

Mass spectrometry in food proteomics: a tutorial[†]

Vincenzo Cunsolo,* Vera Muccilli, Rosaria Saletti and Salvatore Foti

In the last decades, the continuous and rapid evolution of proteomic approaches has provided an efficient platform for the characterization of food-derived proteins. Particularly, the impressive increasing in performance and versatility of the MS instrumentation has contributed to the development of new analytical strategies for proteins, evidencing how MS arguably represents an indispensable tool in food proteomics. Investigation of protein composition in foodstuffs is helpful for understanding the relationship between the protein content and the nutritional and technological properties of foods, the production of methods for food traceability, the assessment of food quality and safety, including the detection of allergens and microbial contaminants in foods, or even the characterization of genetically modified products. Given the high variety of the food-derived proteins and considering their differences in chemical and physical properties, a single proteomic strategy for all purposes does not exist. Rather, proteomic approaches need to be adapted to each analytical problem, and development of new strategies is necessary in order to obtain always the best results. In this tutorial, the most relevant aspects of MS-based methodologies in food proteomics will be examined, and their advantages and drawbacks will be discussed. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: MS; proteomics; food quality; food safety; food adulteration

Introduction to food proteomics

In the last decades, the rapid technological development in molecular biology together with the advances in large-scale protein analysis has contributed to move the attention of scientists from the characterization of individual proteins to a detailed examination of the entire proteome. In this respect, proteomics represents a natural evolution of protein chemistry and has gained impressive development principally as a consequence of the introduction of new ionization methods in MS [i.e. electrospray ionization (ESI) and matrix-assisted laser desorption/ionization for MALDI]. Thanks to the increased performance and versatility of the MS instrumentation, new protein analytical strategies have emerged and MS has become an indispensable tool in proteomic studies.^[1] Moreover, concurrent optimization of the separation techniques and the availability of efficient methods for gene sequencing, resulting in an extensive enlargement of the protein sequences database, have provided an efficient platform for large-scale protein investigation. As a general statement, the overall goal of proteomics is to understand the function of all proteins present in an organism. This aim implies protein identification together with the systematic determination of their different properties, including quantification, characterization of primary structure and post-translational modifications (PTMs), determination of protein interactions and subcellular distribution. The continuous and rapid evolution of proteomics has major implications for the understanding of molecular mechanisms in clinical diseases, but it is already yielding important findings across a wide range of applications in numerous fields such as pharmaceutical, microbiological, agricultural and food-technology. In particular, food proteomics is a part of a global discipline known as Foodomics, also including more working areas such as genomics, transcriptomics and metabolomics, which are aimed to the investigation of micro- and macro-components of foodstuffs.^[2,3]

Food proteomics research is mainly devoted to a detailed investigation of the protein content in foodstuffs.^[4,5] Knowledge of protein

composition is helpful for understanding the relationship between the protein content and the nutritional and technological properties of foodstuff, the production of methods for food traceability, the assessment of food quality and safety or even for the detection of genetically modified products and microbial contaminants. Moreover, considering that most of the more consumed foods (e.g. milk, eggs, cereals etc.) are responsible of allergenic reactions in humans, the detection of allergenic proteins represents a hot topic in the food safety field, because it may contribute to provide the basis for the production of hypoallergenic or to develop nutraceutical foods (Pharma-Foods).^[6]

From an analytical point of view, proteomic investigations are based on the use of highly efficient separation techniques such as reversed phase-high performance liquid chromatography (RP-HPLC) or two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) combined with biochemical methods, enzymatic digestion and MS in order to obtain data suitable for searching by specific software, genomic, protein or expressed sequence tags (ESTs) databases.

Given the high variety of the food-derived proteins and considering their differences in chemical and physical properties, it is evident that a comprehensive proteomic exploration does not consist in the application of a single technique for all purposes. Rather, proteomic studies use multiple technologies aimed to improve proteome resolution and coverage and to provide complementary results. An exhaustive review of such broad field is unfeasible in a journal article and does not represent the aim of this tutorial. Therefore, here, the most relevant applications of

* Correspondence to: Vincenzo Cunsolo, Dipartimento di Scienze Chimiche, Università degli Studi di Catania, Viale A. Doria, 6, I-95125 Catania, Italy. E-mail: vcunsolo@unicat.it

[†] This article is part of the Journal of Mass Spectrometry special issue entitled "3rd MS Food Day" edited by Gianluca Giorgi.

Department of Chemical Sciences, University of Catania, Viale A. Doria, 6, I-95125 Catania, Italy

MS-based methodologies in the analysis of food-derived proteins will be summarized, discussing their advantages and drawbacks. This section will be preceded by a brief description of the most common MS-based proteomic techniques, including the approaches and the bioinformatic tools for protein identification.

MS-based techniques for protein analysis

It is well known that the development in the late 1980s of the two 'soft' desorption/ionization MS techniques ESI and MALDI,^[7] which are capable to ionize polar and nonvolatile compounds such as proteins, DNA and carbohydrates without significant analyte fragmentation, has greatly broadened the applicability of MS to biology and revolutionized the analysis of biomolecules. With respect to the ESI method, MALDI MS is commonly used for the characterization of relatively simple mixtures of peptides or proteins; it is relatively resistant to interference with buffers commonly used in protein chemistry (e.g. phosphate, Tris, urea etc.) and produces mass spectra that are simple to interpret because of the tendency of the method to generate predominantly singly charged ions. On the contrary, the most important feature of the ESI is the generation of multiply charged ions. This feature allows the detection of proteins also using analyzers with limited mass range [e.g. quadrupoles and ion traps]. ESI gained immediate popularity because it can be easily coupled on-line with high-performance separation techniques such as capillary electrophoresis and HPLC and nowadays represents the most used MS-based method in proteomics. In particular, chromatographic separation represents an essential step during proteomic analysis because highly complex samples, such as a total cellular protein fraction or a protein mixture extracted from a food matrix, may contain hundreds of components that must be separated before the MS analysis. Analogously, when a biological sample is digested, the complexity of the protein mixture increases because each protein component yields many peptides and a high-performing separation step is needed. A variety of liquid chromatographic (LC) separation methods, including RP, strong cation exchange, affinity chromatography and size exclusion, have been developed. Recently, ultra-high-pressure liquid chromatography, operating with stationary phases consisting of small particles (size <2 µm), was introduced for more efficient and fast peptide separation.^[8] In order to improve the chromatographic resolving power, it is also possible to combine several separation techniques (i.e. multidimensional separation), in which each technique utilizes different physicochemical properties of molecules as a basis for separation. In this respect, the isoelectric

focusing or ion exchange chromatography may be joined with an RP liquid chromatography. This strategy has become a popular method in shotgun proteomics, and it is known as multi-dimensional protein identification technology (MudPIT; see in the succeeding text). On the other hand, the development of high-resolution ion mobility MS, which allows to separate ions on the basis of their size/charge ratios and their interactions with a buffer gas, adds a degree of orthogonality to MS, and it is now emerging as new powerful tool in the MS arena for the investigation of complex biological samples.^[9]

The implementation of separation techniques in miniaturized formats coupled on-line with high-performance mass spectrometers and the development of miniaturized ESI sprayers (nano-ESI) have reduced the amount of analyte required for complete and routine sequence characterization below the attomole (10^{-18} mol) range. Moreover, the improved performance and versatility of the mass spectrometers nowadays allow to measure the molecular mass of proteins and peptides with high accuracy and to determine additional structural features, which include the primary structure or PTMs. In this respect, the mass analyzer plays a fundamental role in the mass spectrometer technology, together with its key parameters such as sensitivity, resolution, mass accuracy and the ability to generate MS/MS spectra. Fundamentally, the mass analyzers currently used in proteomic studies are five: quadrupole (Q), Ion Trap (IT), time-of-flight (TOF), Fourier transform ion cyclotron resonance (FTICR) and Orbitrap. These mass analyzers differ considerably in design and performance, each with its own strength and weakness (Table 1). They can be stand alone or put together in tandem, in which each type takes advantage of a different physical property of the ion motion for mass-to-charge-based separation. IT, routinely coupled with ESI source, represents the unique analyzer which alone allows to determine either the mass of a given peptide and its sequence by MS/MS analysis. It is robust, sensitive and moderately inexpensive but shows a relatively low resolution and mass accuracy. Stand-alone TOF analyzer requires pulsed ionization methods such as MALDI, whereas it cannot be directly interfaced with ESI sources. On the other hand, hybrid mass spectrometers such as quadrupole-TOF (Q-TOF) and LTQ-TOF can be also coupled with ESI because both the front-end quadrupole and the IT work at the vacuum level compatible with ESI. Q-TOF and TOF-TOF instruments are the most suitable in proteome analysis because of their capability in performing both quantitative analysis and identification of PTMs. In particular, MALDI TOF-TOF technology represents a powerful tool for *de novo* sequencing of peptides coming from in-gel digestion (see in the succeeding text) because it offers the advantages of MALDI ionization and TOF detection such as high sensitivity, resolution and mass accuracy. The resulting fragment

Table 1. Comparison of the features and specifications of the mass analyzers commonly used for proteomics

Mass analyzer	Mass resolution ^a	Mass accuracy (ppm)	Sensitivity	<i>m/z</i> range	MS/MS capability	Ion source
IT	1000–1500	100–1000	Picomole	50–2000; 200–4000	MS ⁿ	ESI
LTQ (LIT)	2000	100–500	Femtomole	50–2000; 200–4000	MS ⁿ	ESI
Q-q-Q	1000	100–1500	Attomole	50–4000	MS/MS	ESI
TOF	10 000–20 000	5–50	Femtomole	No upper limit	n/a	MALDI
TOF-TOF	10 000–40 000	5–50	Femtomole	No upper limit	MS/MS	MALDI
Q-q-TOF	10 000–40 000	5–50	Attomole	No upper limit	MS/MS	ESI; MALDI
LTQ-FTICR	50 000–800 000	1–2	Femtomole	50–2000; 200–4000	MS ⁿ	ESI; MALDI
LTQ-Orbitrap	50 000–500 000	<5	Femtomole	50–2000; 200–4000	MS ⁿ	ESI; MALDI

^aMass resolution at full width half maximum.

ion spectra of TOF/TOF instruments are often more extensive and informative than those generated by trapping instruments. Moreover, the theoretically unlimited mass range of TOF and the high collision energy used to produce fragmentation make possible MS/MS analyses of larger peptides and intact proteins, thus also allowing to perform top-down proteomics.

The FTICR^[10] and Orbitrap^[11] analyzers allow to measure the m/z value of the ions with the highest resolution and mass accuracy of any analyzers used in MS (Table 1). The most common MS platforms equipped with these mass analyzers are the LTQ-FTICR and the LTQ-Orbitrap. In these systems, the LTQ mass analyzer is used for ion accumulation, ion storage and ion fragmentation in MS/MS experiments. The LTQ-Orbitrap instruments show slightly lower resolution and mass accuracy than LTQ-FTICR MS but offer higher scan rate, lower cost and lower maintenance requirements. For these reasons, the LTQ-Orbitrap platform is becoming the universal solution when high resolution and mass accuracy are required in proteomic researches.

On the other hand, it is important to highlight that in proteomic experiments, fragmentation methods for MS/MS play a key role for protein or peptide sequencing and in PTM characterization. Collision-induced dissociation (CID) is the most widely applied fragmentation method for MS/MS analysis, resulting in cleavage of amide bonds to yield *b*- and *y*-type ions. CID is more effective for small, low-charged peptides, whereas it is not suitable for fragmentation of intact proteins and peptides with labile PTMs. In addition, in spite of its prevalence, CID provides limited sequence coverage when applied to peptide sequences carrying many basic residues. The recent development of the new fragmentation techniques of electron-capture dissociation (ECD), electron-transfer dissociation (ETD) and higher-energy C-trap dissociation (HCD) has greatly enlarged the capabilities of MS/MS strategies.^[12] The 'soft fragmentation' methods of ECD and ETD provide better sequence coverage for small-sized to medium-sized peptides, including a more comprehensive characterization of PTMs. Moreover, because these methods lead to the cleavage of the N-C α backbone bonds and generate *c*- and *z*-type fragment ions, they are highly complementary to conventional

CID.^[13] The HCD technique is performed with higher activation energy and shorter activation time compared with the traditional IT CID. The HCD, such as the CID method, generates *b*- and *y*-type fragment ions, with a predominance of *y*-ions; *b*-ions can be fragmented further to *a*-ions or smaller species. Moreover, the lacking low mass cut-off restriction, typical of the traditional IT CID, together with the high mass accuracy of MS/MS spectra, contributes to provide more informative ion series.^[14] In general, several studies demonstrated that ETD is particularly useful for peptides with charge states higher than two, whereas HCD provides more peptide identifications than CID and ETD for doubly charged peptides. Therefore, taking into account the complementary fragmentation information that can be obtained by these methods, the use of CID, ETD or HCD alone, in alternating acquisition or by a decision-tree regulated combination, can provide the best performance for the analysis of distinct peptide populations.^[15]

