Università della Calabria

Dipartimento di Chimica

Tesi di Dottorato di Ricerca in

METODOLOGIE PER LO SVILUPPO DI MOLECOLE D'INTERESSE FARMACOLOGICO
***XX Ciclo* (CHIM/01)**

PROTEIN CHARACTERIZATION FROM NATURAL MATRICES BY MALDI TOF-TOF.

Supervisori.
Ch.ma Prof.ssa Anna NAPOLI

Ch.mo Prof. Giovanni SINDONA

Candidata.
Dott.ssa Donatella AIELLO

Coordinatore.
Ch.mo Prof. Bartolo GABRIELE

A.A. 2005 – 2007

For my thoughts are not your thoughts,
neither are your ways my ways, said the Lord.
(Isaiah 55,8)

To every thing there is a season, and a time to every purpose under the heaven:
A time to be born, and a time to die;
a time to plant, and a time to pluck up that which is planted;
A time to kill, and a time to heal; a time to break down, and a time to build up;
A time to weep, and a time to laugh; a time to mourn, and a time to dance;
A time to cast away stones, and a time to gather stones together
a time to embrace, and a time to refrain from embracing;
A time to get, and a time to lose; a time to keep, and a time to cast away;
A time to rend, and a time to sew; a time to keep silence, and a time to speak;
A time to love, and a time to hate; a time of war, and a time of peace
(Qohelet 3, 1-8)

To my mother,
I know that you are looking at me.

To my family.
To my love.

PROTEIN CHARACTERIZATION FROM NATURAL MATRICES BY MALDI TOF-TOF.

PhD Student

Donatella AIELLO

Contents

Trends in Proteomics.

Proteomics is “the qualitative and quantitative” comparison of proteomes under different conditions to understand cellular mechanism underlying biological processes, with the important objective of achieving an overview of the proteins expressed at a given point in a time in a given tissue and to identify the connection to the biochemical status of that tissue.[1, 2] Protein expression gives the possibility to characterize bioactive markers, to study proteins-proteins interactions or between proteins and other molecular species, to map the presence of bioactive compounds in functional food, in order to characterize proteins with particular properties, for example allergens, structural proteins or carrier in metabolic pattern.[3, 4]

The complexity of any proteome, time- and cell-specific protein complement of the genome, makes all proteome analysis technically challenging. Determination of proteins in either small or large cells requires methods for separation of protein mixtures into their individual components. Three developments changed the biological landscape and formed the foundation of the new biology.[5, 6] The first was the growth of gene, expressed sequence tag (EST), and protein-sequence databases. These resources became ever more useful as partial catalogues of expressed genes in many organisms and culminated in the complete sequence of the human genome. Sequences of plant genomes and those of other widely studied animals also are recently completed or are approaching completion. These genome-sequence databases are the catalogues from which much of our understanding of living systems eventually will be extracted.

The second key development was the introduction of user-friendly, browser-based bioinformatics tools to extract information from these databases. Such database search tools

are integrated with other tools and databases to predict the functions, the locations, the properties of the protein products based on the occurrence of specific functional domains or motifs.

The third key development was the improvement of the oligonucleotide micro-array. The array contains a series of gene-specific oligonucleotides or cDNA sequences on a slide or a chip. By applying a mixture of fluorescently labeled DNAs from a sample of interest to the array, one can probe the expression of thousands of genes at once. One array can replace thousands of Northern-blot analyses and can be done in the time it would take to do one Northern.

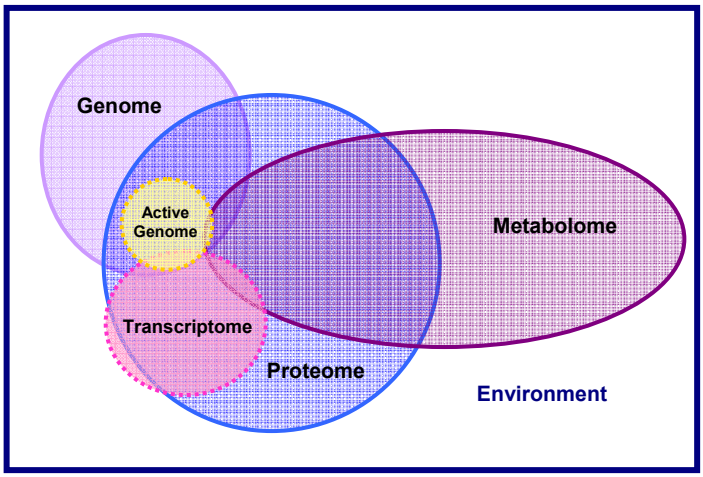


Figure 1. Interrelationships between molecular classes in cells.

Nowadays it is possible "seeing" the whole system, but the information contained in these thousands of data points is beyond our ability to interpret intuitively. New clustering algorithms, self-organizing maps, and similar tools represent the latest approaches to rendering the data in ways that biologists and chemists can comprehend them. All improvements give us the possibility to see the complete system, to think big, to imagine a cell with thousands or tens of thousands of genes, that may be expressed in varying

combinations translating thousands or tens of thousands of different proteins. Each protein, whether a trans-membrane receptor, a transcription factor, a protein kinase, or a chaperone, expresses a function that assumes significance only in the context of all the other functions and activities also being expressed in the same cell.

Currently, proteomic approaches based on the analysis of protein pattern are commonly used, that may provide a more effective evaluated profiling protein for diagnostic purposes, such as two-dimensional polyacrylamide gel electrophoresis (2-DE)[7], surface-enhanced laser desorption ionization (SELDI)[8], matrix-assisted laser desorption ionization (MALDI)[9], liquid chromatography (LC)[10], capillary electrophoresis (CE)[11], followed by computational image analysis and protein identification using mass spectrometry[12]. The use of combined proteomic techniques for protein identification is a powerful approach that can give a better understanding about the mechanism of disease in which proteins play major role[13, 14].

The proteomic approach using different analytical techniques has been successfully used for protein expression analysis, screening, identification and characterization of protein, but some techniques have certain limitations and need to consider the essential factors for solving these problems. Composition of the proteome and analytical methods are the main limitations in the proteomic analysis. In fact, each sample contains a huge diversity of proteins, which show different chemical properties and characteristics.

The sample preparation is the most important factor in the first step of proteomic analysis; ineffective steps can lead to loss of valuable samples, time and cost. However, the sample preparation methods can be affected by some essential factors, such as sample extraction, protein solubilization, protease inhibitors, protein concentration, and non-protein contamination. On the other hand, the limitations of the analytical methods are the detection and the quantification of the proteins, usually it is due to the difficulty in detection of low abundant proteins in biological materials. In addition, some techniques of proteomic analysis show problems about reproducibility, sensitivity and accuracy, for example 2DE, problems that are partially overtaken by the use of mass spectrometry.

For a better understanding about proteomics, it is important explaining the differences between proteomics and protein biochemistry. Both protein biochemistry and proteomics involve protein identification, but while the first involves a complete sequence analysis, structure determination, and modeling studies to explore how protein structure governs function, the second one is the study of multi-protein systems (Table 1).

Protein Chemistry	**Proteomics**
• Individual proteins	• Complex mixtures
• Complete sequence analysis	• Partial sequence analysis
• Emphasis on structure and function	• Emphasis on identification by database matching
• Structural biology	• Systems biology

Table 1. Differences Between Protein Chemistry and Proteomics.

In Proteomics the focus is on the interplay of multiple, distinct proteins in their roles as part of a larger system or network, the analyses are directed at complex mixtures and identification is partial by sequence analysis with the aid of database matching tools. In other words, the point of proteomics is to characterize the behaviour of the system rather than the behaviour of any single component[5, 15, 16].

Proteomics encompasses four principal applications [6].

1) *mining*, 2) *protein-expression profiling*, 3) *protein-network mapping*, and 4) *mapping of protein modifications* (Figure. 2).[17]

1) *Mining* is simply the exercise of identifying all (or as many as possible) of the proteins in a sample. The point of mining is to record the proteome directly, rather than to infer the composition of the proteome from expression data for genes (e.g., by microarrays). Mining is the ultimate brute-force exercise in proteomics: one simply resolves proteins to the greatest extent possible and then uses MS and associated database and software tools to identify what is found.

2) *Protein-expression profiling* is the identification of proteins in particular sample as a function of a particular state of the organism or cell (e.g., differentiation, developmental

state, or disease state) or as a function of exposure to a drug, chemical, or physical stimulus. Expression profiling is actually a specialized form of mining. It is most commonly practiced as a differential analysis, in which two states of a particular system are compared. For example, normal and diseased cells or tissues can be compared to determine which proteins are expressed differently in one state compared to the other.

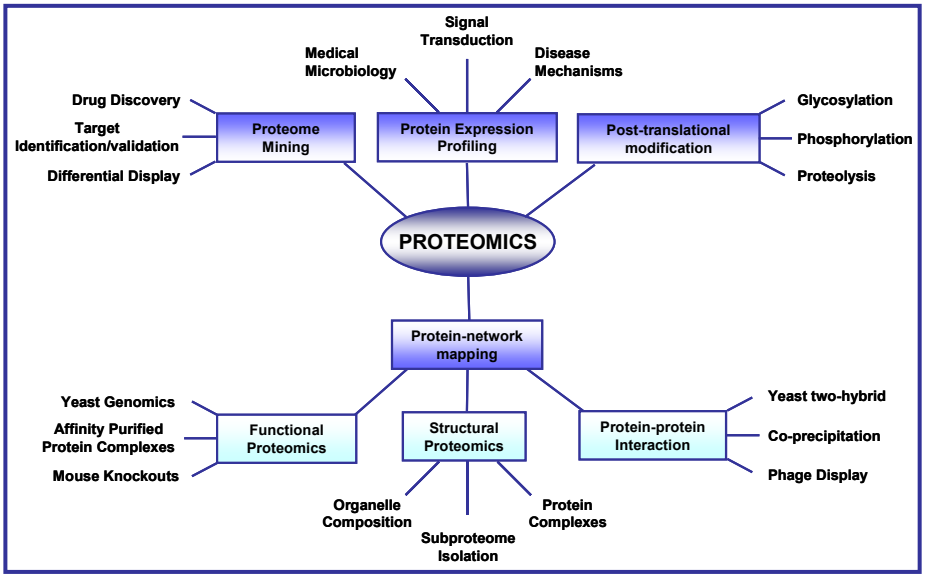


Figure 2. Classifications of proteomics approaches.

3) *Protein-network mapping* is the proteomics approach to determining how proteins interact with each other in living systems. Most proteins carry out their functions in close association with other proteins. It is these interactions that determine the functions of protein functional networks, such as signal-transduction cascades and complex biosynthetic or degradation pathways. However, proteomics approaches offer the opportunity to characterize more complex protein-networks. As such, protein-network profiling represents one of the most ambitious and potentially powerful future applications of proteomics.

4) *Mapping of protein modifications* is the task of identifying how and where proteins are modified. Many common posttranslational modifications govern the targeting, structure, function, and turnover of proteins. In addition, many environmental chemicals, drugs, and endogenous chemicals give rise to reactive electrophiles that modify proteins. Proteomics approaches offer the best means of establishing both the nature and sequence specificity of posttranslational modifications.

Analytical protein identification is built around one essential fact: most peptide sequences of approximately six or more amino acids are largely unique in the proteome of an organism. Put another way, a typical six amino acid peptide maps to a single gene product. Thus, if we can obtain the sequence of the peptide or if we can accurately measure its mass, we can identify the protein it came from simply by finding its match in a database of protein sequences (Figure 3).

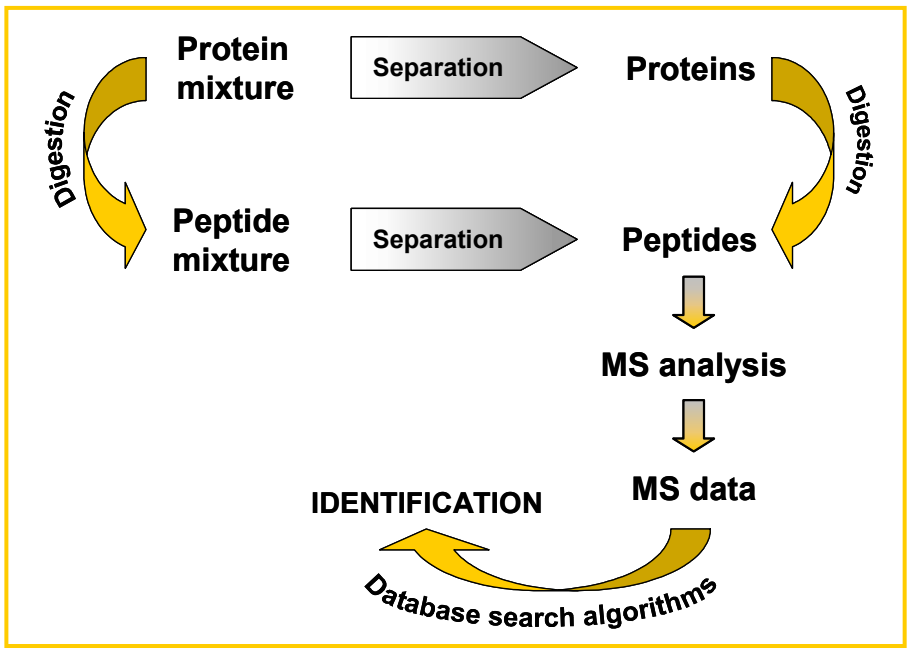


Figure 3. Depicts the essential elements of the analytical proteomics approach.

Most analytical proteomics problems begin with a protein mixture. This mixture contains intact proteins of varying molecular weights, modifications, and solubilities. Before peptide sequences can be obtained, the proteins must be cleaved to peptides. The essence of analytical proteomics is to convert proteins to peptides, obtain sequences of the peptides, and then identify the corresponding proteins from matching sequences in a database. In a typical experiment in proteomics the first step is the sample preparation, the second one is often the separation of mixture proteins, for example by their molecular weight with 1D-gel electrophoresis (SDS–PAGE) or by both their isoelectric point and their molecular weight with 2D-gel electrophoresis (2D-PAGE). Separated proteins are then visualised by staining with silver, Coomassie Blue or fluorescent dyes. In general, proteins are not analysed directly from the polyacrylamide gel although some attempts have been made using direct MALDI-TOF MS analysis from ultra-thin IEF-IPG gel strips[18, 19]. However, the accurate mass of a protein is usually not sufficient to identify it with confidence in sequence databases[20]. Stained bands or spots containing the proteins of interest are usually excised from the whole polyacrylamide gel and digested by specific proteases (Trypsin, LysC or other proteases/chemicals with specific cleavage sites).

The resulting peptide mixture, extracted from the polyacrylamide matrix and further analysed by mass spectrometry, generates an experimental peptide mass profile specific to the protein. This experimental profile, is then compared to the theoretical masses derived from the in silico digestion at the same enzyme cleavage site(s) of all protein sequences of the database. The proteins in the database are then ranked according to the number of peptide masses matching their sequence within a given error tolerance in mass. This process is called peptide mass fingerprinting (PMF)[21,22]. A protein is generally considered identified with sufficient confidence when at least five peptide masses are matched with a mass accuracy better than 10 ppm, 15% of the sequence is covered and the next best database hit shows significant less agreement with the experimental data[23].

An other widespread approach is to digest proteins mix, which is separated by liquid chromatography followed by mass spectrometry analysis.

Sample preparation and purification in Proteomics.

The sample preparation is absolutely essential in many successful experiments and sometimes is not so simple, because biological materials contain either protein of interest or other interfering substances, such as salts, small ionic molecules, ionic detergent, charged molecules, lipid, and non-protein components, and the protein of interest must be isolated by the proper preparation method. Those substances in the sample may lead to difficulty in protein separation and also disturb the detection and identification in proteome studies, so sample preparation is necessary to deplete or entirely remove the interfering substances in the biological samples prior to analysis.

Methods for separating proteins take advantage of properties that vary from one protein to the next, including size, charge, and binding properties. The source of a protein is generally tissue or microbial cells. The first stage, in any protein purification procedure, is to break open these cells, releasing their proteins into a solution called a *crude extract*. Once the extract or organelle preparation is ready, depending on the type of sample, there are various ways to prepare protein sample for further analysis. The general sample preparation methods of greatest interest in proteomics study are pre-fractionation and enrichment of protein prior to further protein separation by preparative electrophoresis or chromatography[24]. The basic methods, including precipitations, dialysis, ultra-filtration and gel filtration, can employ to concentrate the sample and to separate the proteins from potentially interfering substances. Commonly, the extract is subjected to treatments referred to as *fractionation*. Early fractionation steps in a purification utilize differences in protein solubility, which is a complex function of pH, temperature, salt concentration, and other factors. The solubility of proteins is generally lowered at high salt concentrations, an effect called “salting out.” The addition of a salt in the right amount can selectively precipitate some proteins, while others remain in solution. Several are *precipitation methods* relying on different chemical principles, and they can be performed by ammonium sulfate, trichloroacetic acid (TCA), ethanol, chloroform or acetone[25]. Although many protein precipitation methods have the advantages for concentrating and eliminating interferences, they also have the disadvantages of protein irreversible denaturation and protein

insolubilization. Another old established procedure for reducing the salt concentration in samples is *dialysis*; its separation based on principles of diffusion that allows the low molecular weight contaminant removal from sample solutions. Using dialysis method can reduce the maximal interfering substances, but it may have lost the protein in sample, get the high volume of interchanged buffer, need to be concentrated and spend more time than other desalting techniques[26]. In the meanwhile, *ultrafiltration* can remove high molecular weight polysaccharides and salts with a short time and avoid precipitation [27,28]. Although the removal of low molecular weight proteins or interferences can concentrate the protein, but some of high molecular weight interferences are also concentrated that is the disadvantage of this technique. Each sample preparation method has advantages and disadvantages upon the sample composition and the proper preparation method used. Sometime is really fundamental remove the high abundant proteins and enrich the low abundant proteins and enable to increase the quantity of protein identification [29, 30, 31], because the presence of high abundant proteins reasonably obscures the incidence of low abundant proteins that may act as disease biomarkers [32].

Currently, the detection of specific low abundant protein has been studied to increase the dynamic concentration range available for the identification and characterization of proteins by employing commercial removal kits coupled with immune-precipitation technique in different types of antibodies. This techniques combination is sufficient to detect trace proteins, whereas many proteins were less abundant or undetectable. In addition, an immobilized metal affinity chromatography (IMAC) is a separation technique that uses covalently bound chelating compounds on solid chromatographic supports to entrap metal ions, which serve as affinity ligands for various proteins or peptides, making use of coordinative binding of some amino acid residues exposed on the surface, for example exposed histidine residues, which are primarily responsible for binding to immobilized metal ions[33]. This technique has not only proven to be one of the most effective approaches, which is one-way of reducing sample complexity to further enrich the target proteins, but also used for isolating and selective enriching the phosphoproteins or phosphopeptides from complex mixture proteins[34]. On the other hand, an alternative

enrichment approach of low abundant proteins is the gel filtration chromatography that separates the proteins based on size exclusion.

In a typical experiment in proteomics, after the sample preparation, the second really important step is the separation of proteins mixture. *Electrophoresis* is especially useful as an analytical method, it is an important technique for the separation of proteins and it is based on the migration of charged proteins in an electric field. Its advantage is that proteins can be visualized as well as separated, permitting a researcher to estimate quickly the number of different proteins in a mixture or the degree of purity of a particular protein preparation. Electrophoresis allows determination of crucial properties of a protein such as its isoelectric point and approximate molecular weight, it is generally carried out in gels made up of the cross-linked polymer polyacrylamide that acts as a molecular sieve, slowing the migration of proteins approximately in proportion to their charge-to-mass ratio. *Isoelectric focusing* is a specific electrophoretic procedure used to determine the isoelectric point (pI) of a protein. A pH gradient is established by allowing a mixture of low molecular weight organic acids and bases to distribute themselves in an electric field generated across the gel. When a protein mixture is applied, each protein migrates until it reaches the pH that matches its pI. Proteins with different isoelectric points are thus distributed differently throughout the gel. Combining isoelectric focusing and SDS electrophoresis sequentially in a process called *two-dimensional electrophoresis* (2-DE) permits the resolution of complex mixtures of proteins. This is a more sensitive analytical method than either electrophoretic method alone. Two-dimensional electrophoresis separates proteins of identical molecular weight that differ in pI, or proteins with similar pI values but different molecular weights. This separation method has become synonymous with proteomics and remains the single best method for resolving highly complex protein mixtures. Similar to SDS-PAGE proteins separated by 2-DE are visualized by conventional staining techniques, including silver, Coomassie, and amido black stains. Despite the superiority of 2D-SDS-PAGE over other methods as means of resolving complex protein mixtures, the technique presents some problems. The first is the difficulty of performing completely reproducible 2D-SDS-PAGE analyses. This problem becomes important when one wishes to use 2-DE to compare two

samples by comparing the images of the stained gels. Differences in protein migration in either dimension could be mistaken for differences in levels of certain proteins between the two samples. A second problem is the relative incompatibility of some proteins with the first-dimension IEF step. Fox example, many large, hydrophobic proteins simply do not behave well in this type of analysis. A third problem is the relatively small dynamic range of protein staining as a detection technique. Spot densities reflect about a 100-fold range of protein concentrations, at best. This means that staining of 2D-gels allows the visualization of abundant proteins, whereas less abundant proteins frequently cannot be detected.

However an important advantage is the resolution of proteins into multiple, discrete bands due to the presence of multiple protein forms with different isoelectric points. Several protein modifications may affect pI include glycosylation, phosphorylation, oxidation, and exogenous chemical modifications. In some cases, differently modified variants of the same polypeptide may appear as spot "trains". Although this degree of resolution can be useful in establishing what different protein forms are present, it can also complicate the problem of estimating relative protein expression in two samples by 2D-SDS-PAGE.

Anyway, the use of an initial protein separation followed by digestion and analysis is the most widely practiced analytical proteomics approach today. This is based largely on the pre-eminence of 2-DE for protein separations. The biggest single advantage of this approach is the ability of 2D gels to serve as image maps to allow investigators to compare changes in the proteome based on changes in the patterns of spots on the gel. As noted earlier, there are several factors that can confound interpretations of 2-DE gel-spot patterns, but there is no other technique available that provides an intuitive "photograph" of the proteome. However, for lower abundance proteins, 2-DE gels will not prove useful, simply because important proteins cannot be seen. In this case, other separation methods, particularly tandem LC, provide a viable alternative.

