Two MS-based approaches have recently been employed to explore the cows’ milk proteome. In the first, by coupling the use of combinatorial peptide ligand libraries, sodium–dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separation, in-gel digestion, nanoLC/MS/MS analysis and database searching, it was possible to identify a total of 149 unique protein species, 100 of which were not described previously.26 In the second, using an ion-exchange-based protein fractionation method prior to the mass spectrometric analysis, this list was implemented by the discovery of 176 new protein components in both colostrum and mature milk.31 Moreover, semi-quantitative analysis showed a number of up-regulated proteins in colostrum that are involved in the natural defenses of newborns. Finally, by compiling the wealth of data present in the literature dealing with massive proteomic analyses of milk fractions and independent targeted studies on specific groups of proteins, D’Alessandro et al.36 were able to tabulate an extensive list of 573 non-redundant annotated protein entries that could help in understanding the functional properties of both major and low-abundant proteins possibly involved in novel pharmaceutical, diagnostic and biomedical applications.

Investigations of milk proteins have mostly been focused on human27–41 and on the four ruminant species (cow, buffalo, goat and sheep), because of the large use of their milk in human diet. Recently, these studies have also been extended to equine species (i.e., mares and donkeys), especially because their milk is claimed to have special nutritional and therapeutic properties. Indeed, recent clinical trials have shown that, in many cases, equine milk represents a safe and alternative food in cows’ milk protein allergy (cMPA), providing nutritional adequacy and good palatability.57,58 Although the mechanism of this tolerance has not yet been fully clarified, it is reasonable to hypothesize that the reduced allergenic properties of equine milk could explain their better tolerance with respect to the cows’ milk proteins. Therefore, investigation of the allergenic properties of different types of milk is an important goal to achieve. For example, the recent detection of single protein components (for example, β-LG and αs1-CN), recently detected by coupling isoelectrofocusing (IEF) and mass spectrometry analysis,59 might prove to be useful in explaining the unsolved cases observed in clinical trials.

**Milk processing and protein modifications**

As reported in the Introduction, the importance of milk in human nutrition and in the world economy is well known. However, only a small amount of the total world production of
In milk, the reducing sugar is lactose that, reacting with milk proteins, leads to the formation of lactosylated protein species, whose molecular mass increases by 324 Da per lactose unit. Mass spectrometry, alone or coupled with liquid chromatography, has largely been used in the characterization of these derivatives and today plays a fundamental role in the monitoring of this reaction, which mainly involves α-lactalbumin, β-lactoglobulin, αs1- and β-caseins.59,60 In particular, the most abundant whey proteins, α-lactalbumin and β-lactoglobulin, have been used as molecular markers for modifications induced by different heat treatments on milk. Several MS-based studies have demonstrated that the degree of protein lactosylation is strictly related to the thermal procedures used during industrial milk processing.61-63 In fact, the amount of the lactosylated protein forms is low in pasteurized milk, whereas it accounts for almost 30% and 70% of the β-lactoglobulin content in UHT (ultra-high temperature) and dry infant formula samples, respectively.

The identification of modified sites within a protein is of high interest, because modification of amino acid residues may cause a serious decrease in the bioavailability of calcium64 and of the essential amino acid lysine. Moreover, modification of lysine also impairs the overall digestibility of milk proteins since lactosylated lysines are no longer recognized by gastrointestinal proteases.65 Detailed structural investigations of the modification sites carried out by mass mapping strategies have been able to localize the lactosylation sites in α-lactalbumin, β-lactoglobulin and αs3-casein present in either liquid and powder milk samples.62,64 Such approaches revealed the occurrence of preferentially lactosylated residues within the peptidic backbone, which reflects the surface exposure and the chemical micro-environment of the modified residue, but may also depend on a concomitant and significant denaturation, which involves partial tertiary structure unfolding.67