Strategies in proteomics: bottom-up versus top-down

Given the high variety of MS-based techniques available for large-scale protein investigation, proteomic studies are usually carried out by the use of multiple technologies aimed to provide high-throughput protein identification, protein expression and protein-protein interaction, therefore improving proteome resolution and coverage. As previously reported, proteomic experiments involve the simultaneous investigation of several thousands of protein components and require a separation step (i.e. gel electrophoresis and liquid chromatography-based methods) in order to simplify the complex mixtures of peptides or proteins that are delivered to the mass spectrometer. The choice of the separation method or the general strategy applied depends on the experimental question and the scope of the research. Therefore, two complementary MS-based approaches are currently used in proteomics: the so-called bottom-up and top-down (Fig. 1).^[16] The first method uses one or more proteases to digest

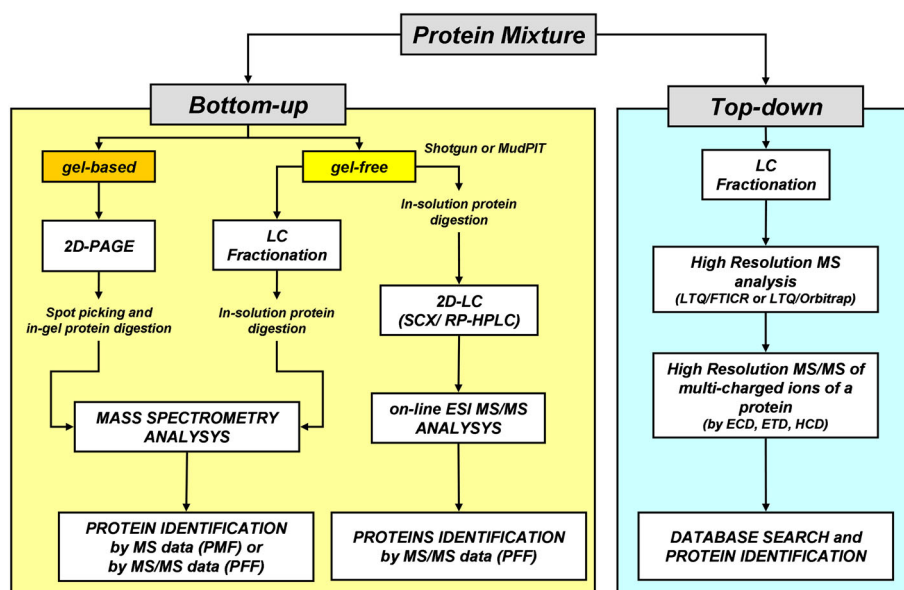


Figure 1. Comparison between the two most common strategies for MS-based protein identification and characterization: the bottom-up and top-down.

the proteins into peptides for subsequent MS and MS/MS analyses. Instead, the 'top-down' strategy directly analyzes intact proteins and their fragmentation ions, generated inside the mass spectrometer (mainly FTICR and LTQ-Orbitrap), without prior digestion.

Bottom-up proteomics

The bottom-up proteomic strategy represents the traditional approach for large-scale or high-throughput analysis of highly complex samples and can be carried out by two different ways: the 'break-then-sort' and 'sort-then-break' approaches. In break-then-sort approach, the protein mixture is directly digested into a collection of peptides, which are separated by multidimensional chromatography (e.g. strong cation exchange followed by RP-HPLC) on-line coupled to ESI MS/MS analysis.^[17] This method is known as 'shotgun' approach or MudPIT.^[18] Although the shotgun approach is conceptually simple, it results in greatly increased complexity of generated peptide mixture. As a consequence, not all peptides resulting from the digestion of a protein can be observed or correctly identified by MS analysis. Moreover, only the peptides present at high relative abundance are preferentially sampled, whereas information regarding the proteins represented as low abundance peptides in the complex mixture is commonly not obtained. However, the MudPIT approach is now emerging in the proteomics field, thanks to the high-resolution power of ultra-high-pressure liquid chromatography systems, the high mass accuracy of LTQ-Orbitrap instruments and the high performing of ETD technique.

In sort-then-break approach, proteins are firstly separated by 2D electrophoresis; then, the isolated proteins are subjected to enzymatic digestion followed by direct peptide mass fingerprinting (PMF)-based acquisition or further peptide separation on-line coupled to MS/MS (i.e. the peptide fragment fingerprinting). The gel-based approach, due to the unsurpassed ability of 2D electrophoresis to resolve complex protein mixtures, constitutes the long-established and more popular separation method in proteomic studies. The popularity of 2D gel-based approaches in proteomics is mainly related to the reproducibility and robustness of the electrophoresis technique, as well as to the ease of performing quantitative analysis and interfacing with immunoblotting techniques. On the other hand, it is well documented that gel-based proteomics shows two main limitations. The first one regards the analysis of membrane proteins, having poor solubility.^[19] The second important drawback is related to very high concentration range of proteins present in a biological system (e.g. up to 10–12 orders of magnitude in human sera). Consequently, only the most abundant proteins may be visualized in 2DE maps, whereas the low-abundant or trace components, including also allergen species, could remain undetected. In this respect, it should be noted that, independently from the methodological approach, the detection and identification of low-abundant proteins represent a major challenge in proteomics. During the last years, several strategies, such as prefractionation and immunodepletion methods, have been carried out to remove the most abundant proteins and for enhancing the sensitivity and efficiency in the identification of low-abundant proteins. Very recently, the characterization of minor components in complex protein systems has been revolutionized by the introduction of the combinatorial peptide ligand libraries technology (CPLL), commercially named ProteoMiner.^[20] The major advantage of this technique, in comparison with other

prefractionation methods, is that it not only reduces the concentration of the more abundant proteins but also concentrates low-abundance and even trace components, thus providing access to the 'invisible' proteome.

As previously mentioned, bottom-up proteomics experiments require a lot of laborious steps in order to get from the protein to the peptide level. The crucial step in such experiments is the protein digestion, which is often the bottleneck in terms of time consumption. The use of trypsin is still considered the gold standard in bottom-up proteomics, but in specific cases, peculiar characteristics of the target proteins (e.g. amino acid composition or hydrophobicity) may require the selection of most suitable proteolytic enzymes or multiple digestion strategies. For example, it is very reasonable that proteins focused in 2DE spots with extremely basic pI will probably contain many basic residues (i.e. arginines and lysines) and few acid residues. Conversely, a very acidic protein will be constituted by a lot of aspartic and glutamic acids and few arginines and lysines. Consequently, the uncritical use of trypsin may cleave a very basic protein into a high number of small-sized peptides and a highly acidic protein into a few large peptides. It is well known that too small and too big-sized peptides are not the best target for MS and MS/MS analyses performed in the most common mass spectrometers (i.e. ITs or quadrupoles). Therefore, alternative proteases, such as Lys-C or Glu-C, or a combination of two different enzymes, may be needed. A detailed comparison of the capabilities of the different protein digestion techniques is out of the purpose of the present tutorial. However, an excellent survey about the currently available digestion strategies and recent developments in the acceleration of digestion processes was recently published.^[21]

Top-down proteomics

Top-down strategy represents the emerging MS-based approach in proteomics, providing information on intact protein mass and its amino acid sequence. It involves gas-phase ionization of intact proteins and their direct MS and MS/MS analysis without prior proteolytic digestion. Particularly, proteins in complex mixtures are firstly fractionated and separated into pure single proteins or less complex protein mixture. This step is followed by off-line static infusion or by on-line LC/MS analysis of sample into the mass spectrometer for high-resolution mass measurement of intact protein ions. MS/MS of the mass-selected multicharged ions of a protein then provides fragment mass values for its structural characterization.^[22] The key to the success of this approach is an efficient fragmentation of intact proteins, as well as the exact selection of the multicharged protein ions and then the identification of the charge states of the resulting fragments ions. These objectives can be achieved by exploiting the high mass accuracy and high resolution of LTQ-FTICR, Q-TOF or LTQ-Orbitrap mass spectrometers, improved further by the electron fragmentation (i.e. ETD or ECD) methods. In general, the top-down approach, with respect to the classical bottom-up, may offer a higher protein sequence coverage, an improved protein quantification and more comprehensive characterization of PTMs.^[23]

However, to date, top-down is not as widespread as bottom-up for some reasons, and therefore, the largest numbers of proteomics experiments employ 'bottom-up' approaches. As mentioned in the previous text, top-down analysis requires a prefractionation step of protein samples in order to obtain more simple protein mixtures. Therefore, even if some studies have shown that large-scale profiling of intact proteins is feasible, most

top-down applications are targeted studies of purified proteins. In this respect, separation of proteins represents one of the most important problem for top-down analysis. Traditional methods such as gel electrophoresis have limitations on the recovery of proteins in a form suitable for analysis by MS and frequently result in sample losses. On the other hand, chromatographic methods used to separate intact proteins may show poor resolving power for large proteins (>50 kDa). Another drawback to the natural evolution of top-down is the limited number of bioinformatic tools for data management and interpretation. Moreover, the indispensable use of high mass accuracy instruments may represent a restrictive requirement in top-down approach. Indeed, these mass spectrometers are expensive to purchase and to maintain if compared with quadrupole and IT instruments usually employed in bottom-up proteomics. Protein size may represent another important limitation in top-down proteomics. Indeed, ECD and ETD methods provide efficient fragmentation for small to medium-sized proteins.^[24] Direct analysis of large proteins (>50 kDa) is more complicated because of increased complexity of the gas-phase protein ion's tertiary structure with many noncovalent interactions. To address this limitation, the so-called middle-down strategy, involving limited digestion to produce larger peptides (>5 kDa), has been explored for very large (>200 kDa) proteins.^[25] This approach joins the best features of top-down and bottom-up approaches. Middle-down proteomics takes its advantage by the high resolution and mass accuracy of MS instrumentation and by the high performance of electron-based fragmentation methods. In ESI ionization, medium-sized peptides (4000–10 000 Da) carry out a higher number of charges with respect to smaller ones, thus enhancing the pattern fragmentation by ECD, ETD and HCD methods and therefore improving protein sequence coverage and identification of PTMs. At the same time, the middle-down approach shows advantages related to the chromatographic analysis of peptides, which are much more efficiently separated than intact proteins because of their narrower mass distribution, charge and hydrophobicity.

In light of their most important features (including advantages and drawbacks), it is obvious that top-down and bottom-up represent complementary approaches, as evidenced by a number of proteomic studies performed by coupling these two strategies. In general, the bottom-up method is widely accepted for the routine identification of proteins in complex mixtures. If, however, more extensive or specific data are needed, such as on polymorphisms or PTMs, the complementary top-down approach can often provide these in a very straightforward manner.

In any case, and independently from the adopted proteomic approach, it is important to highlight that the generation of the MS data is far from being the end of the experiments in proteomics but rather represents an indispensable precondition for undertaking the second level of proteomic studies, the protein identification by the high-throughput bioinformatic methods, which will be in brief discussed in the next paragraph.

MS data analysis and protein identification

If the modern MS instrumentation allows to reach an impressive level of protein characterization, it has not been neglected that handling and storing very large-scale data were made possible by the contemporary improvement of bioinformatic tools.^[26] In this respect, interpretation of MS data represents another fundamental step in proteomics for the successful analysis of proteins,

including their identification, determination of their expression level, characterization of PTMs, determination of protein interactions and subcellular distribution.

In general, the characterization and the identification of a protein may be obtained by comparing MS experimental data with calculated mass values obtained from a sequence database using a dedicated search engine (e.g. Mascot, Ms-Fit, ProFound, MassSearch etc.) (Fig. 2).^[27] In particular, it is well known that the enzymatic digestion of an isolated protein with a protease of known cleavage specificity generates a set of peptides representing its fingerprint and therefore permits the unambiguous protein identification. In bottom-up proteomics, this objective can be reached by two different methods, known as peptide-mass fingerprinting (PMF, or peptide mapping) and peptide-fragmentation fingerprinting (PFF), respectively. In the first method, a list of enzymatic peptide fragment masses present in a MALDI-TOF mass spectrum is matched to that calculated from the same proteolytic digestion of each entry in a sequence database. The success in the identification of the protein depends on several factors, including the specificity of the protease used, the number of experimental peaks observed and the mass accuracy of the mass spectrometer. In the PFF method, database searching is performed using both the measured peptide mass and the list of peptide fragment ions produced by MS/MS. Because the MS/MS spectrum of a peptide is strictly related to its amino acid sequence, the specificity of protein identification is, in principle, much higher than that of peptide mapping. Moreover, because MS/MS data can also be used to search translated ESTs and other sequence databases containing incomplete sequences, the PFF approach is much more comprehensive than PMF. In top-down proteomics, MS/MS data of the intact protein ions can be used in a similar way and, together with the experimentally high resolution measured M_r , allow protein identification.