High-Performance Liquid Chromatography is an important analytical method employed for protein purification. The diversity of stationary phases and separation modes gives to HPLC considerable resolving power. Although HPLC of intact proteins has not become a

widely used technique for analytical proteomics, it is nevertheless highly applicable as an initial step to fractionate protein mixtures. Different chromatographic separations are available, including RP, anion and cation exchange, size exclusion, and affinity chromatography. More frequent is the use of HPLC separation of proteins, after digestion. The main foundation for this approach is that it permits to convert a very heterogeneous mixture of proteins to a more homogeneous mixture of peptides, which can be more easily analyzed. The use of combined separation modes in series is referred "tandem HPLC". For example, strong cation exchange, followed by RP, would apply two completely different separation modes. The tandem LC approach makes possible the identification of peptides from proteins that are present in a mixture at low abundance. This is in contrast to 2-DE, which is inclined to identify more highly expressed proteins. The superiority of tandem LC over 2-DE probably is owing to two factors, one obvious and the other not so obvious. First, proteins are selected from 2D gels for digestion and MS only if they can be visualized by staining. However, the limits of detection of many MS instruments are below the levels at which proteins can be detected by gel staining. Thus, if one cannot see a protein spot to harvest and analyze, no data will be collected on that protein. Second, handling of proteins in mixtures may provide a "carrier effect," in which the presence of more abundant peptides prevents the loss of less abundant peptides. When one works with very dilute samples with little material (such as would be obtained from a 2D gel spot), the fractional loss due to interaction with surfaces and other processing components is relatively high.

MS and Proteomics.

Mass spectrometry is become a valuable technique in protein analysis as a result of the development of two new ionization methods, MALDI [35,36] and electrospray [37], that allow the routine analysis of biopolymers. These methods solved the difficult problem of generating ions from large, non-volatile analytes such as proteins and peptides without significant analyte fragmentation. Because of the lack or minimal extent of analyte fragmentation during the ESI and MALDI processes, they are also referred to as "soft" ionization methods. In fact they are so soft that under specific conditions even non-covalent

interactions may be maintained during the ionization process. ESI gained immediate popularity because of the ease with which it could be interfaced with popular chromatographic and electrophoretic liquid-phase separation techniques[38]. Furthermore, due to the propensity of ESI to produce multiply charged analytes; simple quadrupole instruments and other types of mass analyzers with limited *m*/*z* range could be used to detect analytes with masses exceeding the nominal *m*/*z* range of the instrument.

For different but no less compelling reasons, MALDI also rapidly gained popularity. The time-of-flight (TOF) mass analyzer most commonly used with MALDI is robust, simple, and sensitive and has a large mass range. MALDI mass spectra are simple to interpret due to the propensity of the method to generate predominantly singly charged ions. The method is relatively resistant to interference with matrixes commonly used in protein chemistry.

In particular MALDI has a number of advantages over electrospray in that the majority of generated ions are detected and the process is more tolerant towards salt and detergents. Likewise, the instrumentation and spectra are simpler. As a result of this, the measurement of individual samples is more easily automated and adapted for higher throughput. Furthermore, as a solid sample is used, the acquisition can be paused at any time[39]. Electrospray, on the other hand, is the interface of choice for coupling liquid chromatography with mass spectrometry to allow the analysis of complex mixtures. As a consequence of its strengths, MALDI has been employed as a fast pre-screening tool in proteomic studies in order to identify gel-separated proteins. Due to its sensitivity this leaves the vast majority of the sample for the then optional more time consuming albeit more powerful electrospray-techniques[40].

Numerous reports document the success MS has enjoyed in studies into the four structural classifications of proteins, namely, the *primary* structure or linear sequence of amino acids, the *secondary* structure or the folding of stretches of amino acids into defined structural motifs, the *tertiary* structure or the overall three-dimensional fold, and the *quaternary* structure or the spatial arrangement of folded polypeptides in multiprotein complexes. However the application of MS to proteomics has to date been realized mostly

for the study of protein primary structures, even if there is an increasing role of MS in the systematic study of protein higher order structures, i.e., *structural proteomics*, as well as of protein-ligand interactions. Because of their relative softness of ionization, ESI and MALDI have been used in attempts to generate gas-phase ions of non-covalently associated, apparently intact protein complexes for the purpose of studying these structures by MS.

Traditionally, proteins have been identified by de novo sequencing, most frequently by the automated, stepwise chemical degradation (Edman degradation) of proteins or isolated peptide fragments thereof[41]. These partial sequences were occasionally used to assemble the complete protein sequence from overlapping fragments but more frequently for the generation of probes for the isolation of the gene coding for the protein from a gene library. With the growing size of sequence databases, it became apparent that even relatively short and otherwise imperfect sequences (gaps, ambiguous residues) were useful for the identification of proteins. This was done by correlating information obtained experimentally from the analysis of peptides with sequence databases. The concept of identifying proteins by correlating information extracted from a protein or peptide with sequence databases rather than by de novo sequencing was significantly enhanced when it was realized that mass spectrometers were ideally suited to generate the required data.

Usually mass spectrometric methods, applied on proteomics, are based on *Peptide Mass Mapping Identification* or on *Protein Identification Using Single Peptides*. Peptide mass mapping is based on the insight that the accurate mass of a group of peptides derived from a protein by sequence-specific proteolysis (i.e., a mass map or fingerprint) is a highly effective means of protein identification, quite the reverse protein identification using single peptides depends on tandem mass spectrometry for the generation of sequence-specific spectra for peptides.

The principle behind protein identification by mass mapping is therefore quite simple conceptually[42]; proteins of different amino acid sequence will, after proteolysis with a specific protease, produce groups of peptides the masses of which constitute mass fingerprints unique for a specific protein (Figure 4).

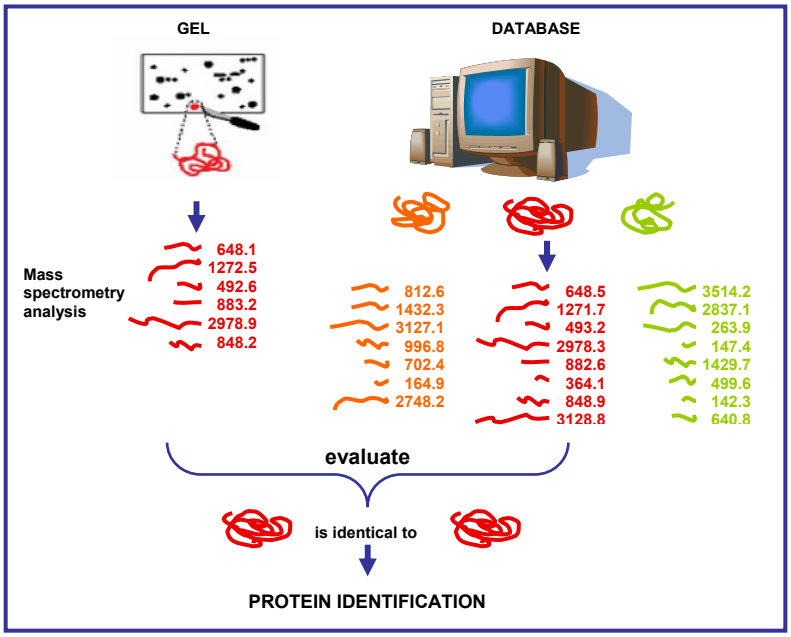


Figure 4. Protein identification by mass spectrometry.

Therefore, if a sequence database containing the specific protein sequence is searched using selected masses (i.e., the observed peptide mass fingerprint), then the protein is expected to be correctly identified within the database. Various methods that automate this process have been developed and reviewed[43]. They vary in specific details but share the following sequence of steps:

(i) Peptides are generated by digestion of the sample protein using sequence-specific cleavage reagents that allow residues at the carboxyl- or amino-terminus to be considered fixed for the search. For example, the enzyme trypsin that is popular for mass mapping leaves arginine (R) or lysine (K) at the carboxyl-terminus, and the N-termini of tryptic peptides (except for the N-terminal one) are expected to be the amino acid following a K or R residue in the protein sequence.

(ii) Peptide masses are measured as accurately as possible in a mass spectrometer. An increase in mass accuracy will decrease the number of isobaric peptides for any given mass in a sequence database and therefore increase the stringency of the search.

(iii) The proteins in the database are "digested" in silico using the rules that apply to the proteolytic method used in the experiment to generate a list of theoretical masses that are compared to the set of measured masses.

(iv) An algorithm is used to compare the set of measured peptide masses against those sets of masses predicted for each protein in the database and to assign a score to each match that ranks the quality of the matches.

Obviously, for a protein to be identified its sequence has to exist in the sequence database being used for comparison. Both protein and DNA sequence databases are equally suited. If DNA sequence databases are being used, the DNA sequences are translated into protein sequences prior to digestion. The approach is therefore best suited for genetically well-characterized organisms where either the entire genome is known or extensive protein or cDNA sequence exists.

Clearly, protein identification by peptide mass mapping depends on the correlation of several peptide masses derived from the same protein with corresponding data calculated from the database. For this reason the method is suited neither for searches of EST (Expressed Sequence Tags) databases nor for identification of proteins in complex mixtures if un-separated mixtures are proteolyzed. ESTs present a problem because they only represent a portion of a gene's coding sequence. Such segments may not be long enough to cover a sufficient number of peptides observed in the mapping experiment to allow an unambiguous identification. Digests of un-separated protein mixtures present a problem for mass mapping because it is not apparent which peptides in the complex peptide mixture originate from the same protein. To overcame this problem PMF is often combined with tandem MS of peptides in an iterative approach where as much information as possible is extracted by mass mapping, and this is followed by tandem MS to resolve the identification of any ambiguous remaining masses. In a MALDI-TOF spectra from real samples, there are typically dozens of *m/z* signals. Peptide mass fingerprinting software can usually match just

about all of these to some entry in a database. However, given errors in *m/z* measurement, frequent sample contamination, and the presence of unanticipated posttranslational modifications, not all of the matches will point to the same proteins. The simplest approach is to assign the highest score to proteins whose predicted tryptic peptides match the greatest number of *m/z* signals in the MS data. If we search only one *m/z* value, then several proteins could be equally good matches. However, as we search a greater number of *m/z* values, more matches correspond to a particular protein and lead to a greater score for that protein vs others. This fairly simple approach works reasonably well with very good MS data. However, it tends to assign higher scores to larger proteins. However larger proteins yield more tryptic peptides, so the chances of a match to one of these is greater for larger proteins than for smaller proteins.

Sponsor (application)	**Uniform resource locator (URL)**
Eidgenossische Technische Hochschule (MassSearch)	http://cbrg.inf.ethz.ch
European Molecular Biology Laboratory (PeptideSearch)	http://www.mann.emblheidelberg.de
Swiss Institute of Bioinformatics (ExPASy)	http://www.expasy.ch/tools
Matrix Science (Mascot)	http://www.matrixscience.com
Rockefeller University (PepFrag, ProFound)	http://prowl.rockefeller.edu
Human Genome Research Center (MOWSE)	http://www.seqnet.dl.ac.uk
University of California (MS-Tag, MS-Fit, MS-Seq)	http://prospector.ucsf.edu
Institute for Systems Biology (COMET)	http://www.systemsbiology.org
University of Washington (SEQUEST)	http://thompson.mbt.washington.edu/sequest

Table 2. Sources for MS-Based Protein Identification Tools.

To address these problems, several of the available peptide mass fingerprinting programs use more sophisticated scoring algorithms (Table 2). These algorithms correct for scoring bias due to protein size, in which larger proteins give rise to greater numbers of peptides. They also correct for the tendency of smaller peptides in databases to have a greater number of matches with searched *m/z* values. Finally, some of these algorithms also apply probability-based statistics to better define the significance of protein identifications. The principal tools available for peptide mass fingerprinting can be grouped into three categories:

• First-generation freeware and subscription software tools that assign scores based on the number of *m/z* values in a spectrum that match database values within a given mass tolerance. These programs include PepSea (http://www.protana.com) and Pept Ident/MultIdent (http://www.expasy.ch/tools/peptident.html).

• Second-generation freeware and subscription software tools that employ scoring algorithms that take into account the effects of protein size and peptide length on the probabilities of matching. These include MOWSE (http://srs.hgmp.mrc.ac.uk/cgibin/mowse) and MS-Fit (http://prospector.ucsf.edu/).

• Third-generation software that employs more extensive probability-based scoring to provide a statistical basis for scores and also to estimate the probabilities that matches may reflect random events, rather than true identities.

These programs include ProFound (http://prowl.rockefeller.edu/cgi-bin/ProFound) and Mascot (http://www.matrixscience.com/).

If a pure protein is digested and the resulting peptide masses are compared with the list of peptide masses predicted for that protein, two observations are typically made. First, not all of the predicted peptides are detected. Second, some of the measured peptide masses are not present in the list of masses predicted from the protein. The first problem, the missing masses, is usually due to a number of problems that can occur both before and during mass spectrometric analysis such as poor solubility, selective adsorption, ion suppression, selective ionization, very short peptide length, or other artefacts that cause sample loss or make specific peptides undetectable by MS. Since a relatively low number of peptide masses are sufficient for the positive identification of a protein, missing peptide masses are not generally considered a problem. In contrast, unassigned peptide masses are a significant problem for protein identification by mass mapping and probably the single biggest source of misidentifications or missed identifications. Thus, to ensure that mass mapping results are reliable, it is important to understand the possible reasons for unassigned masses and to learn how to deal with them[44, 45]. Unassigned masses may be observed for one or more of the following reasons:

(i) Changes in the expected peptide masses by posttranslational modification (e.g., phosphorylation adds a net 80 u to an amino acid mass), art-factual modifications arising from sample handling (such as oxidation of methionine), or posttranslational processing (e.g., amino- or carboxyl- terminal processing). Some of these changes can be anticipated and incorporated into the search algorithm.

(ii) Low fidelity proteolysis due to the presence of contaminating proteases that produce peptides unanticipated by the search algorithm (e.g., the presence of chymotryptic activity in a trypsin preparation) or missed cleavage sites. Again, this can be anticipated to some degree by the search algorithms.

(iii) The presence of more than one protein in the sample. It needs to be stressed that bands in SDS gels frequently and spots in 2D gels occasionally contain more than one protein, even if the respective features appear concise and sharp. In some cases, additionally present proteins can be detected by iterative database searching with the masses left unassigned to the primary target protein. Keratins and other common proteins represent another source of protein contamination.

(iv) The identified protein actually matches a sequence homologue or splice variant of that reported in the database. This must be confirmed using the sequence of genetically well-characterized species.

(v) The protein is misidentified (i.e., false-positive). In this context, the specificity of the enzymes employed for protein digestion should be discussed in more detail. Obviously, the higher the fidelity of the enzyme in hydrolyzing peptide bonds, the more reliably the search can be done with a fixed amino- or carboxyl-terminus. The frequent observation that the protease products are not limited to the ones predicted from the expected enzymatic recognition sites is often due to contaminating protease activity but may also be due to a post-translational modification juxtaposed to the recognition site that blocks access by the enzyme or also to an inefficient proteolysis. If this problem is anticipated, algorithms can be programmed to accommodate missed cleavages by allowing a given number to be entered as a parameter. Furthermore, the success of proteases to cleave proteins is

dependent on accessibility to open stretches of primary amino acid sequence, and the native three-dimensional structure of the substrate protein will block access to many sites.

Data for use with peptide mass mapping are commonly obtained via MALDI-TOF analysis. However, any mass spectrometer capable of generating mass accuracies around 100 ppm or better at 1000 u, in particular ESI-TOF and FT-ICR instruments, can be used to generate a mass map. For MALDI, analytes are spotted onto a metal plate either one at a time or, in a higher throughput format, multiple samples on the same plate. The samples are usually tryptic digests from proteins separated by 2DE, although proteins purified by other separation methods are also compatible with the method. Before deposition of the analytes, the matrix is placed on the plate or mixed in with the sample. The matrix will absorb energy from the laser causing the analytes to be ionized by MALDI (figure 5). The m/z ratio of the ions is then typically measured based on the flight time in a field-free drift tube (as opposed to ion mobility MS where a field pushes ions through a gas) that constitutes the heart of the time-of-flight mass (TOF) analyzer. An additional bonus for samples isolated from biological sources is that MALDI is compatible with biological buffers such as phosphate and Tris and low concentrations of urea, nonionic detergents, and some alkali metal salts. Peptide m/z ratios are calculated based on the energy equation ion $E=1/2mv^2$ that accounts for contributions from kinetic energy, mass, and velocity. At a constant energy, low molecular weight ions will travel faster than high molecular weight ions

An inherent problem with the MALDI process is the small spread of kinetic energy that occurs during ionization. The spread reduces the resolving power and prevents the observation of the natural isotope distribution, even of small peptides. Two approaches, an ion mirror (reflectron) and “time-lag focusing” (delayed extraction), have been implemented in commercial instruments to overcome this problem. A reflectron is a device located at the end of the flight tube opposite from the ion source that decelerates the ions and then re-accelerates them back out of the reflectron toward a second detector. This is achieved by applying a decelerating voltage that is slightly higher than the accelerating voltage at the source. It has been observed that ions of lower kinetic energy do not penetrate

as far into the reflectron as those of higher energy. Consequently, deeper penetrating high-energy ions can catch up, thereby decreasing the initial energy spread.

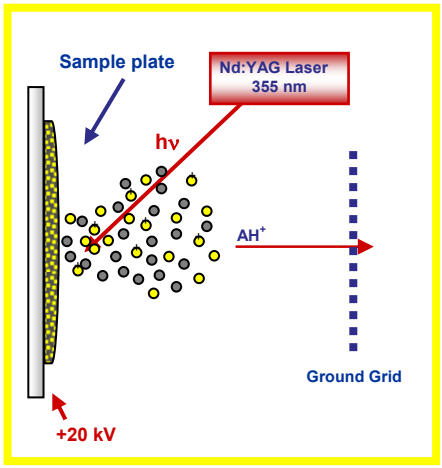


Figure 5. MALDI source.

The second approach to correct the initial spread of kinetic energies during MALDI is the time-lag focusing technique initially developed by Wiley and McLaren, in 1953, and more recently reintroduced as "delayed extraction"[46]. In this method, the MALDI ions are created in a field-free region and allowed to spread out before the extraction voltage is applied to accelerate them for their flight through the drift tube. This results in a significantly decreased energy spread of ions and thus higher resolution. Delayed extraction also limits peak broadening due to metastable decomposition from ions colliding in the source during continuous ion extraction. The effects of these improvements are significant.

Delayed extraction can increase the mass resolution to ≈2000-4000 for peptides in a linear instrument and, if combined with a reflectron instrument resolution, can further increase to ≈3000-6000[47].

Large-scale protein identification critically depends on tandem mass spectrometry for the generation of sequence-specific spectra for peptides, the approach called *Protein*

Identification Using Single Peptides. Different amino acid compositions and permutations of an amino acid sequence can result in isobaric peptides. The amino acid sequence of a peptide is therefore more constraining than its mass for protein identification by sequence database searching[48]. At the mass accuracy achieved with the MALDI-TOF mass spectrometers that are frequently used for peptide mass measurement (10-100 ppm), several peptide masses from the same protein are required for unambiguous identification, whereas the amino acid sequence of even a relatively small peptide can uniquely identify a protein.

Tandem mass spectrometers have the ability to fragment peptide ions and to record the resulting fragment ion spectra. For tandem mass spectrometers such as triple quadrupole, ion trap, quadrupole/TOF or TOF/TOF instruments, fragment ion spectra are generated by a process called collision-induced dissociation (CID) in which the peptide ion to be analyzed is isolated and fragmented in a collision cell, and the fragment ion spectrum is recorded. Typically these types of mass spectrometers are used in conjunction with ESI, exception for the TOF/TOF system that is usually used in conjunction with a MALDI source. This instrument is composed by coupling two TOF mass spectrometers together via a collision cell between them. This new design combines the advantages of MALDI such as high sensitivity for peptide analysis, relative insensitivity to salts, surfactants, and other contaminants, with high-energy CID where amino acids such as isoleucine and leucine can be distinguished by side-chain fragmentation. As with other types of sequencing mass spectrometers, a complete CID spectrum can be acquired in a single acquisition, obviating the need to sum as many as 10 spectra as is necessary with PSD on a single TOF mass spectrometer. Additionally, the MALDI-TOF/TOF mass spectrometer promises to be capable of acquiring tandem mass spectra at a rate that is an order of magnitude above the capabilities of IT and QTOF instruments, which will be significant for proteome studies.

However tandem mass spectra, generated by the fragmentation of peptide ions in the gas phase at low collision energy, are dominated by fragment ions resulting from cleavage at the amide bonds. Very little amino acid side chain fragmentation is observed. Such spectra are much less complex than the high collision energy spectra generated, for example, in TOF/TOF instruments. The low-energy CID spectra generated by the types of

mass spectrometers most frequently used in proteomics are therefore relatively simple to interpret, and a straightforward nomenclature for annotating the MS spectra has been adapted (Figure 6).

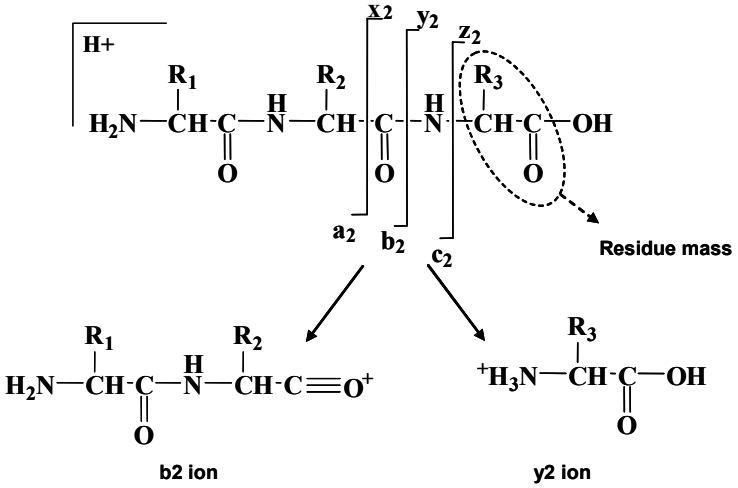


Figure 6. Peptide fragment ion nomenclature.

The nomenclature differentiates fragment ions according to the amide bond that fragments and the end of the peptide that retains a charge after fragmentation. If the positive charge associated with the parent peptide ion remains on the amino-terminal side of the fragmented amide bond, then this fragment ion is referred to as a *b* ion. However, the fragment ion is referred to as a *y* ion if the charge remains on the carboxyl-terminal side of the broken amide bond. Since in principle every peptide bond can fragment to generate a *b* or *y* ion, respectively, subscripts are used to designate the specific amide bond that was fragmented to generate the observed fragment ions. *b* ions are designated by a subscript that reflects the number of amino acid residues present on the fragment ion counted from the amino-terminus, whereas the subscript of *y* ions indicates the number of amino acids present, counting from the carboxyl-terminus. These individual fragment ion *m*/*z* values as

shown in figure 6 can be easily calculated from the amino acid sequence, using the nominal (i.e., monoisotopic value rounded to an integer value) residue masses found in Table 3.