Protein cross-linking is another common feature of heated milk. Disulfide bond formation between β-LG and κ-casein occurs following denaturation of β-LG, which exposes a free sulfhydryl group and leads to both intramolecular and intermolecular thiol-disulfide exchange reactions.68 Non-disulfide cross-linking can occur via the formation of dehydroalanine69 which leads to intra- or intermolecular lysinoalanine, histidinoalanine or lanthionine bonds. Protein cross-linking can determine the formation of aggregates and insoluble precipitates, which also reduce the quality and functionality of milk and milk products. 2-DE coupled with MALDI time-of-flight (ToF) MS was used to examine the effects of storage at elevated temperature on UHT milk.70 The investigation resulted in the identification of three major changes that occur simultaneously in UHT milk during storage: deamidation, non-disulfide cross-linking and lactosylation. Interestingly, there was no evidence for these changes in freshly prepared UHT milk. They were only manifested upon storage and substantial levels were only observed in UHT milk stored at elevated temperatures. As highlighted by the authors, this result suggests that aging rather than processing could be the biggest contributor to the modifications of milk proteins.

It should be highlighted that, during heating treatments, low abundant proteins may also undergo Maillard reaction. Therefore, for a complete evaluation of the biological, nutritional and toxicological properties of differently heated milk samples, it is crucial to investigate the modifications induced by heat treatment in the milk “sub-proteome”. In a recent work, a combination of proteomic procedures based on analyte capture by combinatorial peptide ligand libraries, selective trapping of lactosylated peptides by m-aminophenylboronic acid-agarose chromatography, MS/MS analyses and database search, was used to investigate the lactosylated protein sub-population in powdered milk formula for infant nutrition.71 This approach allowed the identification of 271 modification sites in 33 milk proteins, including low-abundance components involved in nutrient delivery (i.e. lactotransferrin, apolipoproteins etc.), defence response against virus/microorganisms (lactadherin, polymeric Ig receptors, nucleobindin proteins, clusterin, serpins etc.) and cellular proliferative events (osteopontin, fibroblast growth factor-binding protein 1 etc.). These results indicate that the widespread lactosylation of milk proteins, which also involves minor components, may have important consequences on the nutritional and health characteristics of this food. Indeed, since it has been demonstrated that various milk components are resistant to proteolysis in the gastrointestinal tract and could constitute major food allergens,72,73 a reduced digestibility of lactosylated milk proteins may result in an increased allergic response to specific dairy products.

Investigation of adulterations in milk

Milk adulteration represents one of the most common types of food fraud. Among practices not allowed in the dairy industry, the most frequent are the fraudulent addition of low-cost milk to milk of higher costs and the undeclared use of...
admixtures of milk of different species for the production of traditional products protected by denomination of origin (PDO). As an example, a common fraud in the dairy industry is the addition of cheaper and more readily available bovine milk in non-bovine cheese-making. Therefore, the prevention of adulteration and contamination of milk is a matter of vital importance from both an economic and allergenic standpoint, and it is of increasing interest to develop fast and efficient approaches to assess authenticity of dairy products. Thanks to their fast response, MS-based techniques have been largely employed for dairy authentication purposes and they could be proposed as tools for rapid daily controls. Taking into account that the primary structure of homologous proteins shows species-specific amino acid differences which affect both their molecular masses and peptide mass fingerprinting, mass spectrometry-based methods are able to detect milk adulterations by characterizing the protein or peptide profile present in a specific product. On the other hand, as reported above, the thermal treatments used in industrial processes result in structural modifications of milk proteins such as the formation of lactosylated forms by the Maillard reaction. Therefore, the addition of UHT or powdered milk to samples of fresh raw milk can be evidenced by mass spectrometry. As an example, by monitoring the profile of the most abundant whey proteins (i.e. α-LA and β-LG), it has been demonstrated that MALDI-MS may represents a suitable method for the determination of the fraudulent presence of cows’ milk in ewe or buffalo products, or to identify the possible addition of powdered milk to samples of fresh raw milk. Two different approaches, based on RP-HPLC/ESI-MS and capillary electrophoresis (CE)/MS analyses, were recently developed for the quantification of cows’ milk adulteration in non-bovine milks (i.e. caprine and ovine). Using β-LGs as molecular markers and monitoring the adulterants by their retention time and molecular mass, the presence of cows’ milk at levels as low as 5% in the milk of other mammals was quantified. It is worth noting that, even if LC/ESI-MS and MALDI-MS could be considered as two complementary approaches, the great advantage of MALDI-MS over LC/ESI-MS is the speed of analysis and, therefore, it could represent a reliable method for routine analyses of the raw fresh milk samples that the dairy industry receives from producers every day.