Irrespective of the approach adopted, the successful identification of the protein requires the presence of the corresponding protein sequence in the database. If the amino acid sequence of the protein under investigation is not present in the database, the best match will probably be the entry with the closest homology and usually corresponds to a protein belonging to a strictly phylogenetic-related species. If the sequence similarity within protein databases is too low, database search is unsuccessful, and MS/MS data must be interpreted (*de novo* interpretation) manually or through sequencing algorithms (e.g. PEAKS, PepNovo etc.). This approach aims to reconstruct the peptide sequence directly from an MS/MS spectrum without the aid of protein sequence databases. Then, deduced peptide sequences can be employed in identifying proteins by sequence similarity searches using conventional database search algorithms such as BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) or FASTA (<http://www.genome.jp/tools/fasta/>) or by MS-driven sequence similarity searches (MS-BLAST; <http://genetics.bwh.harvard.edu/msblast/>), and may enable, in principle, identification of proteins also for species phylogenetically distant from organisms with completely sequenced genomes.^[28]

Finally, although MS-based approaches are mainly used to address qualitative aspects of proteomics (i.e. the identification and characterization of proteins), quantitative information at the protein level is very helpful to investigate and measure how the expression of proteins under different conditions (differential proteomics) changes. Quantitative proteomics can be performed to obtain both absolute (using internal standards) or relative quantification of proteins by different techniques including

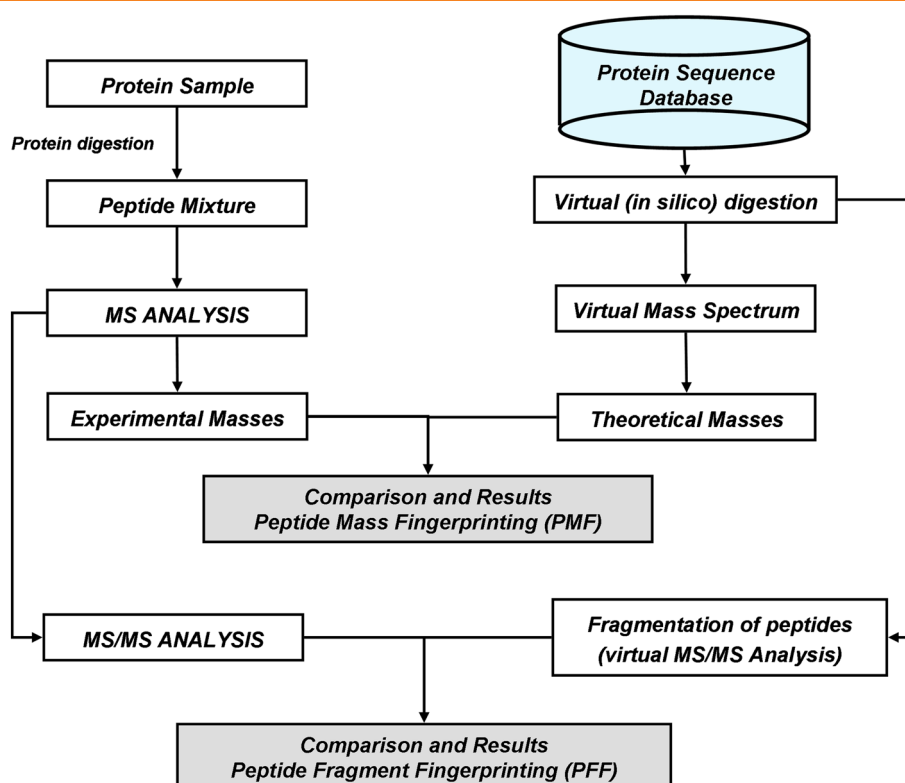


Figure 2. Flowchart of the two database interrogation methods by MS data that can be applied for protein identification.

gel-based, label-based and label-free approaches. Briefly, 2D-PAGE gel quantification is based on the comparison of the intensity of the spot among different samples (control vs case). The 2D fluorescence difference gel electrophoresis, even if does not eliminate the problem of comigrating proteins, gives more accurate and reliable quantification information about protein relative abundance because the samples to be compared are run together on the same gel and avoid potential gel-to-gel variation. Label-based methods involve the labeling of peptides with stable isotopes introduced by either biosynthetic or chemical reactions. Quantification is based on the ratio of heavy/light peptide pairs. On the contrary, MS-based label-free quantitative proteomics avoids the use of isotopes to label the samples under investigation. Quantification is based on the spectral counting or signal intensity of identified proteins after MS/MS analysis of their proteolytic peptides. A detailed description of quantitative proteomics approaches can be found in some review articles.^[29,30]

MS-based strategies in the characterization of food proteins

Protein profiling and food quality

Basic interest in food proteomics relies on the fact that protein pattern influences nutritional, rheological and sensory properties of food products. One of the most important examples of the correlation of protein profile with technological properties is represented by cereal proteins. Proteomic analysis of cereals is currently applied to both major and minor species such as wheat (*Triticum aestivum* and *durum*), oat (*Avena sativa*), rye (*Secale cereale*), barley (*Hordeum vulgare*), maize (*Zea mays*) and rice

(*Oryza sativa* L.). The main drawback in the application of proteomic approaches in cereals, and in general in plant-derived foods, is that the complete genome sequence of many crops is still not known. Therefore, the proteomic investigations need to be performed mostly by means of MS/MS and *de novo* sequencing approaches. In addition, over 80% of the plant genomes are typically constituted by repetitive transposable elements. Moreover, polyploidy is another challenge to overcome for many cultivated crops (e.g. wheat, potato, tomato etc.) and even fruit crops (such as banana or strawberry), thus requiring independent sequencing of the various wild-type haplotypes. Consequently, plant proteomics is at early stage of progress with respect to the animal proteomic studies even if the situation is improving with the numerous ongoing sequencing projects.

Wheat, which represents the most widely grown, processed and consumed cereal by humankind, shows a very complex protein profile, including structural and metabolic proteins, protective proteins and storage proteins (gliadins and glutenins).^[31] Glutenins include the high molecular weight gluten subunits (HMW-GSs; M_r 70–90 kDa) and low molecular weight gluten subunits (LMW-GSs; M_r 30–40 kDa), which are organized in huge polymers by intermolecular or intramolecular disulfide bonds, responsible of the viscoelastic properties of flours. In particular, it was demonstrated that allelic variation in the polypeptide composition of HMW-GSs is closely correlated with the bread-making quality of wheat cultivars and their crosses.^[32] The MALDI mass spectrum of a crude extract with the complete HMW-GSs pattern profile has been proposed as a routine method in breeding programs for rapid identification of lines containing subunits related to quality and for varietal identification.^[33] The early mass spectrometric works on glutenins were focused on the determination of the molecular masses of HMW-GSs, which are easy to isolate as

single subunits. The studies on the identification of HMW-GSs masses by MALDI MS was motivated by the anomalous migration of these subunits in SDS-PAGE, which results in apparent molecular masses higher than those calculated from gene sequencing. However, it is important to highlight that MALDI MS determination of the masses of HMW-GSs, although much more accurate than that obtained by conventional electrophoretic methods, does not provide mass unit accuracy. Therefore, it does not allow either the unequivocal identification or the exclusion of possible PTMs. Direct verification of HMW-GSs cDNA derived sequence, can be performed by tryptic digestion or cyanogen bromide cleavage of the proteins and subsequent analysis of the proteolytic mixtures by MS. The presence of large repetitive domain and the low content of cleavable tryptic sites determine high molecular mass tryptic peptides, which can be characterized by coupling the RP-HPLC/ESI-MS/MS data with MALDI mass spectra acquired both at low and high-molecular mass range. In this respect, a number of papers have clearly demonstrated that high sequence coverage, up to 95%, can be easily obtained, verifying the correctness of the cDNA derived sequences, also allowing detection of some differences, consisting of point modifications, insertion or deletion of short sequences in the repetitive domain.^[34,35] In addition, the obtained results have highlighted the complete absence of glycosylation or other PTMs, thus clarifying a major controversial point in the structural characterization of these proteins.

According to their molecular masses and to the number of Cys residues, the other group of prolamins, the LMW-GSs, are classified in the B, C and D types. Therefore, coupling online RP-HPLC/ESI-MS and offline RP-HPLC MALDI-TOF MS approaches on intact LMW-GSs, it is possible not only to measure the molecular mass of these subunits but also to determine the Cys number by comparison of the molecular masses of the alkylated and unalkylated subunits.^[36]

However, with respect to HMW, characterization of LMW-GSs profile is more complex, because these subunits are encoded by a high number of genes with similar structural characteristics, resulting in the expression of several LMW-GSs with very strong similarity in both primary structure and molecular masses. As a consequence, to investigate such proteins, classical proteomic approach needs to be adapted. In the electrophoretic separation of an LMW-GSs fraction, a single 2D-gel spot may frequently contain more proteins belonging to the same group of LMW-GSs and differing by point substitutions. Moreover, because of a long repeating motif and a scarcity of cleavable tryptic sites, typical of these proteins, the number of the resulting tryptic peptides suitable for MS analysis is scant, and the use of alternative proteases (e.g. chymotrypsin and thermolysin)^[37,38] may be helpful. A practical way to identify LMW-GSs is to perform MS/MS experiments on the few available tryptic peptides and subsequent interpretation of the obtained results in the light of the peculiar primary structure of this class of proteins. Specifically, it should be considered that (1) sequence coverage will obviously be scarce owing to the long repeating motif lacking tryptic cleavage sites, and (2) the few peptides with masses suitable for MS/MS fragmentation (occurring in the N-terminal and C-terminal domains) would not be able to discriminate the single subunits because they are common to groups of LMW-GSs. Studies on LMW-GSs are usually performed in order to gain cultivar characterization^[39] and to better understand the role of LMW-GSs in gluten matrix and seed maturation by means of comparison between common wheat cultivars and translocated or transgenic ones.^[40,41]

The molecular characterization of gliadins by MS is a research topic of increasing interest both for their involvement in the rheological properties of flours and for triggering celiac disease. The same considerations previously reported on LMW-Gs characterization, including the complexity of the protein mixture and the very low number of the resulting tryptic peptides suitable for MS analysis, can be also applied to the study of these monomeric proteins. Therefore, a substantial amount of information on the complex heterogeneity of gliadins was performed using proteases alternative to trypsin (e.g. chymotrypsin) coupled to database searching and *de novo* sequencing of the MS/MS data.^[42]

Another field in which the contribution of MS is of primary interest for food technology is milk protein chemistry. 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. MS approaches have mostly been focused on milk from humans and from four ruminant species (cow, buffalo, goat and sheep) because of the large use of their milk in human diet.^[43] Recently, these studies have also been extended to both major^[44–46] and minor protein components^[47] of equine species (i.e. mare and donkey), especially because their milk is claimed to have special nutritional and therapeutic properties.

Milk proteins can be grouped into three classes according to their different solubility, which reflects their different structures and functional roles. These three groups of proteins, known as caseins (CNs), whey proteins and milk fat globule membrane proteins, can be easily separated and characterized. CNs show an extensive genetic polymorphism, which is also related to the presence at all the encoding loci of at least a 'null' allele, which determines the absence of the respective CN in milk, or may express a truncated CN form. Because changes in milk composition, mainly caused by CN polymorphism, influence deeply both the nutritional and the technological properties of milk and milk-derived products, MS approaches could be used for a fast screening of protein composition and therefore could provide suggestion for the destination of milk for drinking or cheese-making. As an example, investigation of the CN fraction of an individual goat milk sample by means of LC/MS, trypsin digestion and MS/MS analysis allowed the detection and sequence characterization of a truncated β -CN, associated with a "null" β -CN allele and generated by a premature stop codon.^[48] The absence of the full-length β -CN has direct consequences in cheese-making, as milks lacking of this protein show a longer rennet coagulation time to normal milks and their curd firmness is much poorer.

The extensive polymorphism of CN fraction is related not only to genetic polymorphism but also to differential splicing events occurring during primary transcript processing. Consequently, an individual milk sample may contain different CN isoforms with very strong similarity in their primary structure. The complete separation of these isoforms by classical chromatographic methods is therefore often hampered, and a complete sequence characterization may result very tricky. In this respect, a methodological approach, which involves the combined use of RP-HPLC, 2DE and mass spectrometric analyses may be very helpful for the resolution and characterization of protein isoforms present in a complex mixture, for which RP-HPLC separation alone fails,^[49,50] and also for identifying PTMs. This methodology, applied on ovine milk, ensured identification of more than 30 phosphorylated CNs (α_{s1} -, α_{s2} - and β -) and four κ -CN components, including nonallelic, differently phosphorylated and glycosylated forms.^[51]

Several proteomics studies have been performed on the meat protein fraction. The main object in meat proteomics is the elucidation of mechanisms behind meat storage and processing, which affect technological properties, tenderness, color and thus their quality. Because a technique for measuring tenderness on the living animal is lacking, tools to estimate the potential of tenderness from the live animal or carcass are needed. In general, the strategy is based on 2D-gel comparison followed by MS identification of protein muscles from two groups (eg. very tender and not tender). As an example, *longissimus dorsi* tenderness-related parameters of Maremmana, an Italian autochthonous and highly appreciated beef breed, was integrated with proteomics data.^[52] In this work it has emerged that the phosphorylation of muscle proteins plays an important role in the postmortem process and hence in meat quality. Therefore, through TiO₂ preliminary enrichment of phosphopeptides, followed by MS analysis via CID and ETD, it was possible to detect higher levels of glycolytic enzymes, which were less phosphorylated and overall more active (lactate accumulation was higher in the tender group) than in tough counterparts.