While it is relatively simple to calculate the elements of the *b* and *y* ion series from the peptide sequence, it is much less straightforward to read the amino acid sequence from the CID spectrum of a peptide ion. This is mainly because peptide fragmentation under the conditions encountered in the collision cell of a mass spectrometer are sequence dependent, and the rules for fragmentation are not completely understood.

amino acid (3/1 letter codes)	nominal residue mass	immonium ion mass
alanine (Ala /A)	71	44
arginine (Arg/R)	156	129
aspartic acid (Asp/ D	115	87
asparagine (Asn/N)	114	88
cysteine (Cys/C)	103	76
glutamic acid (Glu/E)	129	102
glutamine (Gln/Q)	128	101
glycine (Gly/G)	57	30
histidine (His/H)	137	110
isoleucine (Ile/I)	113	86
leucine (Leu/L)	113	86
lysine (Lys/K)	128	101
methionine (Met/M)	131	104
phenylalanine (Phe/F)	147	120
proline (Pro/P)	97	70
serine (Ser/S)	87	60
threonine (Thr/T)	101	74
tryptophan (Trp/W)	186	159
tyrosine (Tyr/Y)	163	136
valine (Val/V)	99	72
alanine (Ala /A)	71	44

Table 3. Residue and Immonium Ion Masses of 20 Common Amino Acids.

The CID spectrum of a peptide ion acquired at low collision energy can be considered a composite of many discrete fragmentation events. Each peptide tandem mass spectrum will contain *b* and *y* ions as well as other fragment ions that can be used to interpret the amino acid sequence. These include diagnostic ions generated by the neutral loss of specific groups from amino acid side chains (e.g., the loss of ammonia (-17 u) from Gln, Lys, and

Arg or of water (-18 u) from Ser, Thr, Asp and Glu) and low mass ions that result from the fragmentation of amino acids down to a basic unit consisting of the side chain residue and an immonium functionality (Figure 6). The *b* ion series also often shows a satellite ion series in which each signal is 28 u lower than the corresponding *b* ion. These signals result from the neutral loss of carbon monoxide and are referred to as an *a* ion series. CID spectra can be further complicated by the presence of internal fragment ions that represent some contiguous sequence of amino acids in the peptide. These are generated if a specific peptide ion undergoes two or more fragmentation events. Empirical observation shows that internal fragments often occur if either proline[49] or aspartic acid residues are present in a sequence and even more so at any aspartyl-proline bond, indicating that not all peptide bonds have the same propensity to fragment during low energy CID. For the same reason, even if some of the rules that control peptide ion fragmentation in a collision cell have been determined, others remain to be studied, and of course the relative intensity of fragment ions in peptide CID spectra is uneven and somewhat unpredictable[50].

Furthermore the choice of the enzyme used for the proteolysis is very important, if proteins are completely digested with trypsin, then lysine or arginine residues will be present at the carboxyl-terminus of all peptides except for the C-terminal peptide of the original protein. A charge sequestered by lysine or arginine at the C-terminus tends to produce a more complete series of *y* ion fragments than will be generated by peptides produced by protein digestion with chymotrypsin or other protease where lysine and arginine are distributed throughout the sequences rather than at the C-terminus.

For peptide mass mapping, the information collectively contained in the masses of several peptides derived from the same protein is used for protein identification by database searching. In contrast, the CID spectrum of a single peptide can, in principle, contain a sufficient amount of information for unambiguous identification of a protein. Therefore, if a mixture of several proteins is concurrently digested, the components of the mixture can be identified based on the CID spectra, provided that at least one CID spectrum per protein is generated. It is hence no longer necessary to separate proteins to homogeneity prior to proteolysis.

Tandem MS has now become the definitive approach to determination of peptide sequences. There are two ways to identify proteins from peptide MS-MS spectra. The first is *de novo* interpretation of the spectrum to obtain a peptide sequence followed by BLAST searching of the sequence against a sequence database to identify the protein. This is a perfectly reasonable approach as long as there are only a few spectra to deal with. Manual *de novo* interpretation of an individual MS-MS spectrum takes between half an hour and a couple of days, depending on the complexity of the spectrum and the experience of the analyst. As noted earlier, some spectra do not contain complete b- or y-ion series and thus it may not be possible to unambiguously interpret a peptide sequence from these spectra. Unfortunately, the emerging field of proteomics relies on identification of large numbers of proteins from MS-MS spectra. Clearly, the *de novo* sequencing/BLAST searching approach will be too slow for large-scale protein identification. The "slow step" in this case is the manual inspection of MS-MS spectra to determine sequence.

The second approach to protein identification bypasses the "slow step" (manual *de novo* sequence interpretation). In this approach, algorithms are applied to directly correlate MS-MS spectral data with peptide sequences in databases without actually interpreting each MS-MS spectrum individually. The only limitations to such an approach are the quality of the MS-MS spectra and the completeness and accuracy of the databases. If we obtain an MS-MS spectrum of a peptide whose sequence exists in a database, the right algorithm should be able to make the match. The right algorithms can match MS-MS data to protein sequences or to nucleotide (e.g., genome or EST) sequences that are translated to protein sequences. If the sequence of the analyzed peptide does not exist in the database, a correct match cannot be made.

The constraints on database searching of a given stretch of peptide sequence are so powerful that the tandem MS spectrum of a single peptide can be adequate for protein identification in an EST database. The approach is easily automated and can also be adapted to find peptides carrying specified posttranslational modifications by instructing the program to anticipate modification at specific residues [51]. A list of some Internet sites

with protein identification resources developed by these and other investigators can be found in Table 2.

Such algorithms use readily available constraints in a decision-making process that distinguishes the correct match from all other sequences in the database. The availability of complete sequence databases, the development of mass spectrometric methods, and the sequence database search algorithms therefore converged into a mature, robust, sensitive, and rapid technology that has considerably advanced the ability to identify proteins and constitutes the basis of the emerging field of proteomics.

In this research work Mass Spectrometry assumes a likely central place in the application of some proteomics approaches, where Proteomics is synonymous with "*anything to do with proteins*" overtaking throughout *mining*, *protein-expression profiling* and *mapping of protein modifications*.

Some specific advances were used to characterize proteins with particular properties, for example allergens or endogenous proteases, obtaining chemical information about proteins without preventive classical separation, but only by mean of chemical fractionation procedure followed by mass spectrometry MALDI TOF-TOF.

Two natural matrixes were analyzed: olive pollen tree and raw milk from cows affected by mastitis. The first step was that to obtain a reproducible procedure of extraction and fractionation of the total protein content from natural matrixes, followed by the profiling by means of mass spectrometry[52]. Protein and peptide expression provided the possibility to individuate specific and functional markers, and to characterize post-translational modification involved into the protein bioactivity. Peptide Mass Fingerprint, followed by MS-MS experiments were adopted to identify and characterize peptides and proteins [24, 23, 53,54,55].

Protein expression profile by MALDI mass spectrometry was employed to determine the antigenic profile of *Olea europaea* pollen from different Mediterranean cultivars, followed by the full characterization of the major observed allergen, including post-

translational modifications employing the synergic develop of mass spectrometry and bioinformatics tools[56].

An upgrading of this procedure was employed to obtain the protein MS profile of the content of raw bovine milk, revealing the presence of a functional marker for the acute phase of mammary gland inflammation. It was possible to suggest, also, a new biomarker of mastitis and obtain roundabout information to understand the function of several specific endogenous milk proteases[57].

Finally, it was necessary wholly characterization of the most important proteins by means of MSMS experiments and database search, using different algorithms, for example PeptideCutter (www.expasy.org) to simulate specific enzymatic cleavage or GlycoMode (www.expasy.ch/tools/glycomod) to identify the glycan forms of an important pollen allergen.

PROTEIN-EXPRESSION PROFILING:

Olea Europaea Olive Pollens.

Proteins are fundamental and integral food components, both nutritionally and functionally, they are a source of energy and amino acids, which are essential for growth and maintenance. Functionally, they affected the physiochemical and sensory properties of various proteinaceous foods. In addition, many dietary proteins possess specific biological properties, which make these components potential ingredients of functional or health-promoting foods. The proteins playing important role in human diet can be divide into three main groups: animal, plant and microbial proteins. Applied genomics technologies (Transcriptomics, Proteomics, Metabolomics, Nutrigenomics, etc.) contribute to different research areas of the nutritional science and food technology (Table 4).

Combining proteomic technologies with genetics, molecular biology, protein biochemistry, biophysics and bioinformatics will result in accelerated discovery of protein functional information. In the study of the proteomes from natural matrixes, the real difficulty is to identify extensively all proteins of a given organ and analyse the physiological events occurring during a definite stage. In vegetable samples were found changes in abundance of proteins during the time of germination, but variation of protein expression were also found during priming (pre-germination followed by drying) a treatment that allow faster germination, imbibition of seed, dehydratation, mobilization of storge proteins and so on[58].

The proteomes of the different organs of a plant are obviously different. They are often studied separately in proteome database[59], but comparison between them are inadequate,

and most of them are actually related to the study if genetic variations. Several studies have demonstrated that organ-specific proteins are more variable between genotypes than organ-unspecific proteins, and that the level of genetic variant depend on the organ or tissue considered[60]. The higher level of genetic variation of organ-specific proteins amounts is probably related to a higher number of genes controlling their expression. Another important difficulty is represented by the absence of specific proteomics database for plant and vegetable proteins, or for food allergens and also for storage proteins. Since the proteomic study on "*green plant world*" is only at the beginning, the characterization of proteins from some specific natural matrixes (plants, foods, fruits) is more difficult than of proteins from human tissues or cells.

Areas of the nutritional science and food technology	
1. ***Screening for novel functional bioactives.***	• Availability of rapid screening methods for detection of bioactivity.
2. ***Safety evaluation of food constituents.***	• Evaluation of absorption, body distribution and metabolism of food ingredients.
3. ***Detection and control of food.***	• Identification of biomarkers (metabolites, proteins) specific for particular food spoilage and/or pathogenic micro-organism.
4. ***Efficacy testing of bioactive food ingredients.***	• Changes in genes expression and proteome relevant to the states or treatment of certain diseases.
5. ***Food allergy.***	• Identification of allergic proteins through sophisticated proteomics based on recognition of specific posttranslational modification.
6. ***Quality and authenticity of foods.***	• Proteome of certain food (wheat, wine, fish) can be used to authenticate food origin or food quality.
7. ***Production of food ingredients.***	• The yield of bioprocess may be controlled through metabolome/proteome of micro-organism used for such production.
8. ***Food processing.***	• Proteome and/or metabolome of starter culture of fermentation processes (beer, cheese, sausage, etc.) can be used to predict the quality of the fermented end-product.

Table 4.

The term *Allergonomics*, it was coined to designate the use of proteomics approaches to the study of the allergens. Allergens are defined by their ability to cause the induction of

hypersensitivity response when encountered by an immune system of sensitive individuals. Inhalation or ingestion of potential allergens leads to production of allergen-specific IgE antibodies. The incidence and severity of allergic disorders is steadily increasing worldwide. Exposure to common environmental antigens is the cause of allergic conditions such as hay fever, allergic asthma, and eczema affecting up to 25% of the population in developed countries. Most of the inhalant or food allergens of plant origin are proteins ranging from 10 to 50 kDa. Pollen grains of various weeds, trees, and grasses are significant source of inhalant allergens.[61]

Olive (*Olea Europaea*) pollen is considered as one of the most important causes of respiratory allergic disease in the Mediterranean region. In Spain[62], southern Italy[63], Greece[64] and Turkey[65], olive pollen is an important cause of pollinosis. The main pollen season is from April to June. The frequency of olive-induced pollinosis is increasing as a consequence of improved diagnostic procedures and as a result of changes in farming pratices[66]. Olea europaea pollinosis is clinically characterized by rhinoconjunctival symptomatology than bronchial asthma. Moreover, polysensitization to olive pollen is more frequent than monosensitization[50,51]. In sourthern Italy, the frequency of positivity to Olea pollen allergens among all skin prick test-positive patients is 13.49% in adults and 8.33% in children. In pollinosis patients of the Naples area, monosensitization to olive was identified in only 1.33% of children and in 2.28% in adults; in all the remaining patients, sensitization to olive pollen was associated with other allergens, mainly derived from pollen grains[50]. Interestingly, children and adults with monosensitization to olive are frequently affected by year-long symptoms that usually do not increase during the olive-pollen season.

The antigenic profile of *Olea europaea* pollen from different Mediterranean cultivars was obtained by MALDI mass spectrometry using a simple procedure of chemical fractionation of the whole antigen extract. Some of the features of protein structure and distribution probably depend on cultivar adaptation to the environment.

Mass spectrometry is currently applied, with success, in protein profiling of natural matrixes.[67.68] Our group has developed high-tech analytical methods as tools for the assessment of food quality and safety.[69,70] In a survey of all possible allergen candidates

whose profiling could provide clues for the unambiguous identification of olive cultivars and of their subvarieties, we have undertaken a detailed analysis of olive pollen extracted from eight different typical Mediterranean cultivars.

When inhaled, olive pollen of *Olea europaea* is an important as a causative agent for type I allergy in the Mediterranean area.[71,72] More than 30% of the population in this area is affected by type I allergy during the pollination season, and more than 80% of the olive-tree-allergic patients are sensitive to the protein Ole e 1, the major olive pollen allergen.[73] Several separation methods have been employed for the isolation of the allergens, such as SDS-PAGE,[74,75] high performance liquid chromatography (HPLC),[76] immunodetection,[77] and gel filtration.[78] Allergenic candidates of 7, 9, 14, 15, 16, 18, and 36 kDa (Table 5, Part 1 and 2, § Appendix: **A.1**), whose presence in olive pollen has been ascertained as previously mentioned, have been immune-stained with sera from olive-allergic individual patients.[79,80]

Allergen name	MW, kDa, SDS-PAGE[1]	MW (Da)[2]	P.I.	Sequence[3]	Accesion number[4]
Ole e 1	18-21	16330	6.18	C	P19963
Ole e 2	15-18	14489	5.06	C	O24169
Ole e 3	9.2	9356	4.49		O81092
Ole e 4	32	2711	3.77	P	P80741
Ole e 5	16	2973	4.65	P	P80740
Ole e 6	10	5833	4.96	C	O24172
Ole e 7	9-11	9905-10032[5]	3.56	P	P81430
Ole e 8	21	18907	4.51	C	Q9M7R0
Ole e 9	46	48830	5.21	C	Q94G86

Table 5_Part 1. Olive Pollen Allergens with Clinical Relevance Developed and Maintained by Allergen Nomenclature _ Subcommittee of the IUIS (www.allergen.org), Including Allergens Whose IgE Reactivity Has a Prevalence of >5%**. Legend: (1)** Apparent molecular mass in SDS-PAGE. (**2**) Theoretical molecular weight. **(3)** Sequence information obtained by C, cDNA; P, peptide sequence;N, nucleotide sequence. **(4)** Swissprot database. **(5)** Mass spectrometry determination. (**Table 5_Part 2 continues in Appendix, A1**).

The concentration level of the major olive pollen allergens, estimated using monoclonal antibodies alone or in combination with gel scanning densitometry,[53,81] indicates a variation between plant species.[82] Cultivars and probably local variety or sub-varieties of olive trees

present special features that depend on their adaptation to the environment or ecotype. Ecosystem and crop management are factors that are able to induce changes in the allergenic profile of a given variety or cultivar.[83,84] Therefore, pollen protein profiling could be a useful tool for cultivar discrimination. The olive (*Olea europaea*) pollens of the Mediterranean cultivars, Ottobratica (**1**), Carolea (**2**), Dolce di Rossano (**3**), Cassanese (**4**), Coratina (**5**), Nocellara del Belice (**6**), Villacidro (**7**), and Sinopolese (**8**), were selected as case studies to determine a protein profile of the whole extract and to identify and characterize specific proteins without any previous chromatographic or two-dimensional gel separations.

A simple procedure of chemical fractionation of the whole antigen extract was developed, whereby less complex, or pure, fractions of antigen candidate were obtained. Portions (50 mg) of pollen grains (**1**-**8**) were extracted with 1 mL of aqueous 50 mM NH_4HCO_3 for 20 min at room temperature, followed by centrifugation at 14 000 rpm for 2 min. The supernatant portion (saline extract) was separated and stored at -20 °C. A 200 µL portion of whole extract was precipitated with 400 µL of $CHCl_3/CH_3OH$ 1:3 (v/v), and the precipitated protein pellet was partitioned consecutively, under magnetic stirring and at room temperature, for 10 min with (**a**) 150 µL of 50 mM NH_4HCO_3/CH_3OH 1:1 (v/v) and (**b**) 150 µL of 50 mM NH_4HCO_3/CH_3CN 1:1 (v/v) (Chart 1).

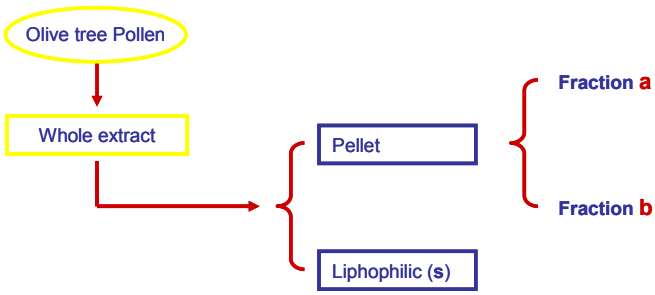


Chart 1. Procedure of chemical fractionation.

All fractions (**a**, **b**)were directly analyzed by MALDI TOF in the linear mode. A 1 μL portion of each fraction was directly analyzed by linear MALDI using α-cyano-4-hydroxy-*trans*cynnamic acid (α-CHCA, 0.3% in TFA) as matrix. MALDI-TOF analyses were performed using a 4700 Proteomics Analyzer mass spectrometer from Applied Biosystems (Foster City, CA) equipped with a 200-Hz Nd:YAG laser at 355-nm wavelength. Linear MALDI MS spectra were acquired averaging 2500 laser shots with a mass accuracy of 500 ppm in default calibration mode that was performed using the following set of standards: insulin (bovine, $[M + H]^+$ avg m/z 5734.59), apomyoglobin (horse, $[M + H]^{2+}$ avg m/z 8476.78, $[M + H]^+$ avg m/z 16 952.56), and thioredoxin (*Escherichia coli*, $[M + H]^+$ avg m/z 11 674.48).

The high solubility in aqueous medium is an important prerequisite for allergen candidates, because their biological activity better correlated to the concentration and rapid release from airborne particle than to their intrinsic properties.[85] The partition coefficient of allergens and the antigenic profile of olive pollen are strongly related to the solvents used for the extraction.[86]

Whole protein extracts can be, therefore, chemically fractionated, and the antigen contents of each fraction can be varied according to the selected experimental conditions. MALDI mass spectrometry, for its specificity and better resolution in comparison with the conventional 2-D (two-dimensional gel electrophoresis) chromatographic approach, represents the methodology to obtain reliable results in the profiling of olive pollen.

The use of saline solution containing sodium chloride, phosphate buffer, and borate buffer either to prepare the whole pollen extract or for its chemical fractionation are not agreeable to direct MALDI-MS analysis. Therefore, ammonium bicarbonate solutions were chosen to prepare the whole antigenic extract from pollen samples of cultivar **1**-**8**. Ammonium bicarbonate should favour the formation of carboxylate/ammonium ion pairs, thus affecting the solubility of the proteins to be extracted in moderately polar solvents, such as acetonitrile/water mixtures. Moreover, ammonium counter-ions have been often used to improve the desorption of high molecular weight protein[87] and does not interfere with the mass spectrometric analysis.

The sample preparation protocol was planned to distribute the amount of information stored in the proteome of each olive pollen in a set of three MALDI spectra that could be independently evaluated and matched to retrieve data for their comparison. The data set displayed by the three spectra provides the entire profiling of a given entity. One lypophilic (**s**) and two hydrophilic fractions (**a**-**b**), respectively, can be obtained from each sample (Chart 1).

Accordingly the antigenic profile of Cassanese **4** displayed four allergens: Ole e 7, Ole e 6, Ole e 2, and Ole e 1 in the 5-20-kDa mass range (Figure 7).

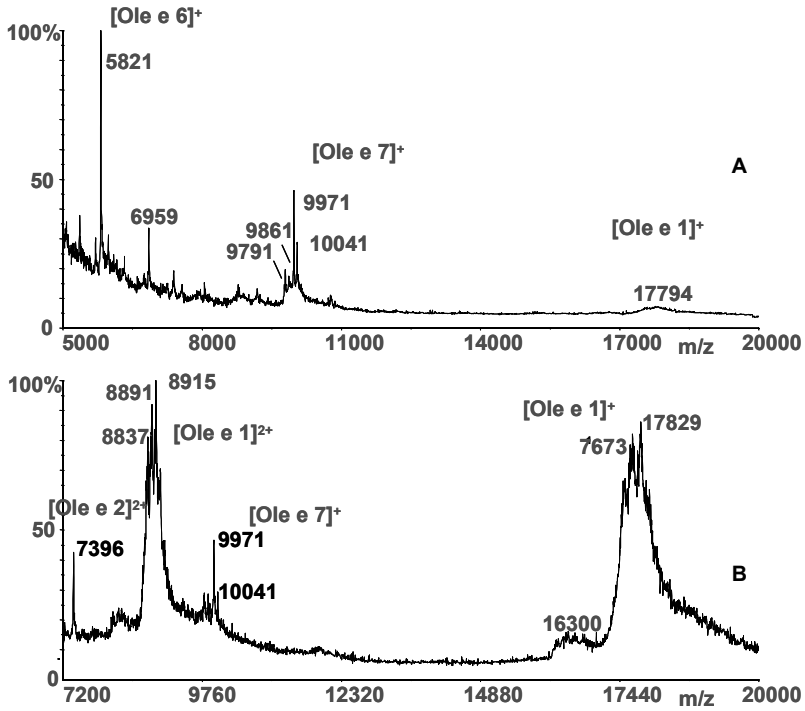


Figure 7. MALDI spectra of fractions (a) **4a** and (b) **4b**. Both fractions represent a mixture of allergens. Part (a) shows the typical pattern of Ole e 7, Ole e 6, and some polypeptides between 5 and 7 kDa; whereas part (b) shows the ion species $[Ole\ e\ 2]^{2+}$, $[Ole\ e\ 1]^{2+}$, $[Ole\ e\ 7]^{+}$, and $[Ole\ e\ 1]^{+}$.

The four ion peaks at *m*/*z* 9791-10041 mass range can be ascribed to Ole e 7. In fact, the apparent SDS-PAGE molecular mass of this allergen is 9-11 kDa (Table 5_Part 1, column 2, row 8). The predicted molecular mass of the Ole e 7 fragment from the peptide sequence is 2199 Da [88]. The only experimental values available are 9905-10302 Da obtained by low-resolution MALDI mass spectrometry (Table 5_Part 1, column 3, row 8).

Therefore, it can be suggested that peaks in the range 9791-10041 Da could be correspond to the expected Ole e 7 (Figure 7).

The ion peak at *m*/*z* 5821 (Figure 7) was attributed to the olive pollen allergen Ole e 6. This allergen has been isolated, purified and biochemically characterized,[89] and its specific cDNA was cloned and sequenced[90,91] (Table 5, row 7). Considering that the value of 5833 Da **(**Table 5_Part 1, column 3, row 7) corresponds to the predicted molecular weight from cDNA and that there are no other known allergens in this range, the observed ion peak at *m*/*z* 5821 can be ascribed to Ole e 6. Meanwhile, Ole e 2, an allergen that consists of 134 amino acids[92,93], probably corresponds to the doubly charged ion at *m*/*z* 7396, since it is known that the predicted average molecular mass is 14.4 kDa (Table 5, row 3). The mono e doubly charged ions at *m*/*z* 16300-17829 and at *m*/*z* 8837-8989, respectively, were ascribed to Ole e 1 (Table 5, row 2).