Adulterations of dairy products can also be detected by a peptidomic approach, using MS/MS analyses to monitor species-specific peptides derived from enzymatic digestion. Recently, a method based on LC/ESI-MS/MS analyses of peptides of a casein extract from cheese was developed to recognize the authenticity and, consequently, to identify fraud in sheep’s and goat’s cheese-making. Specifically, a sheep-specific peptide, produced by in vitro plasmid digestion, was used as a marker for identifying undeclared addition of sheep’s milk to goats’ cheeses. Moreover, by monitoring the same peptide, it was also possible to detect if a sheep’s cheese declared “pure” really contained only sheep’s milk or was produced also using cows’ milk. The quantitative analysis was performed using as an internal standard (IS) an endogenous peptide common to all sample cheeses (goat, sheep and cow) which appeared accurate and was able to detect up to 2% of sheep’s milk in cheese. In a similar approach, based on milk protein trypsinolysis, MS detection of some species-specific casein peptides and quantification by synthetic analog IS peptides has been employed for the simultaneous detection of species-specific casein-derived peptides in ternary (or quaternary) milk mixture. As highlighted in this work, the presence of extraneous caseins in dairy products that are suspected of adulteration could be confirmed using a different set of species-specific proteotypic peptides, which more generally opens up the possibility of using MS-based methods for monitoring any protein of any food as long as its sequence is known.

### Milk proteins allergy

It is well known that milk is the main source of allergens in early childhood and, therefore, constitutes a problem of social relevance. Indeed, the wide use of cows’ milk (CM) in infant diets has shown that approximately 2-3% of infants younger than one year of age are allergic to cows’ milk proteins. This allergy is normally outgrown in the first year of life but 15% of allergic children remain allergic. The proteins most frequently and most intensively recognized by IgE are caseins and β-LG, even if lower abundant (i.e. lactoferrin, IgG and bovine serum albumin) and trace components appear to be potential allergens.

MS-based methods have largely been employed as confirmatory tools for unambiguous identification and/or characterization of milk allergens. In a general approach, allergens of milk protein extracts can be detected by two-dimensional electrophoresis (2D-PAGE) separation, electro-transferring onto a nitrocellulose membrane and IgE immunoblotting analysis with sera from allergic patients. Subsequently, identification of the candidate allergens can be easily performed by their in-gel enzymatic digestion, MS analysis and database searching. This approach has been employed to identify the most abundant cows’ milk IgE-reactive protein isoforms in 20 paediatric patients with documented IgE-mediated CMPA. The authors found that the prevalence of cows’ milk protein allergens was: 95% IgG heavy chain, 70% αs1-casein, 55% αs2-casein, 50% κ-casein, 50% lactoferrin, 45% β-lactoglobulin, 45% serum albumin, 15% β-casein and 0% α-lactalbumin. As highlighted by the authors, the 2D-immunoblot experiments were not in good agreement with the radioallergosorbent test (RAST), showing an evident discrepancy, particularly for α-lactalbumin and αs1-casein. Indeed, while in the immunoblot experiments identification of αs1-lactalbumin as an allergen was not achieved for all subjects, the RAST tests classified this allergen at least as class 2 (moderate level of allergen specific IgE) in seven patients. In contrast, the 2D-immunoblot of six patients with negative RAST results for total caseins, identified αs2-casein as an IgE-immunoreactive protein. The difference observed for α-lactalbumin could be related to the prevalent presence of...
conformational epitopes of this allergen, which are destroyed in the denaturing conditions of the immunoblot experiments. Instead, the different results obtained for α_{2}-casein could be explained by the lack of detection of some α_{2}-casein epitopes in the RAST analysis, which may, therefore, lead to false negative results and appears to not always be reliable for diagnosis purposes.