Another important research topic in food proteomic comprises fruit tree tissues, including leaves, roots and fruits, one of the most complex fields in plant physiology. Since fruit tree crops are agricultural products of relevant economic importance, a great effort has been made to understand the molecular mechanisms involved in fundamental biological processes in fruit tree physiology and fruit biology. Both genomic and proteomic studies are gaining an important impact on fruit research and therefore the number of research papers is considerably growing.^[53,54]

Protein modifications induced by food processing and storage conditions

The manufacturing processes and storage conditions may induce significant food changes, including heat-shock protein responses, proteolytic degradation and many nonenzymatic PTMs (nePTMs), such as condensation, elimination or hydrolysis of side chains, which render the food proteome even more complex with respect to the raw and fresh materials.^[55] Among the different treatments used in the food industry, thermal processing (refrigeration, freezing, cooking, pasteurization and sterilization) are the most widely used. These procedures are employed to improve safety, organoleptic and nutritional characteristics of food, or to reduce bacterial load and extend the shelf life of foodstuffs. However, upon these treatments, chemical reactions can take place among the main components (proteins, lipids and carbohydrates) negatively influencing the product quality and safety. Therefore, investigation of these reactions and localization of the modified amino acid residues is of particular importance. The main modifications induced by heat treatment are structural changes, such as protein unfolding and aggregation, and a complex series of reactions known as 'Maillard reaction'.^[56] It is an nePTM, which involves mainly the lysine ϵ -amino group and the carbonyl group of reducing sugars, forming stable protein-sugar derivatives (Amadori compounds). Because this reaction reduces the bioavailability of the essential amino acid lysine, it can impair the nutritional food value and even may have adverse health effects. Indeed, modification of lysine harms the overall digestibility of milk proteins because lactosylated lysines are no longer recognized by gastrointestinal proteases. In milk, the covalent glycoconjugate-condensation products derived by the reaction between the amino groups of the whey proteins and lactose lead

to the formation of lactosylated protein species, whose molecular mass increases by 324 Da per lactose unit. Using the most abundant whey proteins (i.e. α -lactalbumin, α -LA, and β -lactoglobulin, β -LG) as molecular markers, many studies have demonstrated that MS plays a fundamental role in the monitoring these reactions. MS-based approaches allow to investigate the thermal history of commercial milk samples, evidencing that the degree of protein lactosylation is strictly related to the storage conditions and the thermal procedures used during industrial milk processing.^[57] 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 -LG content in ultra high temperature and in dry infant formula samples, respectively. During heating treatments, less abundant proteins may also undergo Maillard reaction and therefore should be investigated. With this purpose, a combination of proteomic procedures based on capture by combinatorial peptide ligand libraries, selective enrichment of lactosylated peptides by m-aminophenylboronic acid-agarose chromatography and CID and ETD fragmentations has been applied with the aim to investigate the modifications induced by heat treatment in the milk 'sub-proteome' of powder milk formula for infant nutrition.^[58] This methodology allowed to observe an exhaustive modification of proteins, which also included low-abundance components involved in nutrient delivery, defense response against virus/microorganisms and cellular proliferative events and therefore may have important consequences on the nutritional and health characteristics of this food. Even if the lactosylated protein forms (especially for α -LA and β -LG) represent the most monitored markers in heated milks, other different compounds are formed and may be used to reveal the presence of heated or powder milks added to fresh milks. An example of these products is the unnatural amino acid lysinoalanine, which is not present in both raw milk or fresh water Mozzarella cheese, but it is originated upon heating treatments by the reaction of the conjugated carbon-carbon double bond with the nucleophilic amino group of lysine.^[59] As demonstrated in several MS-based trials, high amounts of lysinoalanine are found in calcium caseinates and milk powders, and therefore, it represents a very useful marker for detection of heat treated milk in many dairy products.^[60]

Furthermore, the Maillard reaction between some amino acids (i.e. asparagine, glutamine or methionine) and reducing sugars (e.g. D-fructose, lactose and glucose) can lead to the formation of harmful acrylamide (AA) in food during roasting, toasting and frying processes.^[61] AA is neurotoxic, carcinogenic and also a toxicant for the reproductive tract of animals, binding the cysteine sulphhydryl groups of hemoglobin (Hb) according to a Michael-type addition reaction. HPLC/ESI-MS and MS/MS analyses at protein and peptide levels represent reliable approaches aimed not only to the identification of the specific Cys residues susceptible of alkylation but also to obtain quantitative information about the AA-Hb adducts and permit to monitor the level of human exposure.^[62]

The light exposure and oxygen are other factors that may induce significant food changes such as an oxidative damage of proteins and lipids. As an example, protein carbonyls can be formed directly by oxidation of the side chains of several amino acids, such as lysine, arginine, proline, histidine and threonine.^[63] Therefore, levels of protein carbonyl content could be considered useful markers for monitoring protein oxidation in food products such as milk^[64] and meat^[65] because they are strictly correlated with the severity of the treatment. The classical bottom-up

proteomic approaches were also able to identify protein candidates responsible for deterioration of fish quality upon oxidation conditions, revealing an increased carbonylation of the structural proteins myosin and actin, the integrity of which is essential for the textural characteristics of muscle tissues.^[66]

Finally, it is important to highlight that some of the adducts formed during food processing and storage could lead, in an advanced stage, to cross-linked protein species having a strong resistance to proteolytic digestion.^[67] Consequently, these products, which are resistant to digestion, could also represent potential food allergens. For these reasons, monitoring food protein oxidative modifications could be very useful regarding both product safety and allergenicity issues.

Characterization of food-derived allergens

One of the main goals in food science is to provide accurate and precise methods to detect harmful compounds that might be present in food. In this respect, the detection of food allergens is a hot topic in the food safety field. Indeed, allergic reactions to foods represent an actual and increasing problem in clinical medicine. European legislation recognizes so far 14 major allergenic foods^[68] among which milk, wheat, soybeans, peanuts, tree nuts, fish and shellfish are included. The methods commonly used to detect allergenic proteins in foods are ELISA, PCR and RT-PCR, with the latter two methods detecting DNA markers but not the proteins.^[69] Quantification by immunological

methods is conditioned by the complexity of the food matrix, the presence of allergenic contaminants derived from other foods and the variable specificity of the antibody preparations of commercial ELISA kit. MS overcomes these difficulties, allowing unambiguous detection of single or multiple allergens and their isoforms even in processed products, where physical or chemical modifications might affect protein stability. The application of proteomic strategies to identify food allergens is referred as 'allergenomics'. The known food allergens are frequently identified by a typical bottom-up approach: MS analysis of the peptide mixtures deriving from protein digestion and database searching. The approach aimed to discover new allergens combines electrophoresis separation, electro-transferring onto nitrocellulose membrane and IgE immunoblotting analysis with the sera of allergic patients. The potential allergens can be identified by in-gel enzymatic digestion, MS analysis and database searching (Fig. 3). As for general proteomic application, identification of allergens is complicated by the lack of entries in databases when the genome is not fully sequenced. In such case, *de novo* MS/MS sequencing is very helpful in identifying allergens.

Among allergenic foods, milk is the main source of allergens in early childhood: approximately 2–3% of infants younger than 1 year of age are allergic to cow's milk proteins. This allergy is normally outgrown in the first year of life, but 15% of allergic children remain allergic. The proteins recognized by IgE are mainly the CNs and the β -LG, but even lower abundant components (i.e. lactoferrin, IgG and bovine serum albumin) appear to

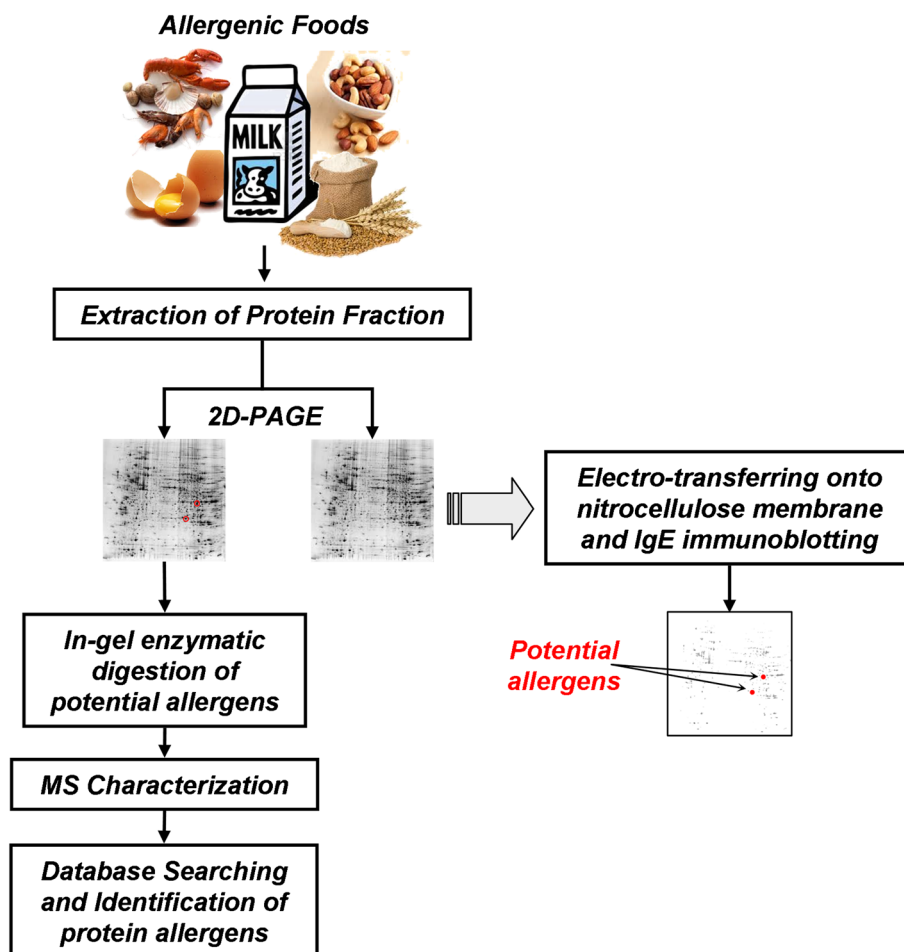


Figure 3. Flowchart for the identification of allergens.

be potential allergens.^[70] Data obtained by proteomic approaches in order to identify the most abundant cow's milk IgE-reactive protein isoforms in pediatric patients with documented IgE-mediated Cow's Milk Protein Allergy (CMPA), when implemented with radioallergosorbent test may show, in some cases, discrepancies with the immunoblotting results.^[71] The observed differences between proteomic and immunological results are probably due to the presence, in some allergens, of conformational epitopes that are destroyed in the denaturing conditions of the immunoblotting experiments. Considering that breast feeding is not always possible, indicated or sufficient, for infants allergic to cow's milk proteins, alternative supply becomes indispensable. Therefore, one of the main goal of the pharmaceutical industries is the production of milk and milk-based food-stuffs (i.e. infant formulae) close to breast milk. 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 allergenic properties of these products are reduced, but never completely suppressed, and adverse reactions have been experienced also with these preparations. A MALDI-TOF MS approach can be easily employed for the protein profile of different infant formula samples in order to control if the declaration on the label is in agreement with the product composition.^[72]

In many cases, donkey's milk represents a safe and alternative food in CMPA, providing nutritional adequacy and good palatability.^[73] It can be reasonable to hypothesize that the reduced allergenic properties of equine milk may be related to structural differences of its protein components with respect to the bovine counterpart. Indeed, the recent sequence characterization of donkey's α_{s1} -CN^[49] evidenced that it shares a low sequence homology (41% of identity and 57% of similarity) with the bovine counterpart. In particular, these differences are remarkable for the IgE-binding linear epitopes of cow's α_{s1} -CN and the corresponding domains present in donkey's α_{s1} -CN. In most cases, donkey's milk represents the unique alternative food in CMPA, even if it is known that a minor number of subjects do not tolerate donkey's milk. Because most of the studies carried out to investigate the allergenic proprieties use bulk milk, instead of an individual one, the investigation of individual milk samples is advisable in order to highlight the specific role of single proteins and the impact of the polymorphism on allergenic traits. Therefore, as recently demonstrated by coupling isoelectrofocusing and MS analysis,^[74] the existence of individual donkey's milk samples lacking of single protein components (e.g. β -LG and α_{s1} -CN) might also explain some unsolved cases observed in clinical trials of milk allergy.

On the other hand, milk proteins are frequently used in the food industry for the preparation of no dairy foods or beverages. Therefore, detection of these so-called hidden allergens in no dairy foods is needed not only for ensuring the correct labeling in food industry but also to protect allergic consumers. As an example, the use of combinatorial peptide ligand libraries was employed for capturing traces of proteinaceous additives (CNs) in wine harvesting as little as 10 μ g CN with 200 μ l of beads from 750 ml of wine, and allowing a signal amplification of 5000 folds. Subsequently, after SDS-PAGE and MS analysis, it was possible to detect in both white and red wine traces of CNs with a detection limit of 1 μ g/l, revealing that red wines had been fined with bovine CN even when the typical fining agent is egg albumin.^[75]

Investigation of 'hidden' allergens can also be performed by monitoring target peptides rather proteins. As an example, by a

peptidomic approach, and employing LC/MS and LC/MS/MS with a Q-TOF analyzer, it was possible to identify and monitor specific biomarkers for the confirmation and quantitative determination of hidden allergenic peanut proteins in different food matrices.^[76] Four peptides obtained from tryptic digestion of Ara h 2 and Ara h 3/4 proteins were identified and investigated as biomarkers using rice crispy and chocolate-based snacks as model food matrices. The selectivity of the method was enhanced by the choice of two peptide targets for allergen identification, thus combining a better limit of detection for the method (down to 1 ppm) with the capability of performing quantitative analysis.