The two **a** and **b** fractions (Figure 7) show a significantly different proteic pattern; a complete pool of Ole e 1 isoforms is, really, predominant in hydrophilic fraction **4b**.

The MALDI spectrum of fraction **4s** is characterized by the presence of ion peaks corresponding to low molecular weight proteins as a consequence of the fractionation procedure which lowers the solubility of lower molecular weight proteins in aqueous ammonium bicarbonate. The MALDI spectra of the first hydro-soluble fractions of **1**, **3**, and **5-8** showed similar protein expression (§ Appendix: **A2**).

In particular, the MS spectrum of fraction **5a** (Figure 8a) shows four peaks in the mass range 9.7-10 kDa, whereas that of **3a** (Figure 8b) displays one additional peak at *m*/*z* 9186. A closer inspection of the main four peaks present in that mass range shows a difference of 70 and 180 mass units between two adjacent peaks and couples, respectively. The allergens

in the 9-10-kDa range (Table 5_Part 1, column 2) that have already been found in olive tree pollen are Ole e 3[94] and Ole e 7.

Ole e 7 consists of one polymorphic polypeptide chain of ≈10 kDa, which accounts for 88 amino acid residues, as deduced from the elution profile on HPLC and determined by mass spectrometry[88] (Table 5_Part 1, column 3, row 8). Data obtained from Edman degradation of NH_2-terminal amino acids [88] indicate that the number of isoforms of Ole e 7 is four and differs by the exchange of three amino acids at the positions 5 (S-G), 10 (L-K), and 18 (I-K), leading to a difference of -30, +15, and +15 mass units, respectively (§ **A1**, Table 5_Part 2, column 4, row 8,). Therefore, the observed difference of 70 mass units (Figure 8b) could be associated with the combined effect of different peptide sequence, to the presence of calcium adducts, or both, whereas the difference of 180 should be associated with a hexose unit. The four ion peaks in the *m*/*z* 9.7-10-kDa range can be attributed to Ole e 7 because of its known polymorphism and its very high solubility in saline solution.

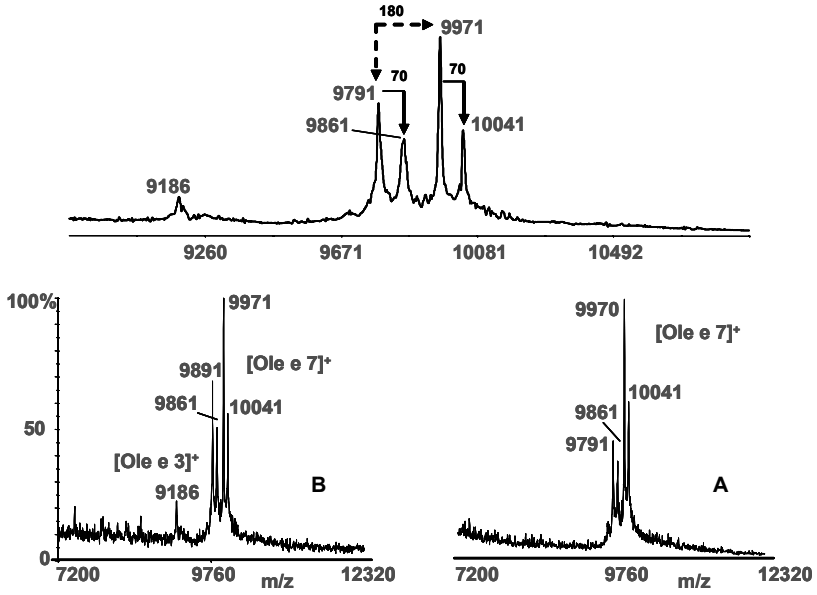


Figure 8. MALDI spectra of fractions (b) **3a** and (a) **5a**.

Additional information was obtained for cultivar **3**, whose pollen was consecutively extracted at room temperature with ammonium bicarbonate for 20, 60, and 120 min. The whole extracts **3I**, **3II**, and **3III** thus obtained were subsequently partitioned into three fractions (Chart 1) generically identified as **I-s**, **I-a**, and **I-b**, respectively.

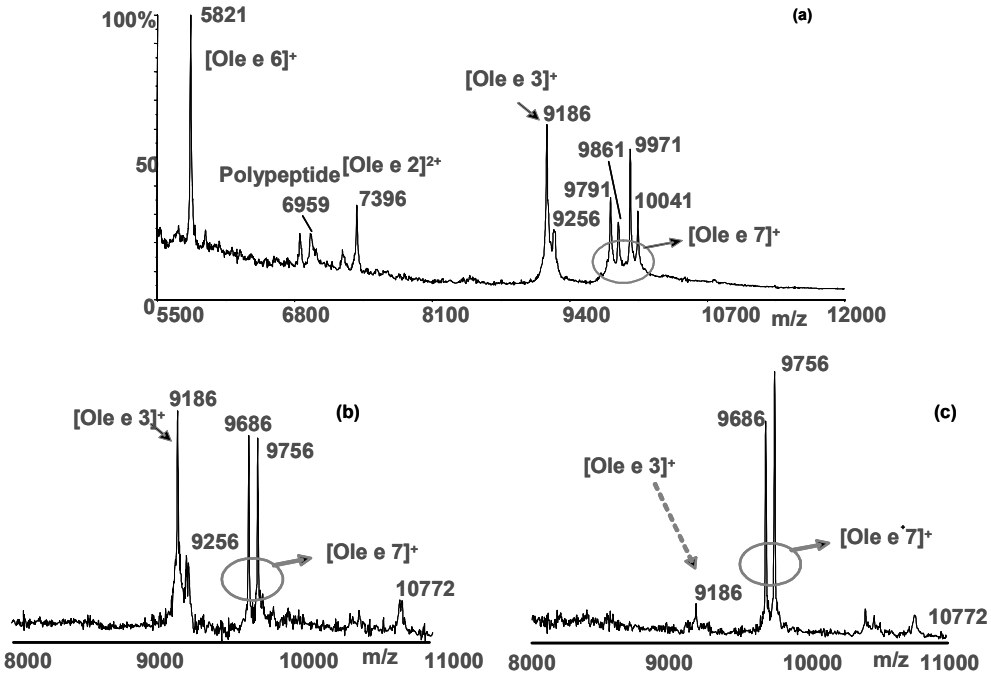


Figure 9. MALDI spectra of fractions (a) **3I-s**, (b) **3II-s**, and (c) **3III-s**.

The MALDI-MS spectrum of lipophilic fraction 3I-s, (Figure 9a) shows ion peaks at *m*/*z* 5821 and 7396, likely corresponding to the species $[Ole\ e\ 6]^+$ and $[Ole\ e\ 2]^{2+}$, respectively; those at *m*/*z* 9186 and 9256 (Δm = 70 Da), which can be attributed to $[Ole\ e\ 3]^+$; and that at *m*/*z* 9.7-10 kDa, which reasonably corresponds to $[Ole\ e\ 7]^+$.

Two other different polypeptides at *m*/*z* 9686 and 9756, observed in fractions **3II-s** and **3III-s**, respectively (Figure 9b and c), belongs to the isoform pool of Ole e 7. The antigen Ole e 3 (Table 5, row 4), known also as Poc3_Oleeu, could exist in the two cDNA-derived isoforms [94] whose 30-u difference is due to the substitution, at the positions 43 and 80, of the amino acid P with L and of V with I, respectively. The MALDI spectra of the two fractions **3I** and **3IIs** displayed two peaks, at *m*/*z* 9186 and 9256, differing by 70 u (Figure 9a and b). Taking into account the calcium-binding properties [94] of Ole e 3, the observed difference likely proves the identification in the natural matrix of both the expected isoforms, as calcium ion adducts.

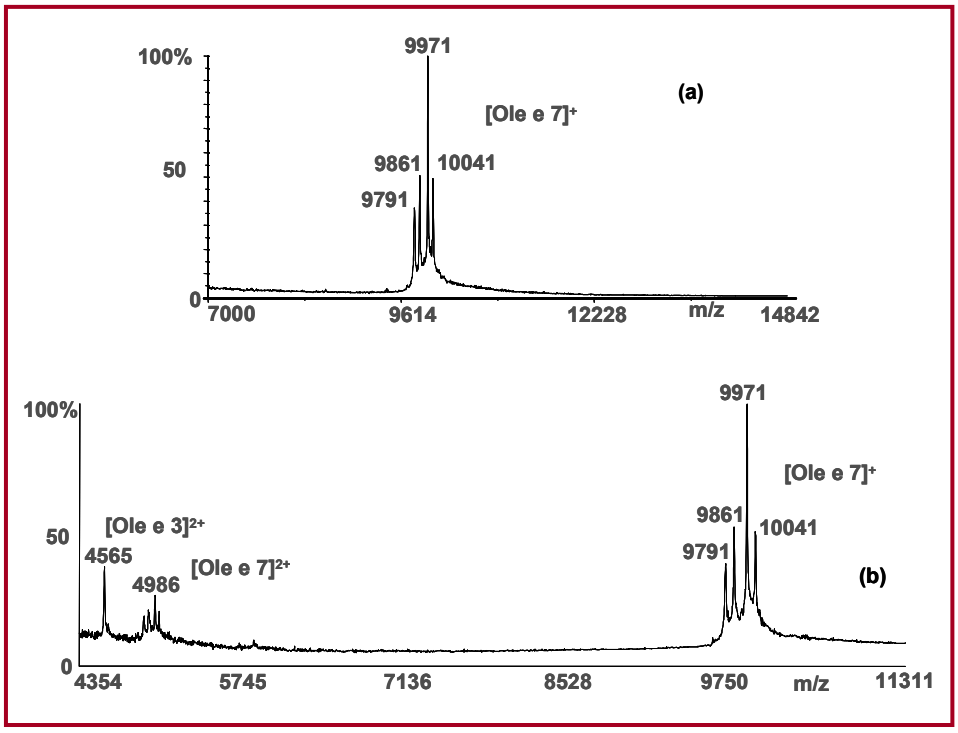


Figure 10. MALDI spectra of fractions (a) **1a** and (b) **1b**.

Ole e 3 and Ole e 7 are the only proteins identified in the whole extract of cultivar Ottobratica (**1**). MALDI spectra of fractions **1a** (Figure 10a) show in the mass range 9.7-10 kDa the typical pattern of Ole e 7, previously discussed, whereas that of fraction **1b** (Figure 10b) contains an additional peak at *m*/*z* 4565, probably corresponding to $[\text{Ole e 3}]^{2+}$.

A 50 μL portion of fraction **1b** (pH 8) was fully reduced by treatment with 10 μL of DTT (50 mM) under magnetic stirring for 6 h at room temperature. The sample was then incubated with 4 μL of PNGase F (0.5 unit/ μL) for 3 days at 37 °C. After 78 h, the resulting mixture was digested overnight with 1 μL (10 pmol/ μL) of Trypsin. The MS spectra were acquired in reflectron mode (20-keV accelerating voltage), with 400-ns delayed extraction, averaging 2000 laser shots with a mass accuracy of 50 ppm. A 1 μL portion of a premixed solution of each fraction and α-CHCA (0.3% in TFA) was spotted on the matrix target, dried at room temperature, and analyzed in the mass spectrometer.

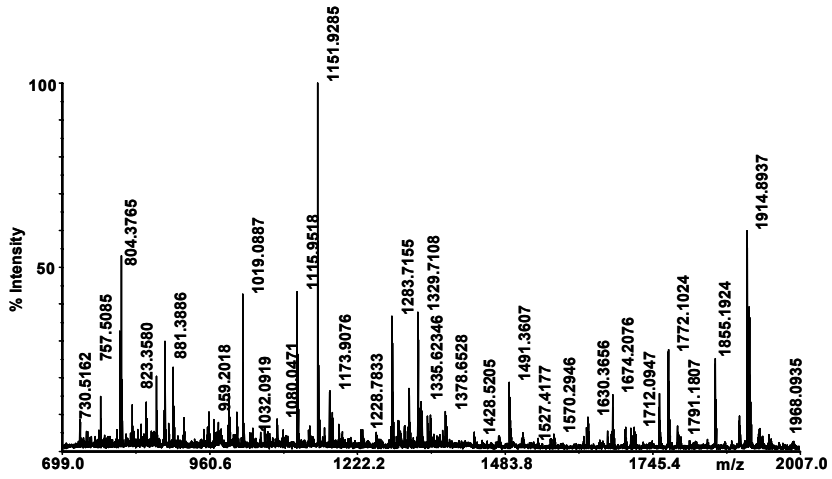


Figure 11. MALDI-MS spectra of peptide mixture of fraction **1b**.

The presence of Ole e 3 in fraction **1b** of pollen Ottobratica has been proved by a peptide mass fingerprinting (PMF) approach by tryptic digestion following deglycosylation by PNGase F. Proteins were identified by searching a comprehensive protein database using

Mascot programs (www.matrixscience.com). One/two missed cleavages per peptide was allowed, and an initial mass tolerance of 50 ppm was used in all searches.

Database searching of the MALDI reflectron spectrum of the tryptic digest (Figure 11) confirms that one component of the fraction **1b** is Polcalcin Ole e 3 (calcium-binding pollen allergen Ole e 3) (sptjO81092) with sequence coverage of 50%. The Ole e 3 identified sequences, detected by MALDI TOF, are shown in Table 6a and b.

(a)

1	**MADDPQEVAE**	**HERIFK**R**FDA**	**NGDGKISSSE**	**LGETLK**TLGS	VTPEEIQRMM
51	AEIDTDGDGF	ISFEEFTVFA	R**ANRGLVK**DV	AKIF	

(b)

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
1 – 16	1914.89	1913.88	1913.89	0.01	1	**MADDPQEVAEHERIFK**
18 – 25	823.37	822.35	822.35	0.01	0	**FDANGDGK**
18 – 36	1968.09	1967.08	1966.95	0.13	1	**FDANGDGKISSSELGETLK**
37 – 48	1329.71	1328.69	1328.69	-0.01	0	**TLGSVTPEEIQR**
72 - 78	757.50	756.49	756.46	0.03	1	**ANRGLVK**

Table 6. Ole e 3 identified sequences: (a) complete sequence of Ole e 3; (b) matched peptides.

A particular case is represented by the cultivar Carolea (2). Its fraction **2a** is characterized by only one abundant doubly charged ion peak at *m*/*z* 7230, probably corresponding to Ole e 2 (§ Appendix: **A.2)**. The antigenic profile of Coratina and Villacidro hydrosoluble **b** fractions, **5b** and **7b**, respectively, (Figure 12 a, b) shows a similar over enrichment of Ole e 1 allergens.

The fraction **7b** (Figure 12b) contains the ionic species at *m*/*z* 8896, 17810, and 35635 only, corresponding to the doubly and monocharged and dimeric forms of Ole e 1, respectively. The peaks at *m*/*z* 8277 and 16447 represent probably the doubly and the monocharged deglycosilated forms of the same allergen.

Ole e 1 is the major allergen of *Olea europaea* (Table 5, row 2) and one of the best characterized allergens of the Oleaceae family; it has been isolated, purified, and bio-

chemically characterized[95] and its specific cDNA was cloned and sequenced[96,97] (GenBank/ EMBL Data Bank accession numbers S75766 and X76395, respectively). Ole e 1 shows 85-95% identity with Lol p 11, Lig v 1, Syr v 1, Fra e 1 and Pla l 1 allergens and 36-38% identity with the deduced amino acids sequences from *LAT*52, *Zmc*13, and *OSPSG* genes from tomato anthers and maize and rice pollens, respectively.[98]

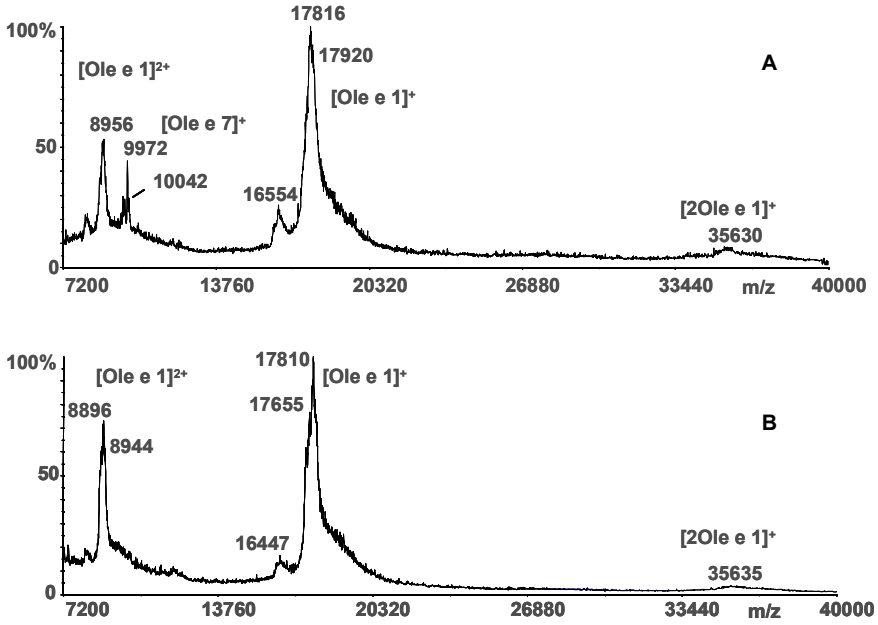


Figure 12. MALDI spectra of fractions (a) **5b** and (b) **7b**.

Fraction **7b** was used for the identification of Ole e 1 by peptide mass fingerprinting after tryptic digestion. A 50 µL portion of fraction **7b** (pH 8) was fully reduced by treatment with 10 µL of DTT (50 mM) under magnetic stirring for 6 h at room temperature. The sample was then digested overnight with 1 µL (10 pmol/*Ì*L) of Trypsin. The MALDI reflectron spectrum (Figure 13) of tryptic peptides of intact Ole e 1 displays ion peaks

corresponding to protonated tryptic peptides of the Ole e 1 family, with mass errors of 50 ppm.

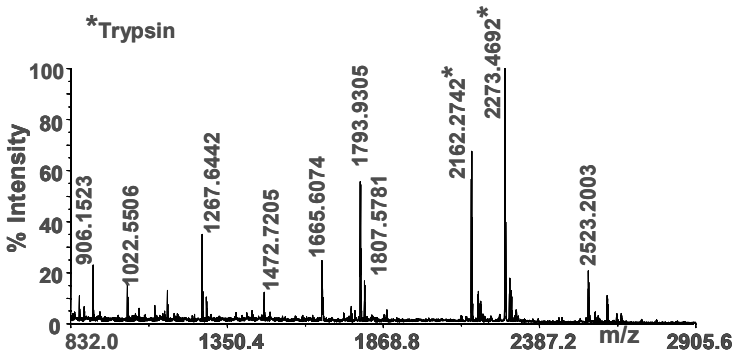


Figure 13. MALDI-MS spectra of peptide mixture of fraction **7b**.

Database searching indicates that the major components of the mixture are: (i) main olive allergen (*O. europaea*), (ii) major allergen OLE26, and (iii) major pollen allergen (allergen Ole e 1) with sequence coverages of 83, 62, and 50%, respectively (Table 7). The tryptic fragments of main olive allergen detected by MALDI-TOF are shown in Table 8.

Accession	description	mass	p.I.	Sequence coverage, %	No.mass values matched
gij13195753	main olive allergen (*Olea europaea*)	14 575	7.74	83	10
gij1362133	major allergen OLE26, common olive	15 288	7.74	62	6
gij14424429	major pollen allergen (allergen Ole e 1)	16 319	6.18	50	6

Table 7.

The fractions **2b**, **3b**, and **8b** did not show any enrichment in Ole e 1, as previously observed for **5** and **7**, whereas they contain the antigenic mixture of Ole e 1, Ole e 2, Ole e 7 and one unidentified polypeptide at *m*/*z* 6959 (Figure 14 a-c).

start-end	obs	*Mr* (exptl)	*Mr* (calcd)	Δ*Mr*	miss	Sequence
1-29	3234.45	3233.45	3233.57	-0.12	2	QVYCDTCRAGFITELSEFIPGASVRLQCK
9-25	1793.94	1792.93	1792.94	-0.01	0	AGFITELSEFIPGASVR
9-31	2523.20	2522.20	2522.32	-0.12	2	AGFITELSEFIPGASVRLQCKEK
32-45	1600.80	1599.80	1599.79	0.01	1	KNGDITFTEVGYTR
33-45	1472.72	1471.72	1471.69	0.02	0	NGDITFTEVGYTR
46-56	1267.64	1266.64	1266.63	0.01	0	AEGLYSMLVER
46-73	3156.46	3155.46	3155.53	0.07	2	AEGLYSMLVERDHKNEFCEITLISSGSK
74-89	1787.91	1786.90	1786.86	0.05	0	DCNEIPTEGWAKPSLK
90-100	1235.67	1234.66	1234.67	-0.00	0	FILNTVNGTTR
101-109	1022.55	1021.54	1021.56	-0.02	0	TVNPLGFFKK

Table 8.

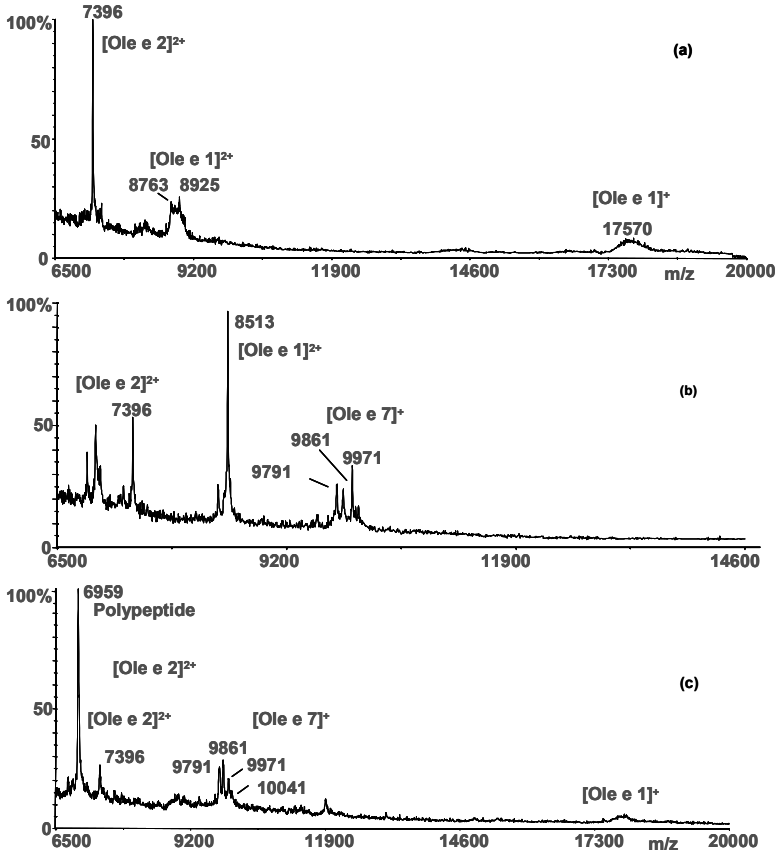


Figure 14. MALDI spectra of fractions (a) **2b**, (b) **3b**, and (c) **8b**.