Taking into account the high incidence of CMPA in infants and considering that breast feeding is not always possible, indicated or sufficient, alternative supplies become indispensable. Therefore, one of the major objectives of the pharmaceutical industries is the production of milk and milk-based foods (i.e. infant formulas) close to breast milk. Usually, infant formulas are products based on bovine milk, which is modified by enzymatic and/or thermal treatments, and represent the preferred choices in the treatment of CMPA. However, it should be considered that allergies in these products is reduced, but never completely suppressed, and adverse reactions have also been experienced with these preparations. As a consequence, an investigation of the protein profiles of these products is needed. A MALDI-ToF MS approach was employed to compare the protein profile of different infant formula samples. Acquisition of the MALDI mass spectra of this kind of sample was more difficult with respect to the raw milk samples, because of the presence of high amounts of interfering compounds such as emulsifiers, vegetable oils, vitamins, metals etc. Therefore, while for MALDI-MS analysis of raw milk practically no sample pre-treatment is required, for the investigation of infant formula samples, a preliminary step of sample dialysis before MS analysis is necessary. This approach was used to investigate the protein profile of 11 infant formula samples and, therefore, to control if the declaration on the product label was in agreement with MS data. In particular, the MALDI-MS spectrum of a sample declared to be suitable for babies allergic to milk proteins, instead showed clearly detectable peaks for caseins. These results revealed that MALDI-MS analysis could be considered a powerful tool to quickly study the relationships between milk protein composition and clinical formula evaluations.

On the other hand, due to their wide range of nutritional, biological and functional properties, milk proteins are frequently used in the food industry for the preparation of different foods or beverages. Therefore, the detection of these so-called “hidden milk allergens” in non-dairy foods is needed for the food industry to ensure the correct labeling of their products and protect allergic consumers. Recently, several LC/MS-based methods have been developed with the aim of assessing the capability of these techniques in the detection of milk protein allergens in complex food matrices, such as fined white and red wines and mixed-fruit juice samples. As an example, a capillary LC separation combined with ESI-Q-ToF mass spectrometry for the detection and identification of α- and β-casein derived peptides in white wine samples fined with a caseinate standard white wine was developed. Identification of caseins was performed by MS/MS analysis and bioinformatic searching of specific peptide ions, arising from tryptic digestion. A specific ion belonging to a β-casein tryptic peptide was used as a caseinate marker to estimate the limits of detection of this method. This approach showed that the lowest concentration for caseinate added to wine was 50 µg mL^{-1}, a limit value that could be potentially lower by implementing peptide pre-concentration approaches and by using more sensitive MS acquisition modes (for example, SIM model). A different approach, based on the concentration of the combinatorial peptide ligand libraries (cPLLs) prior to electrophoresis and MS analysis has been adopted for harvesting and identifying traces of casein (usually used as fining agents) present in white wines. The CPLL technique was extremely efficient in capturing traces of proteinaceous additives (caseins) in white wines and could harvest as little as 10 µg casein with a minute amount of beads [200 µL] from large sample volumes [750 mL], the entire content of a bottle of wine, permitting an amplification of the signal of 5000 times and higher. In addition, it was shown that, in “wine-like” mixtures, added casein as low as 1 µg L^{-1} can be harvested efficiently and detected by coupling SDS-PAGE and MS characterization, after CPLL capture, via a silvering protocol. Similarly, the CPLL technique was adopted for identifying traces of protein present in red wines, leading to surprising results that, although it is stated that red wines are, in general, fined with egg albumin, for all the wines investigated it was found that the only fining agent used was bovine casein, the same as in white wines.