Finally, because nePTMs occurring during technological processes may affect the allergenic properties of a protein, as previously reported, monitoring and characterization of these reactions are needed in order to assess the allergenic properties of processed foodstuffs. As an example, in contrast to other reports, a recent study^[77] showed that deglycosylation did not affect the affinity of β -conglycinin (i.e. the major allergens from soybean) to IgE of patients allergic to soybean, suggesting that the glycosylation should not be involved in allergy. Analogously, many researchers aimed to understand the role and the contribution of glycosylation to allergenicity of egg proteins have provided contradictory results.^[78] These conflicting findings clearly evidence the complexity of explaining and understanding the structural bases of food allergy but at the same time may encourage researchers to enhance their efforts in the study of this challenging task.

Detection of microbial contaminants in foodstuffs

Man is exposed to a wide range of pathogenic microorganisms, which can be introduced in the body mainly through food, water, air and dermal contact. For these reasons, food safety has become a challenging field, and the government of several countries have increased the amount of legislation with regard to this field.^[79] In 2000, the EU Commission published a White Paper on food safety to ensure safe products along every step 'from farm to fork'.^[80] In addition, every stage of the food chain, from raw material production to consumption of the finished product, is under the control of the Hazard Analysis and Critical Control Point system and is monitored scrupulously.^[81] In 2002, the European Food Safety Authority was created to evaluate the risk to human health related to the food and to provide scientific advices. In order to enhance food safety, manufacturers and processors adopt the good manufacturing practices to avert and reduce food borne hazards.

Applications of MS-based approaches in food safety are concerned with the detection, identification and quantification of contaminants, such as pathogens, which can cause food spoilage and can be hazardous to human health. Food spoilage is a process arisen from different biochemical alterations due to microbial activities. More than 250 known pathogens, mostly microbes and their toxins, cause food poisoning. Among them, there are *Listeria Monocytogenes*, *Staphylococcus aureus*, *Clostridium botulinum* A, *Campylobacter jejuni*, some *Salmonella* species and *Bacillus* strains, and *Escherichia coli*. All these pathogenic bacteria excrete a variety of virulence factors into the extracellular medium and to the cell surface, which have essential roles in the colonization and insurrection of the host cells, thus reflecting the degree of bacterial pathogenicity. These toxins, being heat-stable and resistant to proteases, represent a serious danger for the consumer health. Although there are well-established and sensitive immunochemical methods for detection of bacteria

and toxins, MS-based proteomics methodologies have revealed to be a powerful tool for the identification of microbial food contaminants. Indeed, when detection of enterotoxins is not possible with commercial assays or may be difficult by immunological methods, MS-based approaches may represent a competitive alternative in their specific identification. As an example, cytotoxin K1 (CytK1) and nonhemolytic enterotoxin (NHE) are two toxic compounds produced by pathogenic strains of the *B. cereus* group,^[82] for which no commercial assays (CytK1) or specific immunological kits (NHE) are available. In these cases, alternative tools, such as MS, for their detection and identification are highly desirable.^[83]

Most proteomic studies have been directed to the development of MS-based methods for bacterial profiling in order to distinguish among different species and also strains.^[84] In this respect, it was showed that bacteria harbored in fish and other foods, which cause histamine poisoning, exhibited highly specific MS protein fingerprints that enable their unambiguous differentiation.^[85] Consequently, these results evidence that MS may be used for a rapid and specific identification of major biogenic amine-forming bacteria, thus providing a timely prevention of histamine food poisoning. As a consequence, different commercial databases have been developed for bacterial identification by MALDI-TOF MS. Recently, a new public reference library, Spectrabank (<http://www.spectrabank.org>), which contains the mass spectral fingerprints of the main spoilage-related and pathogenic bacteria species from seafood, and includes 120 species of interest in the food sector, was created.^[86]

Several bacteria, including *Pseudomonas aeruginosa*, *S. aureus*, *Staphylococcus epidermidis* or *E. coli* are the major cause of mastitis in dairy cows. This infection is the most costly disease that affects dairy industry because it leads to decreased milk production, changes the protein profile and affects milk quality. Changes in whey protein profile of bovine milk samples during induced *E. coli* mastitis have been investigated by MS-based methods. These approaches allow to reveal a higher expression of some acute phase proteins,^[87] which may represent possible biomarkers for future research into the effects of bacterial inflammation during mastitis and can be useful for the evaluation of the efficacy of new and more specific therapies.

On the contrary, some bacterial species exert various health benefits in humans. In fact, by liberating lactic and acetic acids, they prevent the colonization of potential bacterial pathogens in the gastrointestinal tract, thereby maintaining a balance of normal intestinal flora. Bifidobacteria and lactobacilli are the most popular microorganisms that are added as live bacteria for the production of probiotic dairy foods. MS-based methods allow rapid classification and identification of lactic acid bacteria (LAB) strains from fermented foods,^[88] investigation of the protein expression and verification of the resistance of wild-type or heat-shock tolerant Bifidobacteria to conditions of food manufacturing and of the gastrointestinal tract.^[89,90]

Finally, all the proteomic studies about the microorganism food contamination may be useful in order to enable more rational design of food preservation methodologies or to identify the food chain stages that are most susceptible to microbial contamination.

Characterization of food-derived bioactive peptides

Nowadays, there is a growing interest in promoting functional foods to gain healthy benefits beyond basic nutrition. The term

food-derived peptides indicates different short amino acid chains (2–20 amino acids, in some case up to 40) that are inactive within the sequence of native protein but can become active upon hydrolysis by proteolytic enzymes during processing or gastrointestinal digestion. Some of the common sources of bioactive peptides are milk, egg, fish, soybean, rice sunflower, cereals and maize proteins,^[91] but other organisms, such as algae and organisms originated from the marine environment (eg. sea urchin,^[92] sea cucumber^[93]) have been explored as new source of bioactive peptides.

Bioactive peptides can be produced from the protein precursor by *in vivo* proteolysis (gastrointestinal digestion) or during food processing (ripening, fermentation and cooking), storage or *in vitro* proteolysis due to the presence of endogenous enzymes in the food matrix.^[94] In addition, LAB, the microorganisms most widely used as starter cultures for the production of fermented foods, also contribute to the release of bioactive peptides from dietary proteins. Therefore, microbial fermentation especially by LAB acquires a major role, because the production of key bioactive peptides can be triggered during foodstuffs production.^[95]

Biological activity, attributed to different bioactive peptides, may encompass, among others, antihypertensive, antihypertensive inhibitors of angiotensin-converting enzyme, antioxidant, dipeptidyl peptidase IV, opioid agonistic and antagonistic, hypocholesterolemic, immunomodulatory, antimicrobial, antithrombotic, mineral binding, anti-obesity activity and cancer prevention. An exhaustive description of food bioactive peptides is beyond the scope of this tutorial. However, an excellent survey about the enzymes used for hydrolysis, food processing conditions and the size of the resulting peptides was recently published.^[96]

The approaches currently followed to discover and identify bioactive peptides in food matrix are both *in silico*, on the basis of the prediction of the peptide activities, and *in vitro*. The downstream processing sequence usually consists in (1) food protein selection; (2) enzymatic hydrolysis, fermentation or gastrointestinal digestion; (3) structural characterization of peptide mixture by MS/MS (e.g. nLC/ESI MS/MS); (4) *in silico* screening for the potential bioactivity properties of sequenced peptides; and (5) production of analogous synthetic peptides to validate the activity (Fig. 4). In particular, MS is currently the elective technique to profile food hydrolytic peptides and for the targeted characterization of bioactive peptides. However, it should be taken into account that typical MS-based proteomic approaches need to be adapted. Indeed, a typical bottom-up approach starts with a 2-DE separation analysis which, as previously reported, shows limitation for the characterization of low abundant and trace components, proteins with extreme isoelectric point, or very low molecular weight components such as some bioactive peptides. Therefore, a gel-free approach may be considered a more performing alternative. Another drawback is related to the intrinsic nature of bioactive peptides. In fact, identification and quantification of proteins are usually obtained by matching several peptides that can be unambiguously related to one parent sequence. On the contrary, bioactive peptides are active molecules *per se*, so their identification could be more challenging. For very small peptides, particularly di-, tri- and tetra-peptides, not only HPLC separation has to be adapted, possibly by coupling additional separation techniques, but also MS/MS spectra, if performed at low collision energy, may result of poor quality and less informative.^[97] Therefore, the use of the database searching software (e.g. MAS-COT) may be adapted selecting no enzyme specificity and not

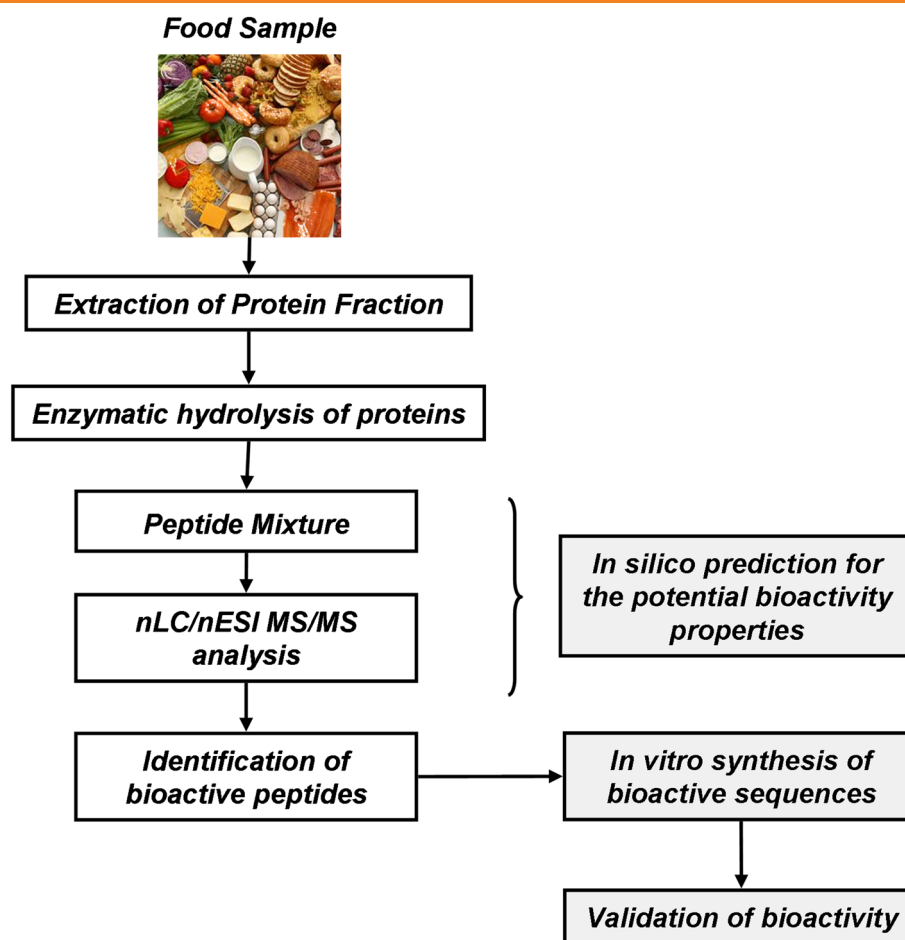


Figure 4. Food bioactive peptides analysis workflow.

considering the scoring results. On this respect, and especially for bioactive food-derived peptides originated by organism with a not fully completed genome annotation, *de novo* sequencing by MS/MS interpretation, followed by BLAST search against organism for which genome is available, probably represents the best way for their characterization.

Some databases, collecting peptide sequences, activity, reference works, source and more, e.g. BIOPEP (www.uwm.edu.pl/biochemia/index.php/pl/biopep), PepBank (<http://pepbank.mgh.harvard.edu>), ADB (<http://aps.unmc.edu/AP/main.php>) and others, are useful resources for bioactive peptides study.

According to recent literature, only few of the food-derived peptides in the gastrointestinal tract can escape the digestion and therefore are absorbed in the intestinal lymphatic system, a step required to carry out some specific activities at systemic level.^[98] Despite extensive information of the *in vitro* properties of food-derived bioactive peptides, the knowledge about the peptide availability and transport is continuously growing. Consequently, mass spectrometric approaches dedicated to the identification of peptide sequences released in blood are only the first step of a more complex system necessary to assess potential bioactivities.

Taking into account all these technical considerations, it is clear that characterization of bioactive peptides and assessment of their biological activity are challenging tasks. Moreover, several difficulties arising from the complexity of food matrix, the large variety of digestion products and the interference of endogenous

compounds in biological fluids have to be considered when performing a 'peptidomic' study.