The significant differences observed in the MALDI-MS spectra of **1b-8b** fractions suggest that the antigenic profile of the analyzed pollens can be divided into three distinct groups: (i) those with low Ole e 1 content, represented by Carolea (**2**), Dolce di Rossano (**3**), and Sinopolese (**8**); (ii) those characterized by the overenrichment of Ole e 1, corresponding to Cassanese (**4**), Coratina (**5**), Nocellara del Belice (**6**), and Villacidro (**7**); and (iii) that containing Ole e 3 and Ole e 7 only, which is represented by the cultivar Ottobratica (**1**). These results demonstrate that the proposed experimental procedure, can supply valuable information on the antigens' micro heterogeneity.

PROTEIN-EXPRESSION PROFILING and PMF: Structure and Function of proteases in Mastitic Milk.

Among the broad applications of proteomics to samples with origins as varied as microbial, vegetable and animal sources, many applications to a variety of nutritionally relevant proteins have also been described. Proteomic tools have permitted the characterisation of food components, the study of their functional, nutritional and biological relevance, the study of protein conformation and of protein interactions, as well as food quality estimation [2, 17] and also are useful to investigate protein heterogeneity in protein-rich foods[99]. Several proteomic techniques have been applied to the study of milk and milk products, allowing the separation of major proteins, including caseins (αs1-, αs2-, β- and κ-casein) and whey proteins (β-lactoglobulin, α-lactalbumin, bovine serum albumin). Milk proteins are characterised by a great heterogeneity and the presence of several molecular forms. Important challenges facing the use of proteomics concern the elucidation of different genetic variants, changes in the degree of phosphorylation or glycosylation and the localisation of post-translational modifications of milk proteins. Milk also contains an important number of low abundance proteins, such as lactoferrin, immunoglobulins, glycoproteins, hormones and endogenous enzymes, which may also be studied by the use of techniques with such high resolution.

Another important field of application of proteomic tools, based on 2DE, concerns protein profiling from different organisms [100]. 2DE procedures were used as high-resolution methods for the separation of milk proteins, 2DE maps of various milk products were established. Two-dimensional patterns of milk proteins of several mammals were obtained

by combining isoelectric focusing (IEF) (pH 3–10) in the first dimension and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) or urea-PAGE in the second dimension, which allowed comparison of the major milk proteins in the different species. The study of the evolution of milk protein profiles during different lactation periods using this technique revealed that casein proportions were reduced throughout the dry period and a number of peptides generated due to casein breakdown were identified[101]. Using the same pH gradient, some of the genetic variants of caseins and whey proteins and their phosphorylated forms were separated and identified

The application of mass spectrometry to the proteomic analysis of food samples constitutes a particularly important advance in milk protein analysis. The coupling of 2DE with the identification of the separated protein by MS analysis by a ‘‘bottom-up’’ approach, in which protein identification is achieved following separation by 2DE, protein digestion with trypsin and analysis by MS, offers a potent strategy for dairy research. The mass fingerprint of the peptides obtained allows comparison of a set of determined molecular mass values of a proteolytic digest against values calculated by theoretical digestion of sequences from a protein database. Using such approach the polymorphism of goat αs1-casein was confirmed using 2DE followed by matrix-assisted laser desorption/ ionisation-time of flight mass spectrometry (MALDI TOF MS)[102]. Differences in the degree of phosphorylation and some of the glycosylated forms from each casein family were identified by MALDI-TOF MS.

Another classic application of proteomic tools to milk protein analysis is the study of proteolysis, a key event that determines the textural characteristics of milk products. Milk proteins contain peptides that are precursors of flavour compounds or exhibit biological properties, which are essential for the development of dairy products with specific sensory characteristics or health-promoting features.

In our research work the proteic profiling of bovine milk produced by cows with subclinical mastitis was obtained by MALDI mass spectrometry. A simple procedure of chemical fractionation of raw milk was developed, whereby less complex protein fractions were obtained prior to mass spectrometric and SDS-PAGE analysis. The profiles of milk

proteins thus obtained could allow the identification of either early markers of the acute phase of mastitis or endogenous peptide of innate immune response. The activity of the endogenous proteases in raw milk produced from each quarter of healthy and mastitic cows was therefore assayed over 24-, 48-, 96-, and 216-h incubation at 37 °C at both physiological and acid pH. Sequence-specific peptides were identified for each fraction by MS/MS experiments, and all tandem mass spectra were evaluated using MASCOT database searching.

The proteolytic activity of the number of endogenous proteases present in milk affects the final properties of texture and flavour in dairy product.[103,104] The most important peptidase families identified in milk are plasmin[105,106] and cathepsin.[107,108] Plasmins are trypsin-like serine proteases with optimal activity at pH 7.5 and 37 °C[109] playing an important role during the ripening of cheese.[110,111,112] The cathepsin family includes cathepsin B, H, L, I, G, and D. The first four proteases are lysosomal cysteine proteases with optimal activity at pH < 7, while Cathepsin G is an alkaline serine proteases and Cathepsin D is an aspartyl protease that display the maximum activity at pH 3.2.[113]

Mastitis is an inflammatory process of the mammary gland and causes anomalies in milk such as an altered consistency[114,115] and appearance of clots. It has considerable impact both on the wellbeing of the individual animals[116] and on the dairy industry business.[117,118] This inflammation is the causative agent of an improved transferring into milk of those blood constituents, such as polymophonuclear leukocytes (PMN) that correspond to more than 90% of somatic cell counting (SSC).[119] An awful consequence of this pathology is represented by the increasing activity of the plasmin/cathepsin protease system as a function of the progression of mastitic infection.[102] Peptide mass fingerprinting (PMF) indicates that neutral and acid proteases, including Elastase and Cathepsin G, B and D, are present.[120] The level of Cathepsin D in milk is significantly correlated to SCC, although it is not clear whether this reflects increased production of Cathepsin D and/or increased activation of its precursor.[121].

As shown in vitro, caseins are very suitable substrates for proteolysis by Cathepsin B, D and G and for Elastase [115,122, 123, 124]. The proteolytic activity of native lysosomal cysteine

proteases was exploited using acid whey fraction against individual caseins followed by SDS-page.[125] Immunological analysis proved that Cathepsin B is one of the active proteases in presence of reducing agents such as dithiothreitol (DTT).[126]

However it is commonly considered that the proteolysis of low SCC milk is dominated by plasmin and cathepsine D, but the role of plasmin decreases in high SCC milk favouring the action of other proteases.

From an industrial point of view the worldwide multibillion losses related with the discharge of mastitic milk seems to be alleviated by the use of these wastes as milk replacement for animal feeding, only. An alternative solution to the problem could be offered by the use of mastitic milk as valuable renewable resource of peptides, in relation to other applications[127], or by the identification of those protein biomarkers for early diagnosis and treatment of mastitis.

Therefore, high SCC bovine milk produced by cows with subclinical mastitis was selected as case studies to investigate milk proteins hydrolysis by endogenous proteases. The activity of the endogenous proteases in raw milk produced from each quarter of healthy and mastic cows was analysed over 24, 48, 96 and 216 h incubation at 37°C in both natural and acid pH. A simple procedure of chemical fractionation of fresh and 24, 48, 96, 216 incubated milk was developed to obtain less complex fractions of peptides and proteins. All fractions were directly analyzed by SDS- PAGE, MALDI MS and MS/MS to endorse the up- and down- regulation of proteins and the alteration of peptide composition in milk during inflammation progress.

Milk from mastitic udders exhibit increased proteolytic activity, the extent of which, at least in the case of milk casein, is undoubtedly associated to the releasing of PMN from blood[87]. The extent of the breakdown of the casein in high SCC milk is a function of the activity of several enzymes and also depends on the length of the time for which the enzymes act on them. Therefore, identification of peptides produced under controlled experimental conditions by the action of PMN proteases could fill many gaps to understand patterns and pathways of proteolysis in milk.

High (**1-4**) and low (**5-8**) SCC milk produced by cows with subclinical mastitis, table 9, were selected to investigate milk proteins hydrolysis by endogenous proteolytic milk enzymes.

A simple preparation protocol was developed to distribute the amount of information stored in the proteome of each fresh and incubated raw milk sample in a set of three MALDI spectra that could be independently evaluated and matched to retrieve data for their comparison.

	SAMPLES	S. C. C. [cell/mL]
1	**F**ront **L**eft quarter **H**ealty cow	$1.1* 10^3$
2	**F**ront **R**ight quarter **H**ealty cow	$0.6* 10^3$
3	**R**ear **L**eft quarter **H**ealty cow	$1.2* 10^3$
4	**R**ear **R**ight quarter **H**ealty cow	$1.3* 10^3$
5	**F**ront **L**eft quarter **M**astitic cow	$0.4 * 10^6$
6	**F**ront **R**ight quarter **M**astitic cow	$1.2* 10^6$
7	**R**ear **L**eft quarter **M**astitic cow	$11.6 * 10^6$
8	**R**ear **R**ight quarter **M**astitic cow	$8.4 *10^6$

Table 9.

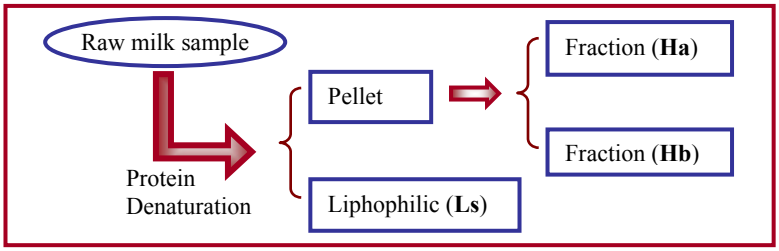


Chart 2.

The data set displayed by the three spectra provides the entire profiling of a given entity. One lypophilic (**Ls**) and two hydrophilic fractions (**Ha-Hb**), respectively, can be obtained from each sample by following a simple procedure.

50µL of fresh raw milk from each quarter taken from healthy and unhealthy cows (table 9) was precipitated with 1mL of $CHCl_3/CH_3OH$ 1:3 (v/v) yielding the supernatant fraction **S** and pellet which was partitioned consecutively, under magnetic stirring and at room temperature with (**a**) 1mL of 50mM NH_4HCO_3 and (**b**) 1mL di H_2O (TFA 2%). Each step (**a - b**) was followed by centrifugation at 11000 rpm for 2 min.

A 1µL portion of each fraction was directly analyzed by MALDI mass spectrometry using α-cyano-4-hydroxy-trans-cynnamic acid (α-CHCA, 0.3% in TFA) as matrix. MALDI-TOF analyses were performed using a 4700 Proteomics Analyzer mass spectrometer from Applied Biosystems (Foster City, CA) equipped with a 200-Hz Nd:YAG laser at 355-nm wavelength. A 0.5µL portion of a premixed solution of each fraction and α-CHCA (0.3% in TFA) was spotted on the matrix target, dried at room temperature, and analyzed in the mass spectrometer. Linear MALDI MS spectra were acquired averaging 2500 laser shots with a mass accuracy of 500 ppm in default calibration mode that was performed using the following set of standards: insulin (bovine, $[M + H]^+$ avg m/z 5734.59), apomyoglobin (horse, $[M +H]^{2+}$ avg m/z 8476.78), ($[M + H]^+$ avg m/z 16 952.56), and thioredoxin (Escherichia coli, $[M + H]^+$ avg m/z 11 674.48).

As extensive example figure 11a-c shows the spectra of the lopophilic (**Ls**) and hydrophilic (**Ha-Hb**) fractions obtained from Front Left (FL) quarter of healthy cow (1, Table 9). Accordingly, the protein profile of **1a** displayed whey proteins and casein pool: α̃lactalbumin, β-lactoglubulin, and caseins in the 5-20-kDa (Figure 15).

The ion peaks at *m*/*z* 7049, 7178, 8628, 14083 and 14150 (Figure 15a) were attributed to the mono and doubly charged forms of α-lactalbumin (α-LCA), since it is known that it consists of 123 -142 amino acids and the predicted average molecular mass is 14.2 (PRO_0000018439) -16.2 (P00711) kDa. The ion peaks at *m*/*z* 10861 – 12196 mass range (Figure 15a) can be ascribed to doubly charged forms of caseins; it is known these protein are characterized by numerous isoforms and the predicted average molecular mass is within 21-26 kDa. While the doubly and mono charged ions at *m*/*z* 9170 and *m*/*z* 18323, respectively, were attributed to β-lactoglobulin.

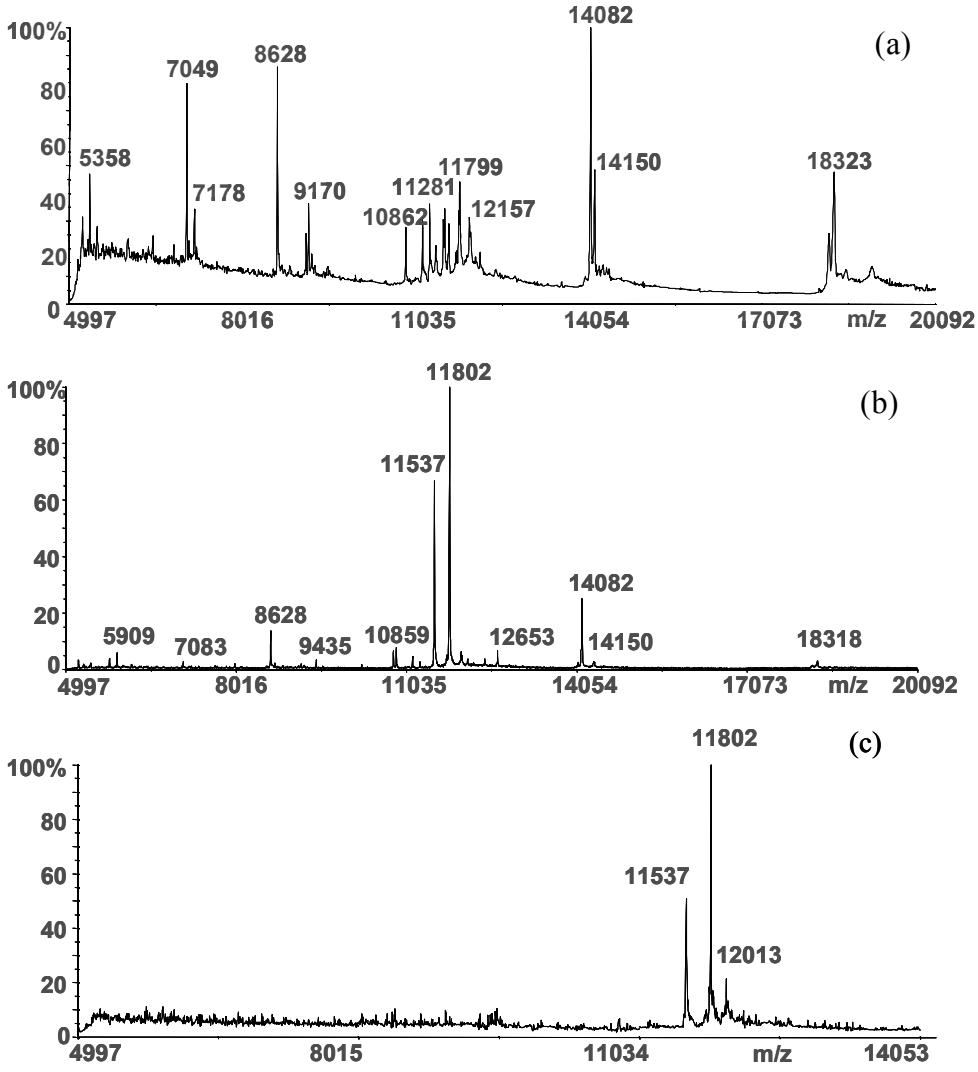


Figure 15. MALDI spectra of fractions (a) **1a**, (b) **1s** and (c) **1b**.

Similar information can be derived from the MALDI spectrum of fraction **1Ls** (Figure 15b) as a consequence of the fractionation procedure which lowers solubility of these

proteins in aqueous solution. The dominant peaks at m/z 11537 and 11802, associated to casein, suggest the fraction **1Hb** represents an enriched fraction of caseins (Figure 15c). The information obtained from linear MALDI spectra (Figure 15a – 15c) was considered as a preliminary indication of the protein composition of the investigated sample. Otherwise, the chemical fractionation approach was developed to isolate subsets of proteins allowing the detection high abundance protein useful to assess integrity and nutritional value of milk.

MALDI spectrum of fraction **7Ha** is dominated by the presence of bovine serum albumin (BSA) (Figure 16a). In fact, the ions at m/z 11374, 22270, 33446 and 67109 were associated to $[BSA]^{4+}$, $[BSA]^{3+}$, $[BSA]^{2+}$ and $[BSA]^{+}$, respectively. The formation of multiply charged BSA ions was considered as a direct consequence of sample preparation.

Dried-droplet preparations with α-CHCA matrix, promotes the formation of doubly and triply charged protein ions as well as protein dimeric ions[128]. The innovative results showed by MALDI spectrum of **7Ha** (Figure 16a) is the presence of a blood protein. Otherwise, **7Ha** is the ammonium bicarbonate fraction derived from mastic raw milk characterized by the most high somatic cell count (Table 9, Figure 16a).

The presence of BSA in this fraction suggests an increased permeability of the blood mammary barrier which is a clear indication of a high degree of quarter inflammation. BSA can be regarded as functional proteineous indicators for acute phase of mammary gland inflammation because, only in this state, serum compounds appear in milk.

Comparison of protein profiling from fraction **Ha,** derived from raw milk of each quarter, is performed in order to find other biomarkers reflecting the kinetics of event during inflammation processes. MALDI spectrum of fraction **6Ha** (Figure 16b) is characterized by the presence of numerous peptides within (3-5) kDa which denotes the occurrence of proteolytic activity also in fresh mastitic milk. While the ion peaks at m/z 7089, 14180 and 8718, associated respectively to lactalbumin and lactoglobulin, represents only a modest percentage of the total ion current. Since the analysis is performed by a desorption / ionization method that could affect the ion peaks intensity, there isn't a direct correlation among data obtained and down regulation synthesis of whey proteins.

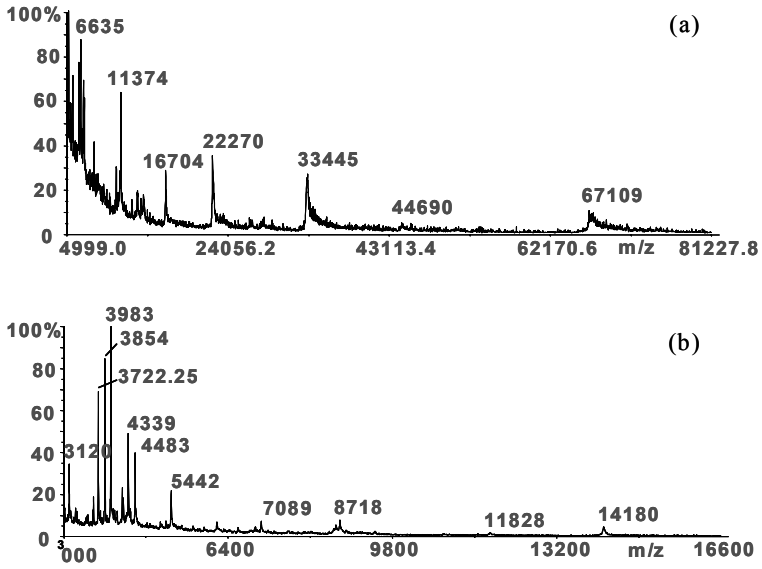


Figure 16. MALDI spectra of fractions (a) **7a** and (b) **6a**.

The MALDI spectra of the lipophilic fractions of **5Ls- 8Ls** showed similar protein expression. In particular, the MS spectrum of fraction **6Ls** (Figure 17) shows numerous polypeptides within (1-5) kDa and ion peaks associated to well known whey proteins.

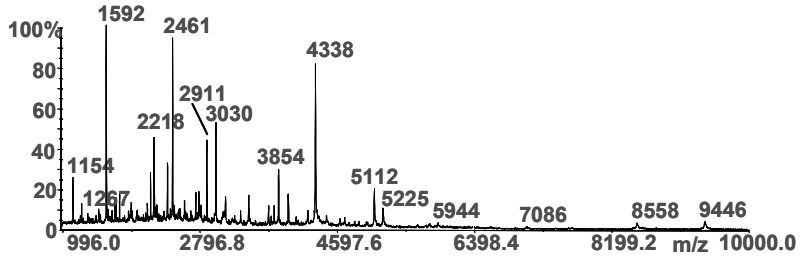


Figure 17. linear MALDI spectrum of fraction **6Ls**.

The novelty of this data is the presence of a polypeptide at m/z 4338 which can be associated to β-defensin. The determined accurate molecular mass (m/z 4339.2275) was used for a simple database search.

The identified β-defensin is an endogenous cationic peptide with well know antimicrobial and bactericidal activity against *E. coli* and *S. aureus* and it is generated by the innate immune response.

SDS-PAGE was performed on each sample in order to compare the data set displayed by the linear MALDI spectra and the protein electrophoretic profile. 400 µL from each chemical fraction, were lyophilized, than the pellets were dissolved in 30µL of solubilization buffer (Hepes pH 7.5 50mM, NaCl 150mM, $MgCl_2$ 1.5mM, EGTA 1mM, Glycerolo 10%, Triton 100 1%).

SDS-PAGE (20%), 18x16 cm, was performed using SE 600 Ruby^{TM} with electrophoresis condition 300V, 60mA, 18W for 6h. The resolving buffer Tris HCl 1.5M pH 8.8, SDS 0.4%, the stacking buffer Tris HCl 0.5M pH 8.3, SDS 0.4% and the Sigma Marker^{TM} Wide Range (M.W: 6.5-205 kDa) were employed for the analysis. Gels were stained with a solution of Comassie Brillant Blue R 0.1% w/v (Sigma-Aldrich) in $H_2O/CH_3OH/CH_3OOH$ (9:9:2) and distained in a CH_3COOH(10%) solution (H_2O/CH_3OH 1:1, v/v).

The bands were identified by their molecular weights, in comparison with the standard protein markers (Figure 18, lane M). The electrophoretic profile resulting from fresh raw milk **1-8** (Figure 18, lane I-IV for sample 1-4 and lane V-VIII for sample 5-8) shows the real over-expression of BSA (protein bands of about 66 kDa), the down-regulation of whey protein (protein bands in the range of 14-20 kDa) and the partial hydrolyzed casein (protein bands in the range of 20-24 kDa). Otherwise, proteolytic activity of endogenous origin was not detect in control samples (Figure 18, lane V-VIII).