A different approach, based on solid-phase extraction and liquid chromatography coupled to mass spectrometry, was developed to detect traces of three allergenic cow milk proteins (α-LA, β-LGs A and B) in mixed-fruit juice samples. Specifically, these proteins were detected in a triple quadrupole mass spectrometer using both the classical full scan mode (FS) and the multiple ion monitoring (MIM) acquisition method. The FS mode was employed recording all the ions in the range 800–1800 m/z. The peak areas of the total ion chromatogram (TIC) were used for the construction of calibration curves. In contrast, the MIM acquisition was based on recording specific masses selected from the most abundant mass-to-charge ions for each protein of interest. The calibration curves obtained in full scan mode showed a good linearity and coefficient of correlation for all three proteins investigated demonstrating, therefore, that the developed method is a valid tool for the quantification of intact milk proteins. Cleaner chromatograms were obtained by using the MIM acquisition mode which resulted in an improvement in method selectivity. The limit of detection of this method was calculated to be around 1 µg mL^{-1} for each protein in mixed-fruit samples. This result represents an important issue since it has been reported that even small amounts of milk/whey proteins in the range of only a few µg mL^{-1} can trigger severe allergic reactions in sensitive consumers. Therefore, this strategy could represent a valid alternative to shotgun approaches, which involve a protein digestion step, a MS/MS analysis and, finally, a database searching for the identification of milk allergens in food products. Indeed, although these methods have proved to be reliable for the detection of milk allergens, their
capacity to provide quantitative information may be limited to
the unknown efficiency of the enzymatic protein digestion step
involved.

Finally, for the first time, a mass spectrometric method has been developed for the simultaneous detection of marker
peptides from seven sources of allergens (milk, egg, soy, hazelnut, peanut, walnut and almond) in an incurred bread
matrix. This method, based on extraction of the allergenic proteins from the food matrix, enzymatic digestion with trypsin
and on liquid chromatography and triple-quadrupole tandem
mass spectrometry operating in multiple reaction mode,
allowed the detection of concentration ranges from 10 µg g⁻¹
to 1000 µg g⁻¹ and below 10 µg g⁻¹ for milk, peanut and almond
respectively, demonstrating that the mass spectrometric
based method represents a useful tool also for simultaneous
multiple-allergen screening.

Conclusion

In this review article, a series of applications of mass spec-
trometry-based techniques employed in the last decade in
the investigation of proteins in fresh milk and processed dairy
products are reported. These examples, although not exhaust-
tive, show the enormous capability and the high versatility of
mass spectrometric methods in this field, especially when
coupled with separation techniques such as electrophoresis
or chromatography. It should be highlighted that the extensive
enlargement of the protein sequences database, made possible
by the increasing efficiency of gene sequencing methods,
together with the advancement of bioinformatics, has provided
an efficient platform to allow the processing of MS data for
large-scale protein investigation. With respect to traditional
methods, which frequently show poor inter-laboratory repro-
ducibility, troubles due to the interference of different compo-
nents present in the complex food matrices and low sensitivity,
mass spectrometry has unequivocally demonstrated that it
is a technique with exclusive features. MS techniques show
high specificity and sensitivity, high accuracy in the molecular
mass determinations and unsurpassed capability to identify
post-translational modifications (PTMs). Together, all these
features have established MS-based approaches as indis-
ispensable tools for protein and peptide analysis in complex
mixtures, such as those from milk and dairy products.

The interest in more detailed investigations of the protein
content of milk has originated for several reasons including a
more precise elucidation of the relationship between protein
composition and its nutritional and technological properties,
the production of methods for milk traceability, investigation of pro-
cessed milk or even for the detection of adulterated dairy products.
Moreover, considering that cows’ milk protein allergy is the most
common disease in early childhood, MS-based approaches are
successful in the detection and characterization of allergenic
proteins and, thus, they may contribute to the elucidation of the
structure-allergenicity relationships. In this respect, MS results
may provide the basis for the production of hypoallergenic or
nutraceutical milk-based foods (Pharma-Foods). On the other
hand, the recent identification of the low-abundant milk proteins
that constitute the so-called “hidden proteome” stimulate a more
comprehensive investigation of milk protein fraction, in order to
clarify the relationship that may occur between a protein profile
of milk and its health implications.

Acknowledgements

This work was supported by a grant from MURST (PRIN
2008, project number 20087ATS57) and FIRB “Italian Human
ProteomeNet” RBRN07BMCT.