Food authenticity and detection of adulterations

The numerous proteomic technologies today available together with the exhaustive investigation of food proteins allow the monitoring of specific components in order to assess the authenticity or to detect food-adulteration. Many foodstuffs are regularly adulterated for short-term economic gain, without any concern for the potential health consequences. Milk adulteration represents one of the most common types of food frauds. Among the 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. Because the primary structure of homologous proteins shows species-specific amino acid differences, which affect both their molecular masses and PMF, MS-based methods are able to detect fraudulent actions in milk and dairy products by the investigation at both protein or peptide level.

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 MS may represent a suitable and rapid method for the determination of the fraudulent presence of cow milk in ewe or buffalo products or for identifying the possible addition of powdered milk to samples of fresh raw milk.^[99] Moreover, MALDI MS is able

not only to identify the fraudulent addition of milk from different species in donkey's milk^[100] but also to recognize unintended contaminations up to limits ($0.5 \div 2.0\%$) comparable with that of the more laborious and time-consuming currently used methods. Adulterations of dairy products can also be detected by peptidomic approaches, using LC/MS/MS analyses to monitor species-specific peptides derived from enzymatic digestion.^[101] The presence of extraneous proteins (i.e. CNs from different species) in protected by denomination of origin dairy products, which are suspected of adulteration, could be confirmed using many sets of species-specific proteolytic peptides and MS-based methods.

Even cereals represent a target for fraudulent actions. In cereals, prolamin profile is determined in order to investigate traceability and authenticity. The molecular characterization of these protein components in local wheat varieties is an irreplaceable way to trace the local plant biodiversity that needs to be recognized, categorized and preserved. As an example, Timilia is an old spring Sicilian durum wheat landrace highly resistant to drought and biotic and abiotic stresses, which shows some valuable qualitative characteristics, such as a long shelf-life of the derived baked products. Semolina derived from Timilia is used to produce a traditional Sicilian bread named Pane Nero di Castelvetro (Black Bread from Castelvetro, Trapani), whose price is higher than that of the usual bread. In light of these peculiar features, Timilia can thus be subject to mixing or substitution with cheaper flours. The characterization of HMW-GSs profile at both protein and peptide levels represents a fingerprinting of each wheat variety and can be applied to reveal the presence of the most common *Triticum durum* Desf. cultivars in a mixture with the more fine Timilia.^[102]

The remarkable progress in plant genomics and proteomics also provided the necessary data for the investigation of plant-derived proteins present in both alcoholic and nonalcoholic beverages in order to certify the genuineness of such products and find out if they contain proteins of the vegetable extracts from which they have been prepared. In this regard, the CPLL technology allowed to explore the trace proteome of orgeat syrup,^[103] detecting a handful of proteins (just 13), all belonging to a bitter almond extract, and consequently verifying the genuineness of such product. On the contrary, the same approach carried out on cheap orgeat syrups produced by local supermarkets allowed to reveal that these products did not contain any residual proteins, suggesting that they were likely produced only with synthetic aromas and no natural plant extracts. The CPLL technology coupled with MS performed in order to investigate the trace proteome of aperitifs, alcoholic beverages (with 15–20% alcohol content) in general made with herbal infusions, highlighted in many cases severe doubts on the procedure declared by the manufacturers.^[104]

Another important issue concerns the geographical origin of food products. In fact, assessment of geographical origin is one of the main requirements for the certification of wine authenticity, grape variety and production technology. However, the assessment, by proteome analysis, whether the geographical origin of wines as stated in the label is accurate or whether the wines have been counterfeited, may be hampered by several factors. One of the main troubles regards the potential marker which may be used to assess the geographical origin: the grapevine proteins. Indeed, grapevine proteins survived to vinification process slowly aggregate and cause turbidity in the final product. A haze or deposit in bottled wine indicates that the product is unstable, with a low commercial value, and is therefore

unacceptable for sale. Therefore, with the aim to prevent haze formation and sediment, especially in white wine, the residual grapevine proteins are deliberately removed using fining agents (e.g. egg albumin or CNs) before the bottling step.

Proteomic characterization of foodstuffs derived from genetically modified organisms

Since the 70s, recombinant DNA technology has established itself as the leading technique in modern biotechnology. Selected individual exogenous genes, named transgenes, can be transferred by genetic engineering techniques from an organism into another, which usually is a nonrelated species, so that the derived organisms (i.e. transgenic organisms) will exhibit a new property and transmit that property to its offspring. More generally, the organisms in which the genetic material has been altered in a way that does not occur naturally are termed genetically modified organisms (GMOs). In this respect, therefore, the term GMO does not always imply, but can include, targeted insertions of genes from one species into another. DNA recombinant technology differs from traditional breeding program, where undefined genes are routinely transferred among breeding lines, species and even genera.^[105] The first commercial trade of GMO in a foodstuff began in 1994, with a delayed ripening tomato. Nowadays, more than 100 GMOs have been approved by regulatory agencies. Most of the genetic modifications have been performed to gain tolerance to total herbicides or resistance to insect in order to increase crop yields and reduce chemical use and soil impacts.^[106] However, introduction of exogenous DNA sequences into the plant genome may result, in addition to the intended effects, in unexpected events such as possible inactivation of endogenous genes or activation of silent genes. In addition, a possible modification of metabolic pathways in genetically modified varieties could lead to alterations in concentrations of nutrients or, in theory, even to the production of new protein allergens or toxins, which could compromise the product safety. Several efforts have been made to establish universal guidelines for the safety assessment of derived GMO food. The assumption that traditional crops consumed for a decade can be considered safe and therefore used as control for the safety assessment of genetically modified varieties is the basis of the concept of 'substantial equivalence', formulated as a comparative tool to identify similarities and differences between the genetically modified food crop and its isogenic counterpart.^[107]

With the aim to identify unintended alterations in the composition of genetically modified food crops, two different analytical approaches can be used: targeted analysis and profiling approaches. Targeted analysis is focused on the detection of primary or intended effect of the genetic modification. Normally, a targeted approach, such as measurements of single known micronutrients or macronutrients, toxins and anti-nutrients, is used to characterize spotting alterations in the composition of a GMO compared with the parent organism. The targeted approach, in the context of substantial equivalence, should cover a number of key nutrients such as proteins, carbohydrates, fats, vitamins and other nutritional/anti-nutritional compounds which, intentionally or unintentionally modified, might affect the safety of the GMO. However, this approach, because of the limitations raised from a restricted selection of compounds, cannot be used for the detection of unknown toxins or anti-nutrients. Few research papers are based on the targeted analysis of transgenic protein in genetically modified crops, mainly due to the low

expression level of the recombinant proteins in addition to the fact that the new protein is often not only not equally distributed in the plant tissues but also difficult to study in consequence of the wide dynamic concentration range of proteins in biological tissues. An example of proteomic targeted analysis was based on gel-filtration chromatographic, anion-exchange chromatography and SDS-PAGE prior mass spectrometric analyses of the tryptic peptides by both MALDI MS and nESI-Q/TOF MS/MS. The application of these approaches allowed the identification and characterization of the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene from *Agrobacterium tumefaciens* CP4 in Roundup ReadyTM soya and maize (a gene which confers resistance to the herbicide glyphosate). The current EU threshold level for labeling containing genetically modified products is 0.9%. Thus, the analytical procedures developed in this study were successful tested for their suitability to detect threshold amounts.^[108]

In order to increase the chances of detecting unintended effects, the European Food Safety Authority recommended the use of profiling investigations by nontargeted approaches such as genomics, transcriptomics, metabolomics or proteomics.^[109] Several studies, based on electrophoresis separation techniques followed by a bottom-up approach, have sought to discriminate if DNA alteration may result in modification of the related character or of the whole proteome. Although shotgun proteomics is a suitable strategy for proteome profile, it has barely been used in GMO analysis. The substantial equivalence of genetically modified crops, applied, as an example, to tomato, maize and potato, showed that environment affects protein expression (in terms of total number of protein spots) more than gene insertion and that the differential protein expression is often variety specific, supporting the idea that the genetic modification does not influence global protein expression.^[110] In durum wheat proteome analysis of kernels of untransformed and genetically modified durum wheat lines, in which genes of the starch branching enzymes were silenced by RNA interference, highlighted subtle differences, most of which were considered as 'predictable unintended effects' because of the primary effect of the transgene on the starch biosynthetic pathway.^[111]

Differently, when wheat seeds from a transformed bread wheat line over-expressing a transgenic LMW-GS were studied both at protein and transcript levels, proteomic analysis revealed, as a response to the over-accumulation of the LMW-GS, a decrease in the amount of proteins belonging to storage proteins. This is a possible result of a mechanism of compensation, which seems to be proportional to the expression level of the introduced glutenin transgene.^[112]

The rapid development and commercialization of a variety of food crops with transgenes have increased the interest on the allergy issue.^[113] Obviously, a protein, well known to be an allergen in one species, remains an allergen when transferred by genetic transformation to a second species and therefore is easy to be detected. The question is whether the protein of interest becomes an allergen as a result of the modification and selection process or whether its presence unintentionally renders a food product more allergenic. Therefore, proteomic approaches are essential to determine the abundance of potential protein allergens in GMOs, examining their PTMs (e.g. glycosylations), heat stability and the presence of disulfide bonds, all features related with known food allergens.

Moreover, biotechnology offers the prospect of utilizing recombinant techniques to eliminate undesirable allergen proteins

to enhance food safety and possibly make use of derived products available for many sensitive individuals. In fact, thanks to gene-silencing technology, it was possible to remove a major or immunodominant soybean (*Glycine max*) seed allergen, the Gly m Bd 30 K protein, providing the evidence for substantial equivalence of composition of transgenic and nontransgenic seeds, after elimination of the dominant allergenic protein.^[114]

Concluding remarks

The present tutorial aims at introducing the readers to the characterization by MS-based approaches of food-derived proteins. The increased performance and versatility of the MS instrumentation, together with the concurrent optimization of the separation techniques, the improvement of gene sequencing methods and the availability of bioinformatic tools, have provided an efficient platform in many proteomic research areas, including food. Even if a comprehensive proteomic exploration does not consist in the application of a single technique for all purposes, MS represents certainly the core technology, providing high-throughput protein identification, protein expression and protein-protein interaction. Nowadays, mass spectrometers are capable of measuring the molecular mass of proteins and peptides with very high accuracy and determining additional structural features such as the primary structure or PTMs.

Food proteomics, in comparison with medical and clinical research, is still at the beginning stage. However, MS-based approaches have clearly demonstrated their enormous potentiality in the characterization of the complex protein mixture contained in food, playing a decisive role to elucidate many relationships between the nature and structure of food proteins and their influence on the health of end-consumers. Today, the various proteomic technologies available, together with in-depth investigation of food proteins, allow to address many questions, encompassing the composition (food quality), the functionality (presence of bioactive peptides), the origin (food traceability) or the safety of foodstuffs (e.g. absence of allergens, pathogens or other contaminants). Moreover, the rapid development and commercialization of a variety of genetically modified food crops have raised the interest on the investigation of alimentary products coming from GMOs. The main objective of proteomic investigation of genetically modified plants is the characterization of possible unintended effects such as the expression of known or new protein allergens, which might be dangerous for the human health.

It should not be neglected that, given the high variety of the food-derived proteins and considering their differences in chemical and physical properties, a single proteomic strategy for all purposes does not exist. Rather, proteomic approaches need to be adapted each time, including the development of new strategies in order to always obtain the best results. It is important to highlight that probably the best is yet to come. Improvements could be achieved by trans-disciplinary collaborations involving researchers tackling food safety issues, health centers, proteomics, metabolomics and micro-components studies. However, researchers should also be supported by adequate economical policies aimed to create health promotion and disease prevention strategies.

Acknowledgement

This work was supported by a grant from MURST (PRIN 2010/11, project number 2010CSJX4F_001).