The evident difference among the two methods is represented by the specificity of the mass spectrometry method, since it allows a preliminary identification of the proteins all the way through the determination of their molecular weights.

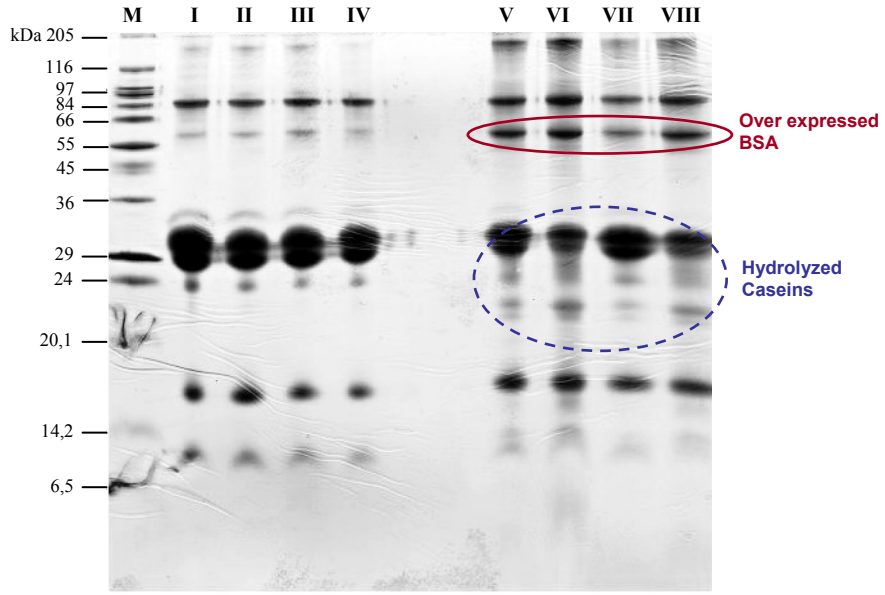


Figure 18. SDS-PAGE (20% separating gel) of fresh raw milk 1-8; lanes **I-IV** for samples 1-4 and lanes **V-VIII** for samples 5-8.

However, the alteration temperature and/ or pH perturb the complex systems constituents/enzymes, inducing the disruption of milk somatic cells and improving the expected proteolytic activity towards caseins. Therefore proteases associated to the PMN, serine, cysteine and aspartic proteases, can be activated by a simple alteration of natural pH of the samples. The pH of all mastitic (**5-8**) and healthy (**1-4**) samples was adjusted to8.0 and 6.5, to activate trypsin- like proteases and or cysteine and aspartic proteases, respectively. Therefore, the proteolytic activity of proteases associated with SC and leukocytes was assayed at pH 8 and pH6.5; in the first one 1mL of milk was incubated at natural conditions, while in the second one 1mL of milk was treated with 10 µl of 3N HCl and 300 µl acetate buffer (CH_3COOH/CH_3COO-). All samples were incubated for 24, 48, 96 and 216 h at 37 °C, than fractionated as reported above (Chart 2) and monitored by mass spectrometry and SDS-PAGE.

On the grounds of the experiments performed on fresh milk, the expected enhancement of the proteolytic activity should be better shown in the fractions **Ls**, **Ha**, and **Hb**, through the appearance of more detectable sequence-specific peptides.

The electrophoretic profile resulting from all fractions of samples **3** and **6** after 216h of incubation in both natural and acid pH (Figure 19, lane I-III for fractions **6s**, **6a** and **6b** at pH 8, lane IV-VI for fractions **6s**, **6a** and **6b** at pH 6.5, lane VII-IX for sample **3s**, **3a** and **3b** at pH 8, lane X-XII for sample **3s**, **3a** and **3b** at pH 6.5) shows the absence of proteolytic activity by endogenous enzyme towards caseins in both pH condition for control samples (Figure 19, lane 7-12 for sample **3**).

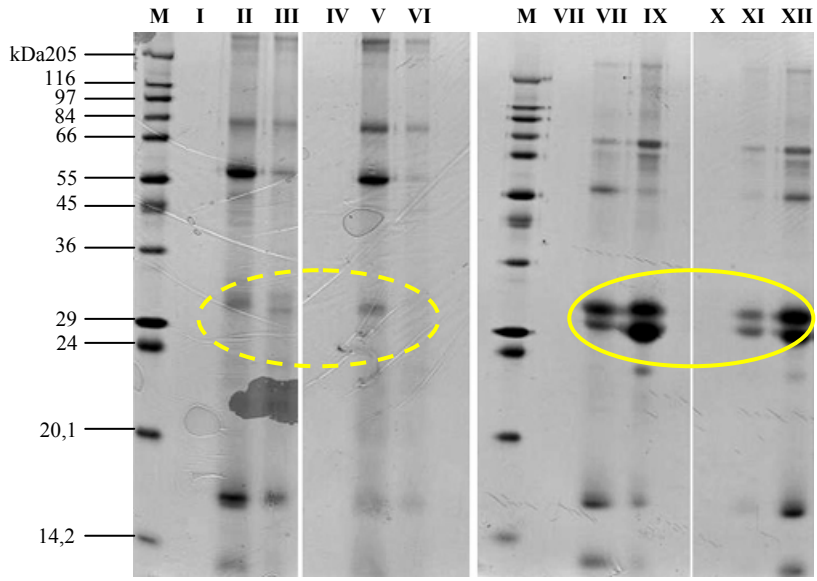


Figure 19.
SDS-PAGE (20% separating gel) from all fractions of samples 3 and 6 after 216 h of incubation in both natural and acid pH: lane I-III for fractions 6Ls, 6Ha, and 6Hb at pH 8; lanes IV-VI for fractions 6Ls, 6Ha, and 6Hb at pH 6.5; lanes VII-IX for samples 3Ls, 3Ha, and 3Hb at pH 8; lanes X-XII for sample 3Ls, 3Ha, and 3Hb at pH 6.5.

While, the action of acid and basic endogenous enzyme is revealed by the disappearance of protein bands of about 24 – 29 kDa associated to caseins (Figure 19, lane I-VI for sample **6**). The appearance of protein bands of about 14-20 kDa for both samples (Figure 19)

indicate that whey proteins were not amenable substrates for SC enzymes. Furthermore, electrophoretic profile resulting from the action of trypsin- like proteases (pH 8, 216h) and cysteine /aspartic proteases (pH 6.5, 216h) towards caseins indicates that the enzymatic activity is equivalent at both experimental conditions over 216h.

The fractions Ls, Ha, and Hb of the high SCC milk sample 8 were chosen as model substrates to determine the peptide composition after 216 h, i.e., after the casein pool was completely digested, at two selected experimental conditions of pH 6.5 and 8.0, respectively. The fractions were chosen to provide complementary information, since 8Ls should include lipophilic or ion pair species, whereas acid and basic peptides should be better found in the fractions 8Ha and 8Hb (Figure 20 and 21).

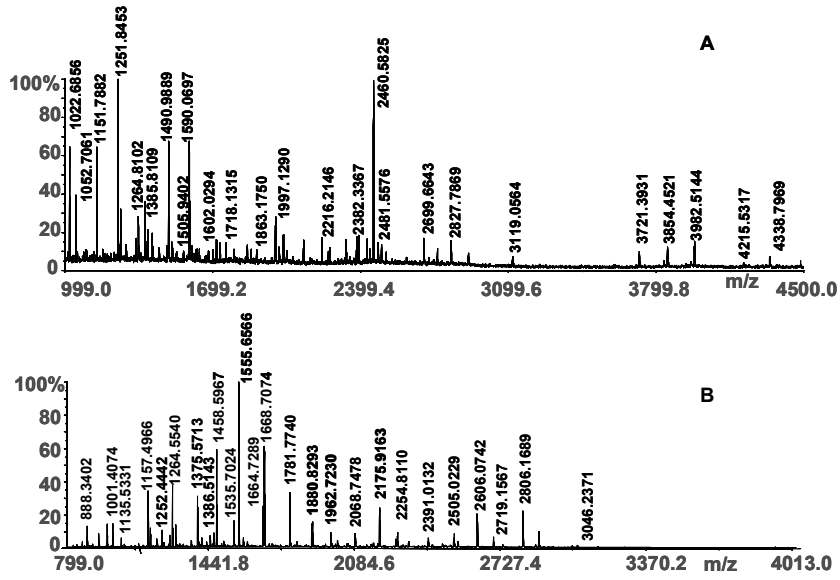


Figure 20. MALDI MS spectra of fractions **8Ls** at pH 8 (a) and **8Ls** at ph 6.5 (b).

The MS spectra were acquired in reflectron mode (20-keV accelerating voltage), with 400-ns delayed extraction, averaging 2000 laser shots with a mass accuracy of 50 ppm.

MALDI spectra of fraction **8Ls** at pH 8 and at pH 6.5 (figure 20a and 20b, respectively) show numerous peptides from 1-4 kDa. The different peptide sets cannot be used for peptide mass fingerprint, since each fraction is originated by an enzymatic digestion performed by different protease families on a complex mixture of proteins. Similar results were obtained with the other two fractions, which display quite a number of peptides.

Therefore, peptides profile can not directly correlated to enzymatic efficiency of each proteases and to degree of inflammation, since it is obvious that different lysosomal enzyme were activated at different pH condition and it is known that some peptides ionize more efficiently than others using MALDI. Moreover, the missing specific peptides of some kind of caseins does not exclude that those are substrates susceptible to hydrolysed.

Peptide groups detected from both fractions **8Ls** and **8Ha** (Figure 20 and 21) are, really, very different confirming protein data distribution.

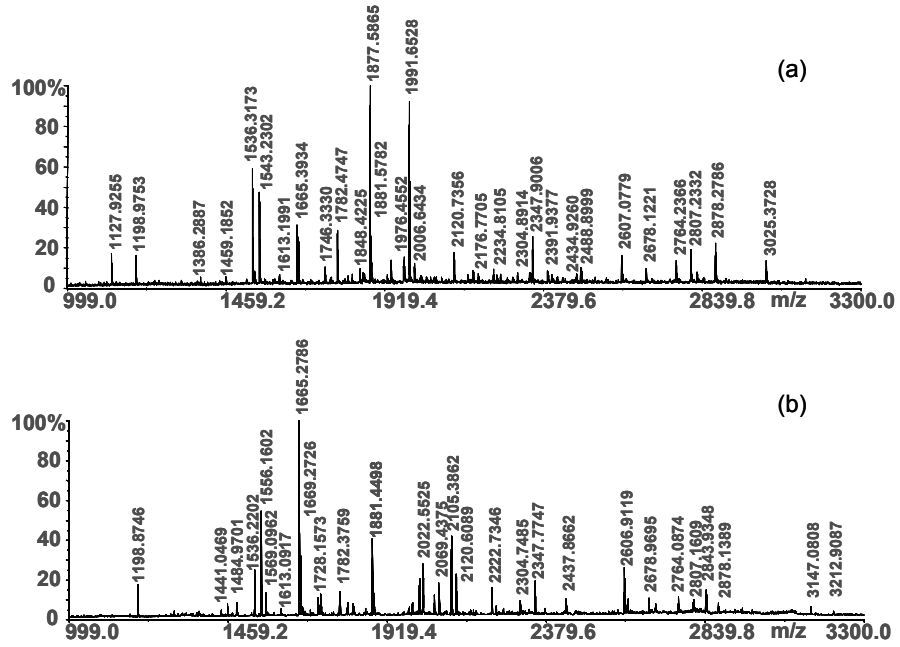


Figure 21. MALDI MS spectra of fractions **8Ha** at pH 8 (a) and **8Ha** at pH 6.5 (b).

Sequence specific peptides were identified for each fraction by MS/MS experiments, selecting the m/z value with a significant signal-noise ratio. MS/MS experiments were performed at collision energy of 1-2 kV, defined by the potential difference between the source acceleration voltage (8 kV) and the floating collision cell (7-6 kV); 3000 laser shots were averaged, while the pressure inside the collision cell was $8x10^{-7}$ Torr.

All tandem mass spectra were evaluated using MASCOT data base searching. Searches were performed against the SWISS PROT data base, with the taxonomy restricted to Bos Taurus, no enzyme cleavage specificity and an initial mass tolerance of 50 ppm.

The results of these searches are shown in Table 9 and 10 (§ Appendix: **A.4**); while figure 22 shows the tandem mass spectrum and interpretation for the ion peak at m/z 1781.77 deriving from fraction **8s** at pH 6.5 (§ Appendix: **A.5**). This ion peak corresponds to β-casein precursor (P02666) sequence fragment 208-223 (YQEPVLGPVRGPFPII; table 10, row 11).

		Protein name	Acc. No.	Peptides sequence	Span	$[M+H]^+$
1	**Ls**	β-casein precursor	P02666	SLSQSKVLPVPQKAVPYPQRDMPIQA	179-204	2877.20
2	**Ls**	β-casein precursor	P02666	SLSQSKVLPVPQKAVPYPQRDMPIQ	179-203	2806.16
3	**Ls**	β-casein precursor	P02666	LSQSKVLPVPQKAVPYPQRDMPIQ	180-203	2719.49
4	**Ls**	β-casein precursor	P02666	SQSKVLPVPQKAVPYPQRDMPIQ	181-203	2606.07
5	**Ls**	β-casein precursor	P02666	SKVLPVPQKAVPYPQRDMPIQ	183-203	2391.31
6	**Ls**	β-casein precursor	P02666	VLPVPQKAVPYPQRDMPIQ	185-203	2175.91
7	**Ls**	β-casein precursor	P02666	YQEPVLGPVR	208-217	1157.50
8	**Ls**	β-casein precursor	P02666	YQEPVLGPVRGPF	208-220	1458.60
9	**Ls**	β-casein precursor	P02666	YQEPVLGPVRGPFP	208-221	1555.65
10	**Ls**	β-casein precursor	P02666	YQEPVLGPVRGPFPI	208-222	1668.75
11	**Ls**	β-casein precursor	P02666	YQEPVLGPVRGPFPII	208-223	1781.77
12	**Ls**	β-casein precursor	P02666	YQEPVLGPVRGPFPIIV	208-224	1880.82
13	**Ls**	β-casein precursor	P02666	EPVLGPVRGPFP	210-221	1264.55
14	**Ls**	α-S2-casein	P02663	YQGPIVLNPWDQVKRN	115-130	1927.02
15	**Hb**	α-S1-casein	P02662	FVAPFPEVFGKE KV	39-52	1593.80
16	**Ha**	α-S1-casein	P02662	RPKHP IKHQGLPQE	16-29	1664.84
17	**Ls**	α-S1-casein	P02662	FRQFYQ	165-170	888.40
18	**Hb**	k-casein precursor	P42155	TMARHPHPHLSF	115-126	1430.65

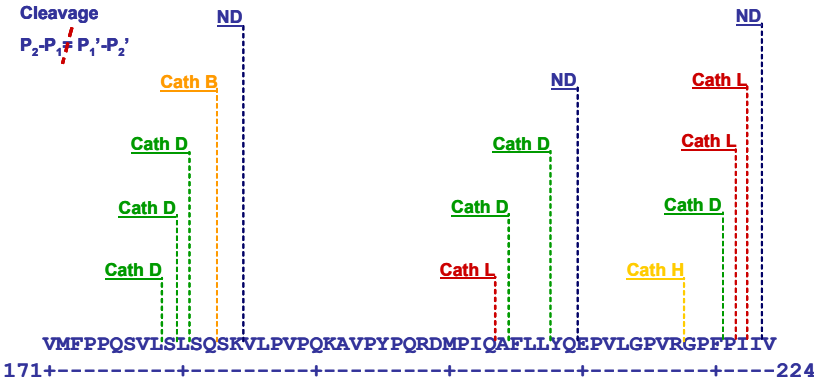
Table 10. Peptides from fractions **Ls, Ha** and **Hb** (pH 6.5) identified by Database search.

Using this strategy, 13, 1, 3 and 1 peptides of β-casein precursor (P02666), α-S2-casein (P02663), α-S1-casein (P02662) and κ-casein (P02668) respectively, were identified from

fractions **8s, 8a** and **8b** at pH 6.5 (Table 10). It is evident from Table 10 that the endogenous proteases have a extensive specificity on β-caseins precursor. The identified peptides of β-casein precursor (P02666) indicate that the cleavage sites are located towards the C-terminus of the protein suggesting that this portion of the protein is a well accessible substrate.

The observed cleavage sites L_{178}-S_{179}, S_{179}-L_{180}, L_{180}-S_{181}, Q_{182}-S_{183}, K_{183}-V_{185}, Q_{203}-A_{204}, A_{204}-F_{205}, L_{207}-Y_{208}, Q_{209}-E_{210}, R_{217}-G_{218}, F_{220}-P_{221}, P_{221}-I_{222}, I_{222}-I_{223} and I_{223}-V_{224} cannot be related to the action of only one specific enzyme. Acid proteases presumable activated, under the adopted experimental condition, would be Cathepsin B, D, H, L and I.

Cathepsin D is a typical mammalian aspartic protease. It is generated auto-catalytically from a precursor, procathepsin D, to pseudocathepsin D and thence by thiol proteases to a number of mature forms. The enzymatic specificity of this protease on β-casein is similar to that of chymosin, but narrower than that of Pepsin A. Otherwise, Cathepsin B, endopeptidase, is a thiol protease and shows pepdyldipeptidase activity liberating C-terminal dipeptides from small molecule substrates. Hence, protonated peptides at m/z 2877.20 (table 10, row 1) 1458.60 (table 10, row 8) and 1880.82 (table 10, row 12) arise from the action of Cathepsin D, while those at m/z 1668.75 (Table 10, row 10) and 1555.65 (Table 10, row 9) are ascribable to the synergic action of both Cathepsins (scheme 1).



Scheme 1. cleavage sites identified on β-casein (P02666) at pH 6.5; ND: non determined.

The Cathepsins L and H belong to cysteine-peptidase family. Cathepsins L has cleavage specificity close to that of Papain, while Cathepsins H remarkable cleaves Arg-/-Xaa bonds. Therefore, the species $[M+H]^+$ at m/z 2806.16 (table 10, row 2), 2719.49 (table 10, row 3) 2606.07 (table 10, row 4) and 1157.50 (table 10, row 7) are attributable to combined action of Cathepsins D/L, D/L, B/L and D/H, respectively (scheme 1).

The high specificity, as well as, the synergic action of Cathepsin B and D towards α-s1 and αs2 caseins is confirmed by the formation of the ion sequences at m/z 1927.02, 1593.80, 1664.84 and 888.40 (table 10, rows 14-17). Acid proteases have a limited action towards k-casein, in fact it is observed only one ion sequence of this protein; furthermore the cleavage sites T_{114}-T_{115} and F_{125}-M_{126} are not directly related to Cathepsins (§ Appendix: **A.6** Scheme 2 and **A.7** Scheme 3).

The endogenous proteases have a wide specificity on β-caseins precursor, also at basic pH conditions (§Appendix: **A.4** Table 10). Presumable activated serine proteases would be Plasmin and Cathepsin G. The first one is a trypsin like protease with specificity restricted to peptides bonds of the type Lys- Xaa to a lesser extent Arg-Xaa [107],while the second one is a neutral serine protease which shows a catalytic activity closer to Chymotrypsin.

Cathepsin G potential cleavage sites for protein sequence P02666 were predicted by PeptideCutter (www.expasy.org) considering a Chymotrypsin like activity (§ Appendix: **A.8** Scheme 4). Therefore protonated peptides at m/z 2106.92 (§ **A.4,** table 10, row 2), 1993.85 (§ **A.4,**table 10, row 3), 1881.18 (§ **A.4,**table 10, row 5) and 3024.79 (§ **A.4,**table 10, row 7) are ascribed to proteolytic action of this enzyme.

The cleavage sites P_{111}-V_{112}, P_{221}-I_{222} and I_{223}-V_{224} can be associated to an enzyme able to cleave sites with bulky residues in position P1', under basic conditions, specifying the site of cleavage like $\mathbf{P_2\text{-}P_1 \neq P_1'\text{-}P_2'}$. This specific action can be performed by Thermolysin, which is a thermostable extracellular metalloendopeptidase from Micrococcus caseolyticus containing four calcium ions. Micrococci enter milk from such sources as the udder of cows, dairy utensils, milking machines, air and dust. These gram-positive bacteria have a incontestable effect on flavour development in cheese, and this may be linked to their proteolytic, peptidolytic and esterolytic activity. Therefore the ion

sequences at m/z 1782.10 (§ **A.4,** table 10, raw 4) and 1038.68 (§ **A.4,** table 10, raw 6) derive from the action of Thermolysin alone or in synergy with Cathepsin G.

MAPPING OF PROTEIN MODIFICATIONS:
Ole e 1 micro-heterogeneity.

A proteome is not a static entity and it is not the simple product of the direct translation of gene information into protein sequences. Therefore, it is fundamental understanding the challenges caused by the different post-translational mechanisms, that process and modify proteomes permanently or reversibly. Among current methods used for the analysis of the products of these mechanisms, MS is an essential component of virtually every current strategy.

All the relevant properties of proteins are predisposed by the gene sequence and, therefore, could be precisely predicted. Proteome analysis is based on the expectation that the information gained by direct protein analysis exceeds or complements that obtained by the more readily available methods for gene sequence analysis. In addition to the sequence and abundance, the properties of proteins that are of particular interest to biochemists include their sub-cellular location, their state of modification, their function, their state of activity and the nature of interacting proteins. It is not obvious how these diverse properties can be determined systematically and quantitatively for a single protein. Extending these measurements to a proteome wide scale is even more challenging and, despite recent advances, remains largely unachievable with current methods.

Most cellular functions are not performed by individual proteins but rather by protein assemblies, also termed multi-protein complexes. It is therefore frequently assumed that proteins that specifically interact also play a part in the same function, and the identification

of specifically interacting proteins is an important component of the proteomics quest to study the function of biological processes.

Post-translational modifications (PTMs) are the chemical modifications of a protein after its translation. It is one of the later steps in protein biosynthesis for many proteins. During protein synthesis, 20 different amino acids can be incorporated in proteins. After translation, the posttranslational modification of amino acids extends the range of functions of the protein by attaching to it other biochemical functional groups such as acetate, phosphate, various lipids and carbohydrates, by changing the chemical nature of a single amino acid or by making structural changes, like the formation of disulfide bridges. Also, enzymes may remove amino acids from the amino end of a protein, or cut the peptide chain in the middle. For instance, the peptide hormone insulin is cut twice after disulfide bonds are formed, and a pro-peptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds. Other modifications, like phosphorylation, are part of common mechanisms for controlling the behavior of a protein, for instance activating or inactivating an enzyme.

Mass Spectrometry is the ideal method of choice for the detection and identification of post-translational modifications (PTMs). In principle, the methods used for protein identification are also applicable to the analysis of PTMs. For a number of reasons, PTMs analysis is however significantly more complex than simple protein identification:

(i) Proteins are frequently modified to a low stoichiometry only. Therefore, a high sensitivity of detection for the modified peptides is required.