References

1. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong and C.M.
Whitehouse, “Electrospray ionisation for mass spec-
trometry of large biomolecules”, Science 246, 64 (1989).
doi: 10.1126/science.2675315
2. K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida and T.
Yoshida, “Protein and polymer analysis up to m/z 100,000
by laser ionisation time-of-flight mass spectrom-
10.1002/rcm.1290020802
3. M. Karas and F. Hillenkamp, “Laser desorption ioni-
sation of proteins with molecular masses exceeding
ac00171a028
and bioinformatics tools for tandem mass spectrom-
etry in proteomics”, Physiol. Genomics 33, 18 (2008). doi:
10.1152/physiolgenomics.00298.2002
5. H. Meisel, “Biochemical properties of peptides
6. H.E. Swaisgood, “Chemistry of the caseins”, in Advanced
(2003).
8. M. Malacarne, F. Martuzzi, A. Summer and P. Mariani,
“Protein and fat composition of mare’s milk: some
nutritional remarks with reference to human and cow’s
milk”, Int. Dairy J. 12, 869 (2002). doi: 10.1016/S0958-
6946(02)010012-6
Dong and F.Z. Ren, “Composition, physicochemical prop-
erties, nitrogen fraction distribution, and amino acid
doi: 10.3168/jds.2006-600


Bibliographic details

Vincenzo Cunsolo graduated in Industrial Chemistry Summa cum Laude at the University of Catania in 1998. PhD in Chemistry at the same University in 2002, for his research on application of ESI mass spectrometry to the investigation of protein unfolding. In 2003, supported by an EMBO Fellowship, he worked in the group of Prof. Peter Roepstorff (University of Southern Denmark - Odense, Denmark) on the application of proteomics methodologies to durum wheat gluten proteins characterization. Assistant Professor of Organic Chemistry at the Department of Chemistry of the University of Catania since 2010. Expert in mass spectrometric applications in proteomics of cereals and milk.

Rosaria Saletti graduated in Chemistry Summa cum Laude at the University of Catania in 1986. In 1989–91 she was a post-doctoral fellow at the Department of Chemical Science, working on applications of FAB mass spectrometry to peptide and protein characterization. Since 1991, she has been an Assistant Professor at the University of Catania. Her current research interests are characterizations of wheat and milk proteins by liquid chromatography combined with ESI and MALDI mass spectrometry.

Salvatore Foti received his degree in Chemistry from the University of Catania, Italy, in 1973. In 1975, he worked in the group of Prof. Helmut Ringsdorf (University of Mainz, Germany) on the application of mass spectrometry to polymer characterization. In 1983, supported by a NATO fellowship, he worked at the Chemical Laboratory of the University of Cambridge, England, on the application of mass spectrometry to peptide characterisation, under the supervision of Prof. Dudley H. Williams. He is currently full Professor of Organic Chemistry at the University of Catania and coordinator of the Interdivisional Group of Protein Chemistry of the Italian Chemical Society. His research activity is mainly focused on applications of mass spectrometry to protein structural characterization and proteomics.
Catania—The Liotru. The symbol of the city is “u Liotru”, or the “Fontana dell’Elefante”, created in 1736 by Giovanni Battista Vaccarini. The Sicilian name “u Liotru” is a phonetic change of Heliodorus, a nobleman who, after trying without success to become bishop of the city, became a sorcerer and was therefore condemned to the stake. Legend has it that Heliodorus himself was the sculptor of the lava elephant and that he used to magically ride it in his fantastic travels from Catania to Constantinople.

Catania—The Ciclopi Islands Natural Marine Reserve. The creation of this archipelago was due to intense submarine volcanic activity, which affected this part of the coast 500,000 years ago. The islands of the Ciclopi are thought by many to be the inspiration for a specific section of “The Odyssey”, wherein the cyclops Polyphemus hurls stones at a fleeing Ulysses. The rocks are said to be the ones tossed by the monster. There is a shrine to the Madonna atop the southern-most island and a university-owned weather observatory on the larger middle island.

Catania—The University. The historical building of the University of Catania (founded in 1434).

Catania—S Agata. St Agatha, the patron saint of Catania.

Catania—Show of Etna. The show of volcano Etna. Although capped with snow for much of the winter, it is one of Europe’s most famous active volcanoes.