References

- [1] J.R. Yates, C.I. Ruse, A. Nakorchevsky. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed.* **2009**, *11*, 49.
- [2] M. Herrero, C. Simó, V. García-Cañas, E. Ibáñez, A. Cifuentes. Foodomics: MS-based strategies in modern food science and nutrition. *Mass Spectrom. Rev.* **2012**, *31*, 49.
- [3] L. Di Donna, F. Mazzotti, A. Napoli, A. Sajjad, R. Salerno, G. Sindona. Secondary metabolism of olive secoiridoids. New micro-components detected in drupes by electrospray ionization and high-resolution tandem mass spectrometry. *Rapid Comm. Mass Spectrom.* **2007**, *21*, 273.
- [4] G. Mamone, G. Picariello, S. Caira, F. Addeo, P. Ferranti. Analysis of food proteins and peptides by mass spectrometry-based techniques. *J. Chromatogr. A* **2009**, *1216*, 7130.
- [5] J.M. Gallardo, I. Ortea, M. Carrera. Proteomics and its applications for food authentication and food-technology research. *Trends in Anal. Chem.* **2013**, *52*, 135.
- [6] A. D'Alessandro, L. Zolla. We are what we eat: food safety and proteomics. *J. Proteome Res.* **2012**, *11*, 26.
- [7] A. Jimmy Ytterberg, J. Dunsmore, S. H. Lomeli, M. Thevis, Y. Xie, R. R. Ogorzalek Loo, J. A. Loo. In *Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities, and Biological Applications*: Second Edition. Part V; chapter 18, In *The Role of Mass Spectrometry for Peptide, Protein, and Proteome Characterization*, R. B. Cole (Ed). J. Wiley and Sons: Hoboken, New Jersey, **2012**, 685.
- [8] S.S. Thakur, T. Geiger, B. Chatterjee, P. Bandilla, F. Frohlich, J. Cox, M. Mann. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *M. Mol. Cell Proteomics* **2011**, *10*, M110 003699.
- [9] A.B. Kanu, P. Dwivedi, M. Tam, L. Matz, H.H. Hill Jr. Ion mobility-mass spectrometry. *J. Mass Spectrom.* **2008**, *43*, 1.
- [10] J.A. Loo, J.P. Quinn, S.I. Ryu, K.D. Henry, M.W. Senko, F.W. McLafferty. High-resolution tandem mass spectrometry of large biomolecules. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 286.
- [11] Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, G.R. Cooks. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005**, *40*, 430.
- [12] L. Quan, M. Liu. CID, ETD and HCD fragmentation to study protein post-translational modifications. *Mod. Chem. Appl.* **2013**, *1*, 1, and references therein.
- [13] D.M. Good, M. Wirtala, G.C. McAlister, J.J. Coon. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteom.* **2007**, *6*, 1942.
- [14] H. Chi, R.X. Sun, B. Yang, C.Q. Song, L.H. Wang, et al. pNovo: de novo peptide sequencing and identification using HCD spectra. *J. Proteome Res.* **2010**, *9*, 2713.
- [15] C.K. Frese, A.F. Altelaar, M.L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A.J. Heck, S. Mohammed. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *Proteome Res.* **2011**, *10*, 2377.
- [16] N.L. Kelleher, H.Y. Lin, G.A. Valaskovic, D.J. Aaserud, E.K. Fridriksson, F.W. McLafferty. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Amer. Chem. Soc.* **1999**, *121*, 806.
- [17] M.L. Fournier, J.M. Gilmore, S.A. Martin-Brown, M.P. Washburn. Multidimensional separations-based shotgun proteomics. *Chemical Review* **2007**, *107*, 3654.
- [18] C.M. Delahunty, J.R. Yates III. MudPIT: multidimensional protein identification technology. *Biotechniques* **2007**, *43*, 563.
- [19] T. Rabilloud. Membrane proteins and proteomics: love is possible, but so difficult. *Electrophoresis* **2009**, *30*, S174.
- [20] P.G. Righetti, E. Boschetti. The ProteoMiner and the FortyNiners: searching for gold nuggets in the proteomic arena. *Mass Spectrom. Rev.* **2008**, *27*, 596.
- [21] L. Switzar, M. Giera, W.M. Niessen. Protein digestion: an overview of the available techniques and recent developments. *J. Proteome Res.* **2013**, *12*, 1067.
- [22] F.W. McLafferty, K. Breuker, M. Jin, X. Han, G. Infusini, H. Jiang, X. Kong, T.P. Begley. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* **2007**, *274*, 6256.
- [23] J. Wiesner, T. Premisler, A. Sickmann. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **2008**, *8*, 4466.
- [24] Y. Ge, B.G. Lawhorn, M. ElNaggar, E. Strauss, J.H. Park, T.P. Begley, F.W. McLafferty. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **2002**, *124*, 672.
- [25] X. Han, M. Jin, K. Breuker, F.W. McLafferty. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* **2006**, *314*, 109.
- [26] L. Martens. Bioinformatics challenges in mass spectrometry-driven proteomics. *Methods Mol. Biol.* **2011**, *753*, 359.
- [27] I. Eidhammer, K. Flikka, L. Martens, S.O. Mikalsen. Protein identification and characterization by MS. In *Computational Methods for Mass Spectrometry Proteomics*. (Eds. Wiley) **2007**, and reference therein.
- [28] P. Waridel, A. Frank, H. Thomas, V. Surendranath, S. Sunyaev, P. Pevzner, A. Shevchenko. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. *Proteomics* **2007**, *7*, 2318.
- [29] A. Panchaud, M. Affolter, P. Moreillon, M. Kussmann. Experimental and computational approaches to quantitative proteomics: status quo and outlook. *J. Proteomics* **2008**, *71*, 19.
- [30] M.H. Elliott, D.S. Smith, C.E. Parker, C. Borchers. Current trends in quantitative proteomics. *J. Mass Spectrom.* **2009**, *44*, 1637.
- [31] V. Cunsolo, V. Muccilli, R. Saletti, S. Foti. Mass spectrometry in the proteome analysis of mature cereal kernels. *Mass Spectrom. Rev.* **2012**, *31*, 448.
- [32] P.J. Payne, M.A. Nightingale, A.F. Krattiger, L.M. Holt. The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *J. Sci. Food Agric.* **1987**, *40*, 51.
- [33] R.G. Dworschak, W. Ens, K.G. Standing, K.R. Preston, B.A. Marchylo, M.J. Nightingale, S.G. Stevenson, D.W. Hatcher. Analysis of wheat gluten proteins by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* **1998**, *33*, 429.
- [34] V. Cunsolo, S. Foti, R. Saletti, S. Gilbert, A.S. Tatham, P.R. Shewry. Structural studies of the allelic wheat glutenin subunits 1Bx7 and 1Bx20 by matrix-assisted laser desorption/ionization mass spectrometry and high-performance liquid chromatography/electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2004**, *39*, 66.
- [35] V. Cunsolo, S. Foti, R. Saletti, S. Gilbert, A.S. Tatham, P.R. Shewry. Structural studies of glutenin subunits 1Dy10 and 1Dy12 by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Comm. Mass Spectrom.* **2003**, *17*, 442.
- [36] V. Muccilli, V. Cunsolo, R. Saletti, S. Foti, S. Masci, D. Lafiandra. Characterization of B- and C-type low molecular weight glutenin subunits by electrospray ionization mass spectrometry and matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* **2005**, *5*, 719.
- [37] W.H. Vensel, F.M. Dupont, R. Chan, W.J. Hurkman. **2007**. Mass spectrometry based identifications of LMW glutenin subunits. In *Gluten Proteins 2006*, G.L. Lookhart, P.K.W. Ng (Eds). AACCI: Minneapolis, MN, 347–351.
- [38] W.H. Vensel, F.M. Dupont, S. Sloane, S.B. Altenbach. Effect of cleavage enzyme, search algorithm and decoy database on mass spectrometric identification of wheat gluten proteins. *Phytochemistry* **2011**, *72*, 1154.
- [39] G. Mamone, S. De Caro, A. Di Luccia, F. Addeo, P. Ferranti. Proteomic-based analytical approach for the characterization of glutenin subunits in durum wheat. *J. Mass Spectrom.* **2009**, *44*, 1709.
- [40] V. Muccilli, V. Cunsolo, R. Saletti, S. Foti, B. Margiotta, F. Scossa, S. Masci, D. Lafiandra. Characterization of a specific class of typical low molecular weight glutenin subunits of durum wheat by a proteomic approach. *J. Cereal Sci.* **2010**, *51*, 134.
- [41] S. Gobaa, E. Bancel, G. Kleijer, P. Stamp, G. Branlard. Effect of the 1BL1RS translocation on the wheat endosperm, as revealed by proteomic analysis. *Proteomics* **2007**, *7*, 4349.
- [42] G. Mamone, F. Addeo, L. Chianese, A. Di Luccia, A. De Martino, A. Nappo, A. Formisano, P. De Vivo, P. Ferranti. Characterization of wheat gliadin proteins by combined two-dimensional gel electrophoresis and tandem mass spectrometry. *Proteomics* **2005**, *5*, 2859.
- [43] V. Cunsolo, V. Muccilli, R. Saletti, S. Foti. Review: applications of mass spectrometry techniques in the investigation of milk proteome. *Eur. J. Mass Spectrom.* **2011**, *17*, 305.