(ii) While proteins can be identified by the sequence or the CID spectrum of a single peptide, the identification of PTMs requires the isolation and analysis of the specific peptide that contains the modified residue/s.

(iii) The bond between the PTMs and the peptide is frequently labile. It may therefore be difficult to find conditions that maintain the peptide in its modified state during sample work up and ionization.

(iv) Numerous different types of protein modifications have been described. The total sequence space containing all the potential modified protein sequences is therefore enormous.

For these reasons sometimes is necessary combining mass spectrometry with other techniques to characterize PTMs. The two mainly diffuse posttranslational modifications are *Phosphorylation* and *Glycosylation*.

Phosphorylation is the addition of a phosphate (PO_4) group to a protein or to a small molecule. There are thousands of distinct phosphorylation sites in a given cell since there are thousands of different kinds of proteins in any particular cell and phosphorylation often occurs on multiple distinct sites on a given protein. Phosphorylation of any site on a given protein can change the function or the localization of that protein, therefore the understanding of cell "state" requires knowing the phosphorylation state of its proteins. For example, if amino acid Serine-473 ("S473") in the protein AKT is phosphorylated AKT is generally functionally active as a kinase. If not, it is an inactive kinase.

Glycosylation is the process or result of addition of saccharide to proteins and lipids. The process is one of four principal co-translational and post-translational modification steps in the synthesis of membrane and secreted proteins and the majority of proteins synthesized in the rough *Endoplasmic reticulum* undergo glycosylation. Two types of glycosylation exist: *N*-linked glycosylation to the amide nitrogen of asparagine side chains and *O*-linked glycosylation to the hydroxy oxygen of serine and threonine side chains.

The *N*-linked glycosylation process occurs in eukaryotes and widely in archaea, but very rarely in bacteria. There are two major types of *N*-linked saccharides: high-mannose oligosaccharides, and complex oligosaccharides. High-mannose is, in essence, just two *N*-acetylglucosamines with many mannose residues, while complex oligosaccharides are so named because they can contain almost any number of the other types of saccharides.

Ole e 1 (accession number P19963) is the major allergen of *Olea Europaea* and one of the best-known allergens of Oleaceae family. This is a pollen specific protein playing an important role in pollen hydration and / or germination[129]. The primary structure of Ole e 1 has been determined by amino acid [75] and cDNA sequencing [99]. It is a polymorphic

glycoprotein of 145 amino acid residues and shows two main variants differing to each other in the glycosylation state. Both glycosylated and non-glycosylated variants, High Molecular Weight (HMW) and Low Molecular Weight (LMW), are recognized by the sera of hypersensitive individuals[79,130]. The HMW variant contains only one potential N-glycosylation site at Asn_{111} according to SWISS-PROT database (www.expasy.org).

The major N-glycan structures present in samples of *Ole e 1* from Spanish Cultivars were suggested by comparative approach among HPAEC-PAD and ^{1}HNMR experiments. This study concluded that Ole e 1 contains mainly high mannose-type ($Man_7GlcNAc_2$) and xylosylated complex-type ($GlcNAcMan_3XylGlcNAc_2$)[131, 132].

Ole e 1 is a single peptide chain, characterized by an extensive microheterogenity in 37 positions [97]. This microheterogenity arises from the presence of genetic variants and post-translational modifications [97, 99, 133]. At least 7 isoallergen forms of Ole e 1 have been described; however no method to properly resolve individual variants has been reported.

Moreover, Ole e 1 displays several interesting challenges with multiple isoforms and variants, that are present at both high and low abundance.

Mass spectrometry is a rapid and sensitive method for characterization of the primary structure of proteins and their posttranslational modifications including glycosylation, which plays an important role in the biological properties of many proteins[134]. The conventional methods for proteome analysis involve two dimensional gel (2D) as the final step of purification[135], followed by overnight in-gel enzyme digestion and mass spectrometry analysis[136]. Nevertheless, it has been well documented the limited separation capability of 2D-gel to resolve particular classes of proteins[137].

Other approaches such as immunoprecipitation, protein complexes[138] and protein profiling experiments lead to a protein mixture. Hence, the trend in proteomic is to work with protein mixtures limiting the protein purification steps prior to analysis. For these reasons our research group have developed a simple procedure of chemical fractionation[58,59] of the whole saline proteome extract of natural matrixes which leads to a sample containing only one protein family and/or chemically homogeneous proteins. In this report we present a new approach for the resolution of multiple forms of Ole e 1 from

whole antigen extract and identification of its structural modifications using MALDI MS and MS/MS.

The high solubility in aqueous medium is an important prerequisite for allergen candidates; their allergenicity is linked either to the concentration either to a rapid release from airborne [87]. *Ole e 1* is the major allergen of *Olea Europaea* and it is characterized by a very high solubility in saline solution[139]. According to the know hydrophilic and acid nature of this pollen antigens, a fraction of *Ole e 1* family was obtained by partition of the whole Villacidro's proteome extract (Chart 1). A 1 μL portion of each sample was directly analyzed by linear MALDI mass spectrometry using α-cyano-4-hydroxy-*trans*-cynnamic acid (α-CHCA, 0.3% in TFA) as matrix. MALDI-TOF analyses were performed using a 4700 Proteomics Analyzer mass spectrometer from Applied Biosystems (Foster City, CA) equipped with a 200-Hz Nd:YAG laser at 355-nm wavelength. The protein content of the fraction was monitored by linear MALDI MS.

The linear MALDI MS shows the formation of the ion species at m/z 8896, and 17810 corresponding to doubly and monocharged form of required antigen, respectively, and confirms the enrichment of *Ole e 1*. Whereas the ion peacks at m/z 8277 and 16447 represent probably the double and mono-charged low molecular weight form of the same allergen (Figure 23b).

The protein profiling displayed by linear MALDI experiments is equivalent to the data obtained by SDS-PAGE chromatography (Figure 23a). The obvious difference between the two methods is represented by the specificity of the mass spectrometric method which allows a preliminary identification of the proteins pool through the determination of the molecular weight of each compound.

The electrophoretic profile resulting from this sample shows only two protein bands within 14 ÷20 kDa (Figure 23a, lane 1) corresponding to LMW and HMW forms of Ole e 1, respectively. The protein bands were identified by their molecular weights, in comparison with the standard protein markers (Figure 23a, lane M).

According to MALDI experiments, SDS-PAGE protein bands intensity suggests that the HMW form of Ole e 1 (protein band ~ 19kDa) is over expressed.

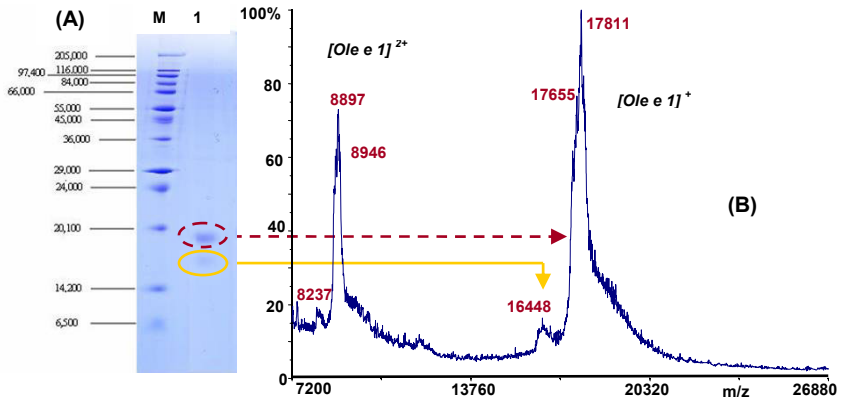


Figure 23. **(A)** SDS-PAGE (lane **M**, marker, lane **1** Ole e 1 pure fraction) **(B)**Linear MALDI Ole e 1 pure fraction.

The complete characterization of Ole e 1 requires: i) microheterogeneity determination in the natural matrix, ii) glycosylated isoallergens individuation and iii) glycans characterization. Therefore, it is necessary detect the glycopeptides, identify amino acids sequence and then characterize carbohydrate moieties. The glycopeptides can be analyzed by mass spectrometry without clean-up procedure. Peptides backbone generally has higher ionization efficiency than carbohydrates, producing sharp mass spectral signals and providing straightforward accurate mass measurements[140]. The conventional approach to establish the specific isoforms and the definite glycosylation pattern requires the isoforms separation by HPLC/2D and the glycopeptides analysis on the digested isolate isoforms[141,142].

Our chemical fractionation procedure is an alternative approach, since the hydrophilic and acid nature of the observed antigen limit separation capability of 2D-gel.

This procedure can be considered a 1-D liquid phase separation technique, which resolves the proteins by hydrophylicity/hydrophobicity and leads to a pure fraction of Ole e 1 that is ideal for mass spectrometry identification of the intact allergen[149] and for the enzymatic hydrolysis of the observed protein.

50 µL of the pure sample of *Ole e 1* (pH 8) was fully reduced by treatment with 10 µL of DTT (50 mM) under magnetic stirring for 1 h at 57 °C. The sample was digested overnight with 1 µL (4 pmol/µL) of trypsin and than analysed by MALDI TOF/TOF MS.

The MS spectra were acquired in reflectron mode (20-keV accelerating voltage), with 400-ns delayed extraction, averaging 2000 laser shots with a mass accuracy of 50 ppm. MS/MS experiments were performed at collision energy of 1kV (low energy CID) or 2kV (high energy CID), defined by the potential difference between the source acceleration voltage (8 kV) and the floating collision cell (7-6 kV); 3000 laser shots were averaged, while the pressure inside the collision cell was $8x10^{-7}$ Torr.

A 1 µL portion of a premixed solution, of each fraction and α-CHCA (0.3% in TFA) or SA (Sinapinic acid, 0.3% in TFA), was spotted on the matrix target, dried at room temperature, and analyzed in the mass spectrometer. Proteins were identified by searching a comprehensive protein database using Mascot programs (www.matrixscience.com). One/two missed cleavages per peptide were allowed, and an initial mass tolerance of 50 ppm was used in all searches.

The mass spectrum Ole e 1 peptides mixture (figure 24) was processed to assign candidate peptides in SWISS PROT and NCBI database using MASCOT search program. Peptide Mass Fingerprinting (PMF) interrogation unambiguously identified the protein as *Major pollen allergen (Allergen Ole e 1) (Ole e 1)* (SWISS PROT entry: P19963; NCBI ENTRY: gi[14424429]) using eight masses corresponding to three possible sequences (table 12).

	Accession	Mass	Description
1.	**gi|14424429**	16319	Major pollen allergen (Allergen Ole e 1) (Ole e I)
2.	**gi|13195753**	14575	main olive allergen [Olea europaea]
3.	**gi|1362133**	15288	major allergen OLE26 - common olive (fragment)

Table 12.

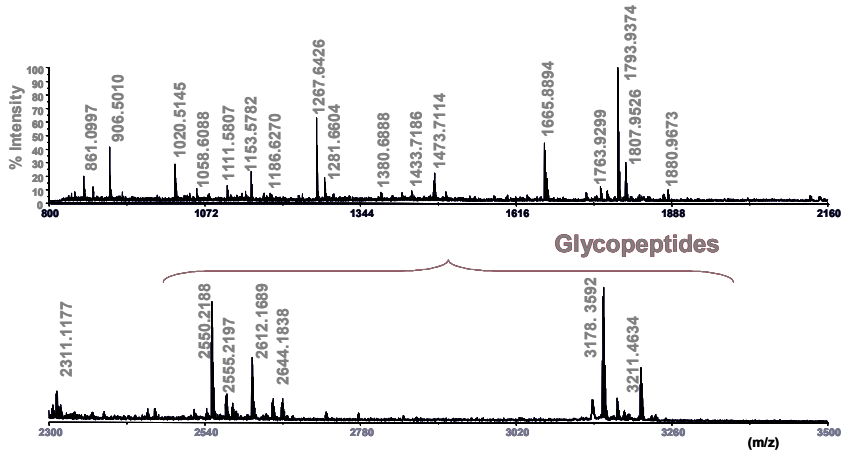


Figure 24. MS Ole e 1 peptides mixture.

The potential site of post-translational modification is located on the C-terminal portion of the protein corresponding to tryptic peptides detectable in reflectron ion mode, using sinapinic acid as matrix. MALDI MS spectrum of tryptic mixture showed the characteristic pattern of glycoforms within mass range (2-3) kDa (Figure 24). Any way, the absence of more large glycopeptides and digestion efficiency were monitored in linear mode.

A closer inspection of the characteristic patterns observed (Figure 25) provided some structural information. It can be observed the presence of three set of glycoforms and several extra peaks arising from the losses of one or two hexose residues. The fragmentation of glycopeptides affects the mass spectra interpretation and sets hurdle the evaluation of glycan-microheterogeneity. The glyco peptide at m/z 3182,6440 loses 527 and 1054 Da, forming the ion pecks at m/z 2655,0987 and at m/z 2128,0031, respectively, corresponding probably to a diantennary N-glycan specie. The ion peak at m/z 2776.1842 probably came from the loss of 2HexNAc, but it could be correspond to a simple natural glycoform.

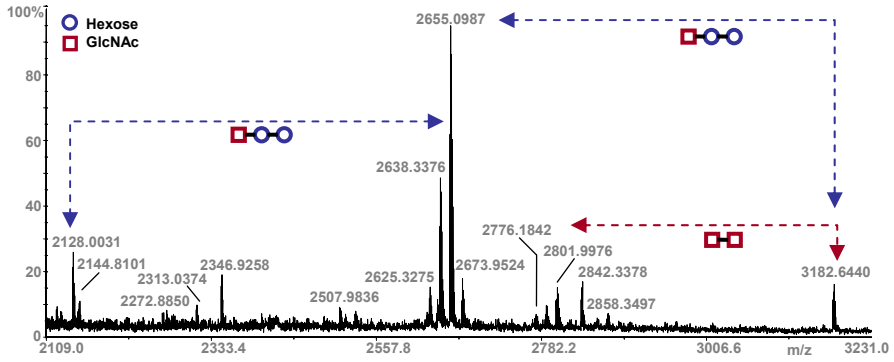


Figure 25. MS Ole e 1 glycopeptides mixture.

In plants, N-glycan-protein can bear high-mannose- and/or complex-type glycan moieties. High-mannose-N-glycans are usually characterized by the general structure $Man_{5-9}GlcNAc_2$; while complex-type N-glycans are carbohydrates having terminal GlcNAc residues or large antennae β(1, 2) linked to the α(1,6)- or α(1,3)-mannose constitutive of conservative core (Man_3-$2GlcNAc_2$). The conservative core could take an extra α(1, 3)-fucose and/or β(1, 2)-xylose residue[143].

Spectral interpretation can be facilitate by a free available program, GlycoMode (www.expasy.ch/tools/glycomod), using as input literature data about conservative core and peptide information obtained by mass spectrometry. Therefore, the hypothetical glycans structure of the different glycopeptides can be suggested on the base of the congruence of their molecular weights by means of a bioinformatics tool.

Hybrid complex and high mannose structure type were assigned to the observed glycopeptides, hypothesizing that the peptide backbone carrying glycans moiety was LNTVNGTTR (m/z 975, 5223, missed cleavage 0), instead of **FK**LNTVNGTTR (m/z 1250, 6847, missed cleavage 1). The selected peptide is derived from the extensively cleavage of site K_{106} (P19963). Using the chosen peptide, the conservative core composition (three Hexoses, Hex, and two N-acetylhexoses, HexNAc), and one of the observed mass, i.e. that at m/z 3182,644, the program GlycoMode calculated the

hypothetical oligosaccharides. On the basis of the data, a hybrid complex $(Hex)_3$ $(HexNAc)_2$ $(Deoxyhexose)_2$ $(Pent)_1$ + $(Man)_3(GlcNAc)_2$ with an error of +524ppm, was assigned to the glycopeptides at m/z 3182.644.

MS/MS experiments were used to confirm the hypothetical carbohydrates composition and the AA peptide sequence matched by the GlycoMode algorithm. Although the extent of fragmentation of large ion precursor at m/z 3182.644 is low, direct cross ring fragmentation leads to the release of $[(HexNAc)_2(Hex)_2+43]$ and consecutive lost of two hexose residues; therefore the observed fragmentation pathway does not confirm the suggested hybrid complex structure. However, the lost of $[(HexNAc)_2(Hex)_2+43]$ can be only justified suggesting a biantennary glycan structure, but no information is obtained bring about the identity of peptide itself $[P+H]^+$ and the related ion species $[P+H-17]^+$/ $[P+GluNAc+H]^+$ / $[P+H+83]^+$, that would be observed in the MS/MS spectrum of the glycopeptides, as reported in literature[144].

To overcoming the lack of information we decided to derivatize peptides at the N-terminal amino function with a well known group to improve their ionization and to increase structural information. Dansyl chloride, as a masking group, represents the best choose, because the corresponding derivatized peptides are preferentially protonated at aromatic amine in MS experiments[145]. The derivatized peptides should have great signal intensities and an increasing of the N-terminal charged direct fragmentation[146].

20 µL of tryptic digested *Ole e 1* sample (pH 8) were mixed with 16 µL of a Dansyl-Cl solution (0,22mol/L, in CH_3COCH_3), 10 µL of $NaHCO_3$ (0,1M) and 5 µL of saturated Na_2CO_3. The solution was stirred for 3 hours at room temperature, the corresponding Dansylated peptides were obtained, and the solution was directly analyzed by MALDI TOF/TOF workstation in single MS mode.

The obtained spectrum is showed in Figure 26 and the interrogation of PMF identifies Ole e 1 using 10 masses corresponding to eleven possible sequences. This result is not surprising because the mass range observed became broader, since the masses are increased by the addition of Dns-group.

The mass spectrum (Figure 26) shows more peaks than from unmodified peptides (Figure 24). Peaks are from peptides with one or two dansyl modifications; while unmodified peptides were not observed indicating that the dansyl derivatization takes place for all peptides. Dansylation of the N-terminal amino group increases desorption/ionization efficiency of peptides as well as glycopeptides, using a-CHCA as matrix.

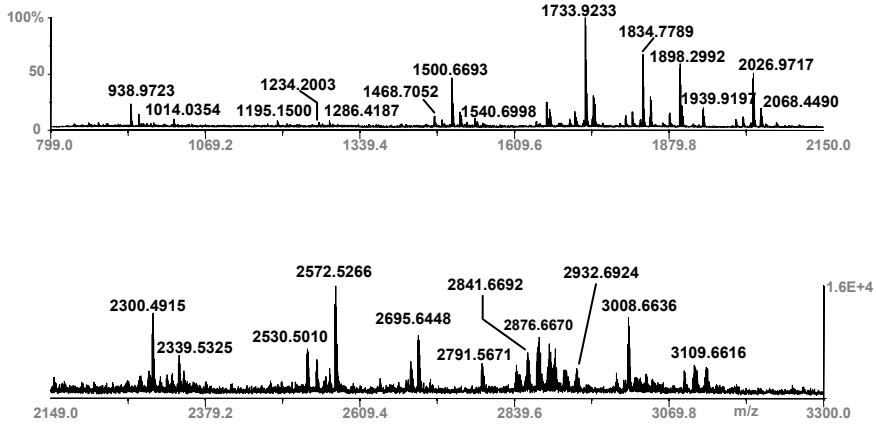


Figure 26. MS Ole e 1 derivatized peptides mixture.

Peptide peaks showing characteristic glyco-patterns were selected as ion precursors. In all cases, low energy CID of derivatized peptides produced mainly three ion peaks at m/z 1469, 1553 and 1672. The first is peptides itself $[Dns\text{-}P+H]^+$, while 1553 and 1672 arise from cleavage at the first N-acetylglucosamine residue $[Dns\text{-}P+83+H]^+$ and from Y-type cleavage of chitobiose core $[Dns\text{-}P+203+H]^+$ with the retention of the peptide moiety (P). The latter suggested that all glycopeptides investigated does not contain a fucosylated core.

High energy CID spectrum of one dansylated-glycopeptides is shown in Figure 27 as representative example for all the glycopeptides investigates. In all cases, the spectra showed fragmentation of the peptide chain and also is possible observing the significant pecks at m/z 1469.53, 1553.19 and 1672.12, whereas, under the chosen condition, the fragmentation of the glycan structure was negligible. The interpretation of derivatized

spectra reveals that the sequence of N-Dns-peptide, carrying glycan moiety, is **Dns-FILNTV*N*GTTR** (theoretical mass 1469.7337). MS spectrum shows an ion peak at m/z 1468.7052 probably corresponding to **Dns-FILNTV*N*GTTR** (-20 ppm error) arising from the release of the neutral carbohydrate moiety during the ionization process.

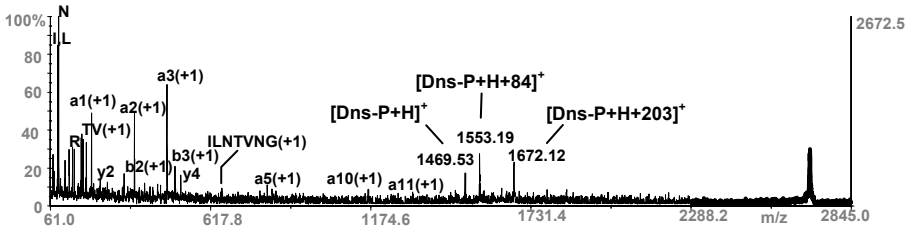


Figure 27. High Energy CID MSMS of the ion at m/z 2695,6448.

Fragments of the proposed sequence match the ion peaks visible in the MS/MS spectrum (Figure 28). Direct submission of MS/MS data for protein identification against other green plants, NCBI database allows to identify the isoallergens Ole e 1.0102, 1.0103 and 1.0105 (Swiss-Prot entry P19963). These are the only three known isoallergens of Ole e 1 family, containing a punctual modification $K_{106} \rightarrow I$.

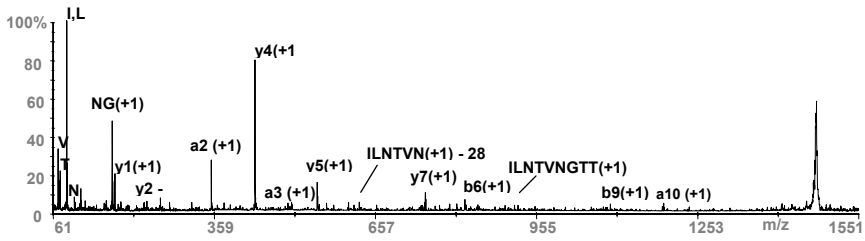


Figure 28. MSMS of the ion at m/z 1468,7052 (Dns–**FILNTVNGTTR).**

Therefore the peptide carrying glycan moiety is F*I*LNTVNGTTR (m/z 1234,6448) and not the hypothized LNTVNGTTR (m/z 975,5223) or FKLNTVNGTTR (m/z 1250,6847).