- [44] T. Uniacke-Lowe, T. Huppertz, P.F. Fox. Equine milk proteins: chemistry, structure and nutritional significance. *Int. Dairy J.* **2010**, 20, 609.
- [45] V. Cunsolo, E. Cairone, R. Saletti, V. Muccilli, S. Foti. Sequence and phosphorylation level determination of two donkey β -caseins by mass spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, 23, 1907.
- [46] V. Cunsolo, R. Saletti, V. Muccilli, S. Foti. Characterization of the protein profile of donkey's milk whey fraction. *J. Mass Spectrom.* **2007**, 42, 1162.
- [47] V. Cunsolo, V. Muccilli, E. Fasoli, R. Saletti, P.G. Righetti, S. Foti. Poppea's bath liquor: the secret proteome of she-donkey's milk. *J. Proteomics* **2011**, 74, 2083.
- [48] V. Cunsolo, F. Galliano, V. Muccilli, R. Saletti, D. Marletta, S. Bordonaro, S. Foti. Detection and characterization by high-performance liquid chromatography and mass spectrometry of a goat β -casein associated with a CSN2 null allele. *Rapid Comm. Mass Spectrom.* **2005**, 19, 2943.
- [49] V. Cunsolo, E. Cairone, D. Fontanini, A. Criscione, V. Muccilli, R. Saletti, S. Foti. Sequence determination of α_{s1} -casein isoforms from donkey by mass spectrometric method. *J. Mass Spectrom.* **2009**, 44, 1742.
- [50] R. Saletti, V. Muccilli, V. Cunsolo, D. Fontanini, A. Capocchi, S. Foti. MS-based characterization of α_{s2} -casein isoforms in donkey's milk. *J. Mass Spectrom.* **2012**, 47, 1150.
- [51] G. Mamone, S. Caira, G. Garro, A. Nicolai, P. Ferranti, G. Picariello, A. Malorni, L. Chianese, F. Addeo. Casein phosphoproteome: identification of phosphoproteins by combined mass spectrometry and two-dimensional gel electrophoresis. *Electrophoresis* **2003**, 24, 2824.
- [52] A. D'Alessandro, S. Rinalducci, C. Marrocco, V. Zolla, F. Napolitano, L. Zolla. Love me tender: an Omics window on the bovine meat tenderness network. *J. Proteomics* **2012**, 75, 4360.
- [53] V. Muccilli, C. Licciardello, D. Fontanini, M.P. Russo, V. Cunsolo, R. Saletti, G. Reforgiato Recupero, S. Foti. Proteome analysis of *Citrus sinensis* L. (Osbeck) flesh at ripening time. *J. Proteomics* **2009**, 73, 134.
- [54] V. Muccilli, C. Licciardello, D. Fontanini, V. Cunsolo, A. Capocchi, R. Saletti, B. Torrisi, S. Foti. Root protein profiles of two citrus rootstocks grown under iron sufficiency/deficiency conditions. *Eur. J. Mass Spectrom.* **2013**, 19, 305.
- [55] M. Pischetsrieder, R. Baeuerlein. Proteome research in food science. *Chem. Soc. Rev.* **2009**, 38, 2600.
- [56] Q. Zhang, J.M. Ames, R.D. Smith, J.W. Baynes, T.O. Metz. A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease. *J. Proteome Res.* **2009**, 8, 754.
- [57] I. Losito, T. Carbonara, L. Monaci, F. Palmisano. Evaluation of the thermal history of bovine milk from the lactosylation of whey proteins: an investigation by liquid chromatography–electrospray ionization mass spectrometry. *Anal. Bioanal. Chem.* **2007**, 389, 2065.
- [58] S. Arena, G. Renzone, G. Novi, A. Paffetti, G. Bernardini, A. Santucci, A. Scaloni. Modern proteomic methodologies for the characterization of lactosylation protein targets in milk. *Proteomics* **2010**, 10, 3414.
- [59] M. Friedman. Chemistry, biochemistry, nutrition and microbiology of lysinoalanine, lanthionine, and histidinoalanine in food and other products. *J. Agric. Food Chem.* **1999**, 47, 1295.
- [60] M.G. Calabrese, G. Mamone, S. Caira, P. Ferranti, F. Addeo. Quantitation of lysinoalanine in dairy products by liquid chromatography–mass spectrometry with selective ion monitoring. *Food Chem.* **2009**, 116, 799.
- [61] D.S. Mottram, B.L. Wedzicha, A.T. Dodson. Acrylamide is formed in the Maillard reaction. *Nature* **2002**, 419, 448.
- [62] A. Basile, P. Ferranti, R. Moccaldi, G. Spagnoli, N. Sannolo. Proteomic approach for the analysis of acrylamide–hemoglobin adducts. Perspectives for biological monitoring. *J. Chromatography A* **2008**, 1215, 74.
- [63] F. Fenaile, V. Parisod, J.C. Tabet, P.A. Guy. Carbonylation of milk powder proteins as a consequence of processing conditions. *Proteomics* **2005**, 5, 3097.
- [64] A. Scaloni, V. Perillo, P. Franco, F. Fedele, R. Froio, L. Ferrara, P. Bergamo. Characterization of heat-induced lactosylation products in caseins by immunoenzymatic and mass spectrometric methodologies. *Biochim. Biophys. Acta* **2002**, 1598, 30.
- [65] A. Promeprat, T. Sayd, E. Laville, C. Chambon, B. Lebre, P. Gatellier. Early post mortem sarcoplasmic proteome of porcine muscle related to protein oxidation. *Food Chem.* **2011**, 127, 1097.
- [66] M. Pazos, A. Pereira da Rocha, P. Roepstorff, A. Rogowska-Wrzesinska. Fish proteins as targets of ferrous-catalyzed oxidation: identification of protein carbonyls by fluorescent labeling on two-dimensional gels and MALDI-TOF/TOF mass spectrometry. *J. Agric. Food Chem.* **2011**, 59, 7962.
- [67] P.C. Burcham, Y.T. Kuhan. Diminished susceptibility to proteolysis after protein modification by the lipid peroxidation product malondialdehyde: inhibitory role for cross-linked and non cross-linked adducted proteins. *Arch. Biochem. Biophys.* **1997**, 340, 331.
- [68] European Commission. OJ L 109, 6.5.2000. European Commission; Bruxelles, Belgium. **2009**. Directive 2000/13/ec of the European parliament and of the council of 20 march 2000 on the approximation of the laws of the member states relating to the labelling, presentation and advertising of foodstuffs; p. 29.
- [69] A.J. van Hengel. Food allergen detection methods and the challenge to protect food-allergic consumers. *Anal. Bioanal. Chem.* **2007**, 389, 111.
- [70] L. Monaci, V. Tregioat, A.J. van Hengel, E. Anklam. Milk allergens, their characteristics and their detection in food: a review. *Eur. Food Res. Technol.* **2006**, 223, 149.
- [71] M. Natale, C. Bisson, G. Monti, A. Peltran, L.P. Garoffo, S. Valentini, C. Fabris, E. Bertino, A. Coscia, A. Conti. Cow's milk allergens identification by two-dimensional immunoblotting and mass spectrometry. *Mol. Nutr. Food Res.* **2004**, 48, 363.
- [72] S. Sabbadin, R. Seraglia, G. Allegri, A. Bertazzo, P. Traldi. Matrix-assisted laser desorption/ionization mass spectrometry in evaluation of protein profiles of infant formulae. *Rapid Commun. Mass Spectrom.* **1999**, 13, 1438.
- [73] G. Monti, E. Bertino, M.C. Muratore, A. Coscia, F. Cresi, L. Silvestro, C. Fabris, D. Fortunato, M.G. Giuffrida, A. Conti. Efficacy of donkey's milk in treating highly problematic cow's milk allergic children: an in vivo and in vitro study. *Ped. Allergy Immunol.* **2007**, 18, 258.
- [74] A. Criscione, V. Cunsolo, S. Bordonaro, A.M. Guastella, R. Saletti, A. Zuccaro, G. D'Urso, D. Marletta. Donkey's milk protein fraction investigated by electrophoretic methods and mass spectrometric analysis. *Int. Dairy J.* **2009**, 19, 190.
- [75] A. D'Amato, A.V. Kravchuk, A. Bachi, P.G. Righetti. Noah's nectar: the proteome content of a glass of red wine. *J. Proteomics* **2010**, 73, 2370.
- [76] M. Careri, A. Costa, L. Evi, J.B. Lagos, A. Mangia, M. Terenghi, A. Cereti, L. Perono Garoffo. Use of specific peptide biomarkers for quantitative confirmation of hidden allergenic peanut proteins Ara h 2 and Ara h 3/4 for food control by liquid chromatography–tandem mass spectrometry. *Anal. Bioanal. Chem.* **2007**, 389, 1901.
- [77] H.B. Krishnan, W.S. Kim, S. Jang, M.S. Kerley. All three subunits of soybean beta-conglycinin are potential food allergens. *J. Agric. Food Chem.* **2009**, 57, 938.
- [78] S.A. Hasan, R.D. Wells, C.M. Davis. Egg hypersensitivity in review. *Allergy Asthma Proc.* **2013**, 34, 26.
- [79] A.K. Malik, C. Blasco, Y. Pico. Liquid chromatography–mass spectrometry in food safety. *J. Chromatogr. A* **2010**, 1217, 4018.
- [80] Commission of the European Communities, White Paper on Food Safety, COM. **1999**. 719 final, Brussels, 12 January 2000 (available at: http://ec.europa.eu/food/food/intro/white_paper_en.htm).
- [81] H.J.P. Marvin, G.A. Kleter, L.J. Frewer, S. Cope, M.T.A. Wentholt, G. Rowe. A working procedure for identifying emerging food safety issues at an early stage: Implications for European and international risk management practices. *Food Control* **2009**, 20, 345.
- [82] E. Venir, M. Del Torre, V. Cunsolo, R. Saletti, R. Musetti, M.L. Stecchini. Involvement of alanine racemase in germination of *Bacillus cereus* spores lacking an intact exosporium. *Arch. Microbiol.* **2014**, 196, 79.
- [83] V. Tsilia, B. Devreese, I. de Baenst, B. Mesuere, A. Rajkovic, M. Uyttendaele, T. Van de Wiele, M. Heyndrickx. Application of MALDI-TOF mass spectrometry for the detection of enterotoxins produced by pathogenic strains of the *Bacillus cereus* group. *Anal. Bioanal. Chem.* **2012**, 404, 1691.
- [84] B.L.M. van Baar. Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiol. Rev.* **2000**, 24, 193.
- [85] I.C. Fernandez-No, K. Bohme, J.M. Gallardo, J. Barros-Velazquez, B. Canas, P. Calo-Mata. Differential characterization of biogenic

- amine-producing bacteria involved in food poisoning using MALDI-TOF mass fingerprinting. *Electrophoresis* **2010**, *31*, 1116.
- [86] K. Bohme, I.C. Fernandez-No, J. Barros-Velazquez, J.M. Gallardo, B. Canas, P. Calo-Mata. SpectraBank: an open access tool for rapid microbial identification by MALDI-TOF MS fingerprinting. *Electrophoresis* **2012**, *33*, 2138.
- [87] J.L. Boehmer, D.D. Bannerman, K. Shefcheck, J.L. Ward. Proteomic analysis of differentially expressed proteins in bovine milk during experimentally induced *Escherichia coli* mastitis. *J. Dairy Sci.* **2008**, *91*, 4206.
- [88] N.T.L. Doan, K. Van Hoorde, M. Cnockaert, E. De Brandt, M. Aerts, B.L. Thanh, P. Vandamme. Validation of MALDI-TOF MS for rapid classification and identification of lactic acid bacteria, with a focus on isolates from traditional fermented foods in Northern Vietnam. *Lett. Appl. Microbiol.* **2012**, *55*, 265.
- [89] L.F. Marvin-Guy, S. Parche, S. Wagniere, J. Moulin, R. Zink, M. Kussmann, et al. Rapid identification of stress-related fingerprint from whole bacterial cells of *Bifidobacterium lactis* using matrix assisted laser desorption/ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1222.
- [90] B. Sanchez, M.C. Champomier-Verges, P. Anglade, F. Baraige, C.G. de Los Reyes-Gavilan, A. Margolles, et al. Proteomic analysis of global changes in protein expression during bile salt exposure of *Bifidobacterium longum* NCIMB 8809. *J. Bacteriol.* **2005**, *187*, 5799.
- [91] H. Korhonen, A. Pihlanto. Food-derived bioactive peptides – opportunities for designing future foods. *Curr. Pharm. Des.* **2003**, *9*, 1297.
- [92] D. Schillaci, V. Arizza, N. Parrinello, V. Di Stefano, S. Fanara, V. Muccilli, V. Cunsolo, J.A. Haagenen, S. Molin. Antimicrobial and antistaphylococcal biofilm activity from the sea urchin *Paracentrotus lividus*. *J. Appl. Microbiol.* **2010**, *108*, 117.
- [93] D. Schillaci, M.G. Cusimano, V. Cunsolo, R. Saletti, D. Russo, M. Vazzana, M. Vitale, V. Arizza. Immune mediators of sea-cucumber *Holothuria tubulosa* (Echinodermata) as source of novel antimicrobial and anti-staphylococcal biofilm agents. *AMB Express*. **2013**, *3*, 35.
- [94] J. Carrasco-Castilla, A.J. Hernández-Álvarez, C. Jiménez-Martínez, G. F. Gutiérrez-López, G. Dávila-Ortiz. Use of proteomics and peptidomics methods in food bioactive peptide science and engineering. *Food Eng Rev.* **2012**, *4*, 224.
- [95] L. Saavedra, E.M. Hebert, C. Minahk, P. Ferranti. An overview of “omic” analytical methods applied in bioactive peptide studies. *Food Res. Int.* **2013**, *54*, 925.
- [96] A. Panchaud, M. Affolter, M. Kussmann. Mass spectrometry for nutritional peptidomics: how to analyze food bioactives and their health effects. *J. Proteomics* **2012**, *75*, 3546.
- [97] B. Paizs, S. Suhai. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005**, *24*, 508.
- [98] G. Picariello, G. Mamone, C. Nitride, F. Addeo, P. Ferranti. Protein digestomics: integrated platforms to study food-protein digestion and derived functional and active peptides. *Trends Anal. Chem.* **2013**, *52*, 120.
- [99] R. Cozzolino, S. Passalacqua, S. Salemi and D. Garozzo. Identification of adulteration in water buffalo mozzarella and in ewe cheese by using whey proteins as biomarkers and matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* **2002**, *37*, 985.
- [100] V. Cunsolo, V. Muccilli, R. Saletti, S. Foti. MALDI-TOF mass spectrometry for the monitoring of she-donkey's milk contamination or adulteration. *J. Mass Spectrom.* **2013**, *8*, 148.
- [101] M. Cuollo, S. Caira, O. Fierro, G. Pinto, G. Picariello and F. Addeo. Toward milk speciation through the monitoring of casein proteotypic peptides. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 1687.
- [102] V. Muccilli, M. Lo Bianco, V. Cunsolo, R. Saletti, G. Gallo, S. Foti. High molecular weight glutenin subunits in some durum wheat cultivars investigated by means of mass spectrometric techniques. *J. Agric. Food Chem.* **2011**, *59*, 12226.
- [103] E. Fasoli, A. D'Amato, A.V. Kravchuk, A. Citterio, P.G. Righetti. In-depth proteomic analysis of non-alcoholic beverages with peptide ligand libraries. I: almond milk and orgeat syrup. *J. Proteomics* **2011**, *74*, 1080.
- [104] V. Saez, E. Fasoli, A. D'Amato, E. Simó-Alfonso, P.G. Righetti. Artichoke and Cynar liqueur: two (not quite) entangled proteomes. *Biochim. Biophys. Acta* **2013**, *1834*, 119.
- [105] V. García-Cañas, C. Simó, C. León, E. Ibáñez, A. Cifuentes. MS-based analytical methodologies to characterize genetically modified crops. *Mass Spectrom. Rev.* **2011**, *30*, 396.
- [106] C. James. Global status of commercialized biotech/GM crops: 2010, ISAAA brief No. 42; ISAAA, Ithaca, NY, **2010**.
- [107] Organization of Economic Cooperation & Development (OECD) in Safety evaluation of foods derived by modern biotechnology: concepts and principles. Organisation for Economic Co-operation and Development, Paris: **1993**. <http://www.oecd.org/pdf/M00034000/M00034525.pdf>
- [108] M. Fernandez Ocaña, P. Fraser, R.K.P. Patel, J.M. Halket, P.M. Bramley. Mass spectrometric detection of CP4 EPSPS in genetically modified soya and maize. *Rapid Commun. Mass.* **2007**, *21*, 319.
- [109] European Food Safety Agency. Guidance document of the scientific panel on genetically modified organisms for the risk assessment of genetically modified plants and derived food and feed. EFSA Communications Department, Parma, Italy. **2006**.
- [110] L. Zolla, S. Rinalducci, P. Antonioli, P.G. Righetti. Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. *J. Proteome Res.* **2008**, *7*, 1850.
- [111] F. Sestili, F. Paoletti, E. Botticella, S. Masci, R. Saletti, V. Muccilli, D. Lafiandra. Comparative proteomic analysis of kernel proteins of two high amylose transgenic durum wheat lines obtained by biolistic and Agrobacterium-mediated transformations. *J. Cereal Sci.* **2013**, *58*, 15.
- [112] F. Scossa, D. Laudencia-Chinguanco, O.D. Anderson, W.H. Vensel, D. Lafiandra, R. D'Ovidio, S. Masci. Comparative proteomic and transcriptional profiling of a bread wheat cultivar and its derived transgenic line overexpressing a low molecular weight glutenin subunit gene in the endosperm. *Proteomics* **2008**, *8*, 2948.
- [113] B.B. Buchanan. Genetic engineering and the allergy issue. *Plant Physiol.* **2001**, *126*, 5.
- [114] E.M. Herman, R.M. Helm, R. Jung, A.J. Kinney. Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol.* **2003**, *132*, 36.