The identification of peptide FILNTVNGTTR (missed cleavage 0), generated by the extensive triptic cleavage site K^{104} on Ole e 1 (P19963), provides the possibility to assign hybrid complex and high mannose structure types (Table 13). The structure 1 (Table 13), now easily confirms the neutral loss previously discussed. While 2, 3, 5 and 6 (Table 13) probably arise from 1 by the consecutive lost of Hex and HexNAc residues during the ionization processes, although most plant glycoproteins until now studied contain small "truncated structures" terminating in mannose or HexNAc[147].

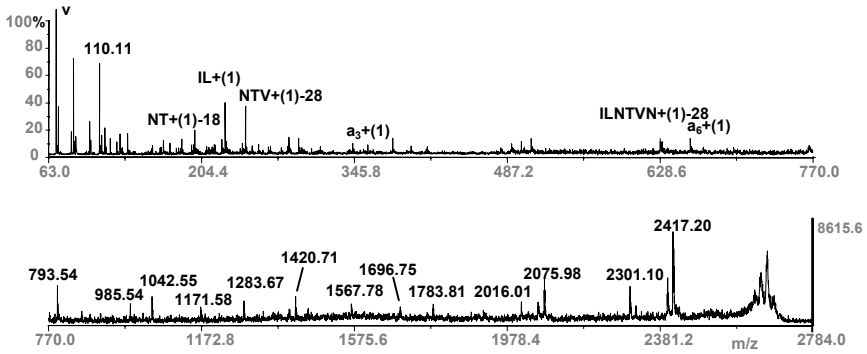


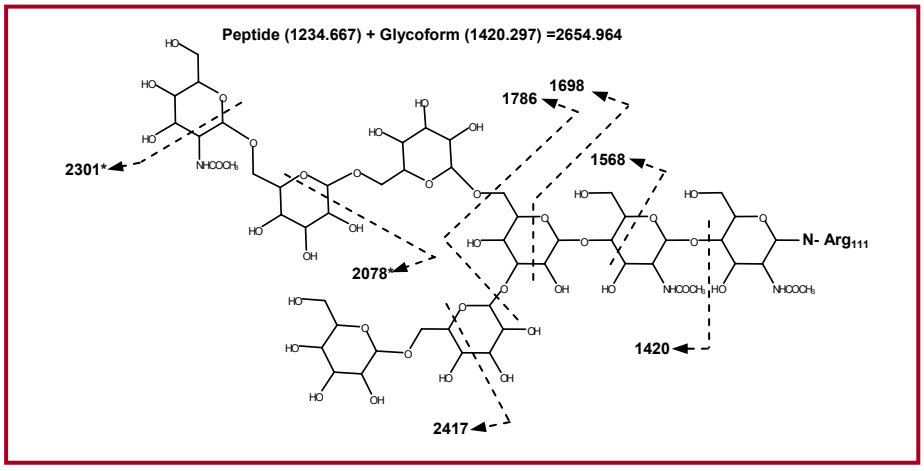
Figure 29. High energy CID MSMS of the ion at m/z 2655,0987.

High energy CID spectrum of one glycopeptides is shown in Figure 29 as representative example for all the glycopeptides investigates. The spectrum shows proteolytic fragments on either sides of the amino acids carrying the carbohydrate and fragments belonging to the cross ring cleavages, mainly (Scheme 5). These are not commonly observed in MS/MS spectra of glycopeptides and these features are generally related to the lack of sufficient energy.

The presence of cross ring fragmentation in a MALDI TOF/TOF spectrum can be attributed to the rearrangements of metastable ions produced during ionization process[148].

	glycoform mass	Structure	type	Mr(cal)	Mr(cal)	Observed
1	1947.777	$(Hex)_4$ $(HexNAc)_2$ + $(Man)_3(GlcNAc)_2$	hybrid/complex	1234.667	3182.444	3182.644
2	1623.492	$(Hex)_2$ $(HexNAc)_2$ + $(Man)_3(GlcNAc)_2$	hybrid/complex	1234.667	2858.159	2858.3497
3	1607.493	$(Hex)_1$ $(HexNAc)_2$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$	hybrid/complex	1234.667	2842.160	2842.3378
4	1541.387	$(Hex)_4$ + $(Man)_3(GlcNAc)_2$	high_man	1234.667	2776.064	2776.1842
5	1420.297	$(Hex)_2$ $(HexNAc)_1$ + $(Man)_3(GlcNAc)_2$	hybrid/complex	1234.667	2654.964	2655.0987
6	1403.507	$(Hex)_1$ $(HexNAc)_1$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$	hybrid/complex	1234.667	2638.174	2638.3376
7	1389.492	$(Hex)_1$ $(HexNAc)_1$ $(Pent)_1$ + $(Man)_3(GlcNAc)_2$	hybrid/complex	1234.667	2625.166	2625.3275
8	1227.439	$(HexNAc)_1$ $(Pent)_1$ + $(Man)_3(GlcNAc)_2$	hybrid/complex	1234.667	2463.113	2463.2734
9	892.317	$(Hex)_3$ $(HexNAc)_2$	-	1234.667	2127.991	2128.0031

Table 13.



Scheme 5. Fragmentation of Glycoform 1420,297 Da.
(*consecutive fragmentation from the ion at m/z 2417,20).

This fragment ion analysis confirms the canonical structure of an N-linked biantennary glycan for the most abundant glycopeptides. The glycoforms 4 and 8 (Table 13) are in agreement with the structures published by Van Ree [131]. While the structure 7 (Table 13) may considered as a higher analogue than the well known glycoform 8 (Table 13).

A.N.	Sequence		
14424429	2.	DIPQPPVSQFHIQGQVYCDTCRAGFITELSEFIPGASLRLQCKDKENGDVTFTEVGYTRA	61
2465129	3.	V.....V.....I.....	62
2465127	3	.V.....V.....I.....	62
1362131	2	.V....I...YV.....TR.....F.....GV.....G...KI.....	61
33329739	16.	----------S.....GV.....G...NI.....	62
14424429	62	EGLYSMLVERDHKNEFCEITLISSGRKDCNEIPTEGWAKPSLKFKLNTVNGTTRTVNPLG	121
2465129	63	.....I.S.....	122
2465127	63	.....I.....R	122
1362131	62	.....I.....L..S....D.....V...V..I.....I....	121
33329739	62	.....D.....V.....I.....I....	121
14424429	122	FFKKEALPKCAQVYNKLGMYPPNM	145
2465129	123	.Y.....	146
2465127	123	.Y.....	146
1362131	122	.....P..F.....	145
33329739	122		145

Figure 29. Alignment of the 5 most important variants of Ole e 1.
gi[14424429 Major pollen allergen (Allergen Ole e 1)(Ole e I); **gi[2465129** Ole e 1.0103; **gi[2465127** Ole e 1.0102, **gi[1362131** Ole e 1c; **gi[33329739** Major pollen allergen Ole e 1 (Q5DVQ9_OLEEU).

The amino acid sequence of five variants of Ole e 1 with potential PTMs is shown in Figure 29. The sequences 38–51 and 83–99 (Figure 29) are responsible for the high micro heterogeneity of the deduced cDNA forms of Ole e 1 [95]. These regions are to the most hydrophilic and antigenic areas of Ole e 1 [97].

While the post-translational modification of Ole e 1 and the specific amino acid residues were identified from underivatized and derivatized peptides, the microheterogeneity can be evaluated from the analysis of tryptic peptides after removal of N-glycans. In attempt to simplify spectra and to obtain the separation of potential sites from punctual modification, Ole e 1 fraction was digested after N-deglycosilation by PNGase F.

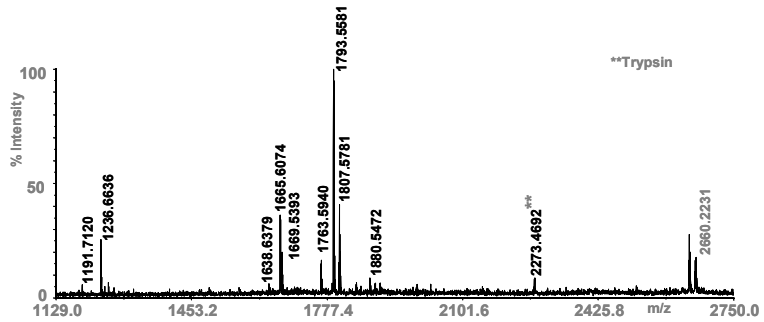


Figure 30. MALDI MS of Ole e 1 triptic peptide mixture after N-deglycosilation by PNGase F.

The tryptic digests was examined in reflectron mode and MS/MS experiments were used to peptide identification.

Ion peaks corresponding to protonated tryptic peptides of Ole e 1 were easily detected, with mass errors of 8 ppm (Figure 30). The interrogation of the PMF (MASCOT search program, NCBI database) unambiguously identified Ole e 1 using 6 masses (missed cleavage 1) corresponding to seventeen possible sequences (Table 14).

	A. N.	Mass (Da)	Pep. Match.	Score	Protein Name
1	gi|145313986	15750	6	67	major pollen allergen Ole e 1 [Olea europaea]
2	gi|14424429	16319	6	66	Major pollen allergen (Allergen Ole e 1)(Ole e I)
3	gi|37724593	15191	6	66	major pollen allergen Ole e 1 [Olea europaea]
4	gi|37724597	15065	6	64	major pollen allergen Ole e 1 [Olea europaea]
5	gi|33329750	14759	5	50	major pollen allergen Ole e 1 [Olea europaea]
6	gi|33329744	14715	5	50	major pollen allergen Ole e 1 [Olea europaea]
7	gi|33329730	14831	5	50	major pollen allergen Ole e 1 [Olea europaea]
8	gi|33329734	14904	5	50	major pollen allergen Ole e 1 [Olea europaea]
9	gi|37548749	14801	5	50	major pollen allergen Ole e 1 [Olea europaea]
10	gi|33329742	14844	5	48	major pollen allergen Ole e 1 [Olea europaea]
11	gi|33323443	14873	5	48	major pollen allergen Ole e 1 [Olea europaea]
12	gi|33325111	14847	5	48	major pollen allergen Ole e 1 [Olea europaea]
13	gi|13195753	14575	4	36	main olive allergen [Olea europaea]
14	gi|2465131	16605	5	50	Ole e 1 protein [Olea europaea]
15	gi|1362129	15491	5	49	major allergen OLE17 - common olive (frag.)
16	gi|145313982	15920	5	48	major pollen allergen Ole e 1 [Olea europaea]
17	gi|145313988	15775	4	37	major pollen allergen Ole e 1 [Olea europaea]

Table 14.

Protein identification by PMF is considered a rapid and useful method to identify proteins previously separated by 2D-gel. However, PMF can be used to identify the conservative domain matching proteins, in particular in the analysis of a family of the same antigen. In fact, Ole e 1 (gi[14424429]) was identified matching 6 peptides, with score 66 and 48% sequence coverage.

An enlargement of figure 30 (figure 31) shows a difference of 3.95 mass units between the peaks at m/z 1665.6074 and 1669.5393, and a more complex pattern of isotopic peaks indicating that Ole e 1 fraction represents a mixture of several components. In literature, there are two sequences published differing by the exchange of two amino acids at the positions 31 (L→ F) and 39 (S →G), respectively, leading to a mass difference of only 3.95

mass units. The theoretical calculated isotopic distribution and the sum of isotopic distributions of both segments suggest that the variant Ole e 1 (gi: 14424429, Figure 29) and Ole e 1c (gi: 1362131, Figure 29) exist in the mixture of natural material.

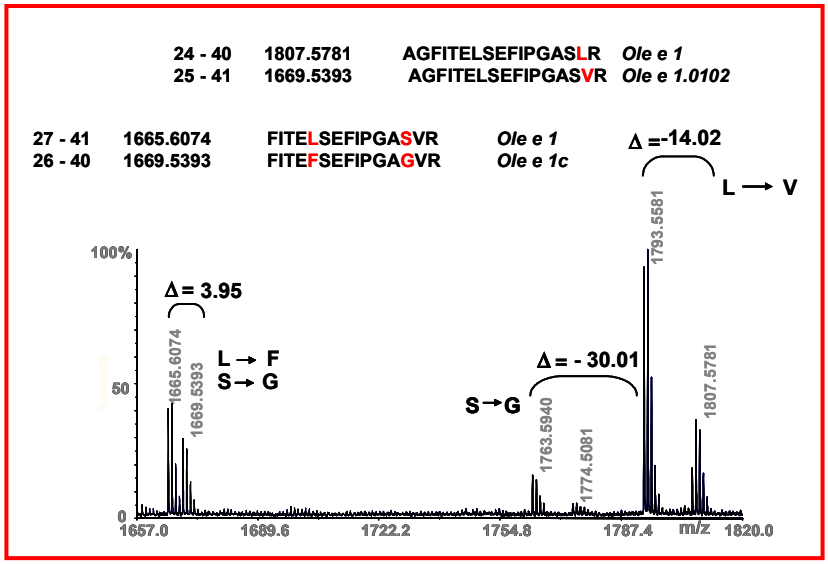


Figure 31.

The MS/MS experiments of these peptides validated the sequences (27-41) FITE*L*SEFIPGA*S*VR and (26-40) FITE*F*SEFIPGA*G*VR belonging to Ole e 1 (gi:14424429, Figure 29) and Ole e 1c (gi:1362131, Figure 29), respectively.

The peptide mass 1793.4300 identified by data base search and validated by MS/MS experiment as the peptide fragment AGFITELSEFIPGASVR (Figure 32) matched two different variant of Ole e 1. This sequence is contained in Ole e 1.0102 (gi: 2465127, Figure 29) and Ole e 1.0103 (gi:2465129).

A mass difference of 14.02 units between the adjacent peaks at m/z 1793.4300 and 1807.5781 can be justified by the exchange of one amino acids at the positions 40 (L→ V)

indicating the presence of an ulterior variant. The MS/MS confirms the proposed sequence AGFITELSEFIPGAS*I*/R identifying Ole e1 (P19963, gi:14424429, Figure 31).

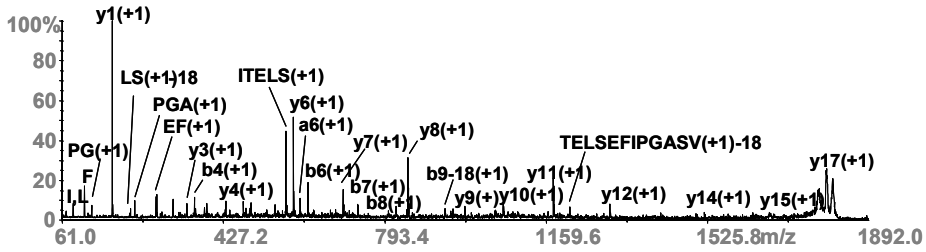


Figure 32. MALDI MS/MS spectra of the ion at m/z 1793.43

The exchange of one amino acids at the positions 39 (S→ G) leading to a mass difference of 30.01 mass units allows to attribute the sequence AGFITELSEFIPGA*G*VR to the experimental mass 1763.9187 Da. The theoretical mass of the candidate is 1763.9326 Da, thus the experimental mass error difference is 8 ppm. Though, PMF interrogation does not match the proposed sequence peptide AGFITELSEFIPGA*G*VR, an MS/MS spectrum was acquired from these peptides and all peaks were manually matched by the believed sequence.

A free available program, BLAST (www.expasy.org/tools/blast) was used to identify the protein source of the observed specific peptide. The variant Major pollen allergen Ole e 1 (Q5DVQ9_OLEEU, 33329739) was assigned as the protein source of the analyzed peptide (AGFITELSEFIPGA*G*VR) by comparison with sequences available in the UniProt Knowledgebase (Swiss-Prot+Trembl, using complete database).

Conclusions.

In this research work we have adopted some specific proteomics advances to characterize proteins with particular properties, obtaining chemical information about them without preventive classical separation, but only by mean of chemical fractionation procedure followed by mass spectrometry MALDI TOF-TOF.

Two natural matrixes were analyzed: olive pollen tree and raw milk from cows affected by mastitis. The identification and the structural characterization of proteins and peptides were performed employing *mining*, *protein-expression profiling* and *mapping of protein modifications* approaches.

The first step, for each sample, was represented by the attainment of reproducible procedure of extraction and fractionation of the total protein content from matrices, followed by linear MS profiling.

Mass *protein-expression profiling* of the examined pollens samples, Ottobratica (**1**), Carolea (**2**), Dolce di Rossano (**3**), Cassanese (**4**), Coratina (**5**), Nocellara del Belice (**6**), Villacidro (**7**), and Sinopolese (**8**), should take into account the presence of different allergenic and antigenic patterns in the pollen of olive tree pollen, from the same and, also, from different cultivars[149]. Furthermore, qualitative and quantitative antigenic content of a given pollen can be affected by environmental adaptation and factor management. The lack of sensitivity of classical methods, like SDS-PAGE or 2-D chromatographic separation, was overcoming by the employing of mass spectrometry combined with an suitable chemical fractionation procedure. The previous separation of proteins of different hydrophilicity from the lipophilic ones prevents those typical suppression effects frequently

observed when desorption methodologies are applied and improves the efficiency of matching MALDI spectra of different pollen fractions. The antigenic profile of the considered cultivar (1-8) has given the possibility to suggest a distinction among the pollen of the set of the examined cultivars by matching the relative amount of Ole e 1, according to the following criteria:

(i) Low Ole e 1 content, such as *Carolea*, *Dolce di Rossano* and *Sinopolese*;

(ii) Over-enriched in Ole e 1, such as *Cassanese*, *Coratina*, *Nocellara del Belice* and *Villacidro*;

(iii) Those containing Ole e 3 and Ole e 7 only, such as *Ottobratica*.

The exception, represented by *Ottobratica* (1), reveals that two isoforms of Ole e 3 are present in the natural matrix and that Ole e 7 consists of four isoforms characterized by a different glycosilation degree. These results demonstrate that the proposed experimental procedure can supply valuable information on the antigens' micro heterogeneity.

An upgrading of this procedure was employed to obtain mass *protein-expression profiling* of the content of raw bovine milk. Also in this case, the separation of proteins by their different hydrophilic/lipophilic properties prevents the suppression effect typical of desorption methodologies and makes more effective the matching of MALDI spectra of different milk fractions. The protein MS profiling of the examined samples (1-4, milk samples from healthy cows and 5-8, milk samples from cows affected by mastitis) indicates that different proteic patterns can be found in mastitic milk. In particular, the profiling of milk proteins thus obtained allowed to identify both functional protein marker and endogenous peptide of the innate immune response, as a function of the inflammatory event. The presence of BSA in one of the hydrophilic, from mastitic cow milk fraction suggests an increased permeability of the blood mammary barrier, which is a clear indication of a high level of inflammation.

Therefore BSA can be considered as a functional marker for the acute phase of mammary gland inflammation. Furthermore the presence of a polypeptide at *m*/*z* 4338 can

be associated with *β*–defensin, an endogenous cationic peptide with well know antimicrobial activity, generated by the innate immune response.

Furthermore, mastitic milk samples reveal increased proteolytic activity, which is certainly associated to the releasing of PMN cells from blood. The presence of several enzyme families, generated by an high number of SC, is confirmed by the extent of the breakdown of caseins. The experimental data confirm that those protease families have a particular predilection for caseins, in fact the SDS-PAGE was used to check the partial or total idrolysis of caseins over 24-, 48-, 96-, and 216-h incubation at 37 °C at both physiological (pH~8) and acid pH. The observed results were confirmed by mass spectrometry.

Sequence-specific peptides were identified for each fraction by MS/MS experiments, and all tandem mass spectra were evaluated using MASCOT database searching. The results show a specific proteolytic activity of endogenous enzyme toward β-casein precursor (P02666), α-S2-casein (P02663), α-S1-casein (P02662), and κ-casein (P02668). However to obtain roundabout information on the presence of endogenous proteases, it was necessary using free algorithms, PeptideCutter (www.expasy.org), to simulate specific enzymatic cleavages. Data showed that these proteolytic activities was performed preferentially by Cathepsins D and G under acid and basic conditions, respectively.

Hence, the results thus obtained suggest that the proposed experimental procedure can provide important information on the presence and activity of numerous cell-derived proteolytic enzymes in milk [59].

The exploitation of the *mapping of protein modifications* approach, allowed the evaluation of the primary structures of the *N*-glycans of the major pollen allergen Ole e 1, of *Olea Europaea*. *Ole e 1* (P19963) is a polymorphic glycoprotein showing several variants differing each others in the glycosylation state. It contains only one potential N-glycosylation site at Asn_{111} according to SWISS-PROT database.

Classical method adopted to characterize posttranslational modification of proteins comprise 2D gel, in-gel enzyme digestion and mass spectrometry analysis.

The alternative approach, here presented, requires a simple procedure of chemical fractionation of the whole saline proteome extract of pollen, leading to a sample containing Ole e 1 family, only, followed by digestion with PNGase F/Trypsin, and by the identification of peptides mixture by means of MALDI MS and MS/MS combined with the software tool Mascot (www.matrixscience.com).

Mass spectrometry is a rapid a sensitive means for characterizing the primary structure of proteins and their posttranslational modifications including glycosylation degree playing an important role in the biological properties of many proteins. Accordingly, it was worth to employ a free available program GlycoMode (www.expasy.ch/tools/glycomod), to determine the hypothetical glycan portion and its anchor site within the peptide backbone.

The latter, however, does not provide straightforward results, therefore a derivation step with the know dansyl group of glyco-peptide mixture. was attempted to fulfil the goal of structure identification by MS and MS/MS measurements. Fortunately, this approach, was appropriate to identify the peptide sequence FILNTVNGTTR (m/z 1234.667) as the backbone segment where the glycan moieties are linked. This new procedure could be exploited in further applications.

In the case here examined, the direct loading of MS/MS data of the dansylated tryptic peptides on Mascot program (Taxonomy: other green plants, Database : NCBI) allowed us to identify the isoallergens Ole e 1.0102, 1.0103 and 1.0105. The latter are the only three known isoallergen of Ole e 1 containing a punctual modification $K^{106} \rightarrow I$, therefore are the only three possible isoforms originating the peptide FILNTVNGTTR that carries the glycan moiety. Moreover hybrid complex and high mannose structure type were assigned to the observed glycopeptides on the base of the sequnce FILNTVNGTTR, by a manual interpretation of MS/MS spectra of the glyco-peptides. Eight glycan moiety were identified, but probably the principal forms are the two hybrid complex structures $(Hex)_4$ $(HexNAc)_2$ + $(Man)_3(GlcNAc)$ and $(Hex)_1$ $(HexNAc)_1$ $(Pent)_1$ + $(Man)_3(GlcNAc)_2$, and the high mannose structure $(Hex)_4$ + $(Man)_3(GlcNAc)_2$, the other glyco-moietys most likely arise from the consecutive lost of specific sugar residues.

The results presented in this report demonstrate that the proposed innovative experimental procedure can supply valuable information on the proteins' micro heterogeneity providing an alternative approach to resolve "proteomics troubles". Moreover this procedure can afford important roundabout information on the presence and activity of numerous cell-derived proteolytic enzymes in milk, in order to characterize their activity and function.

Appendix

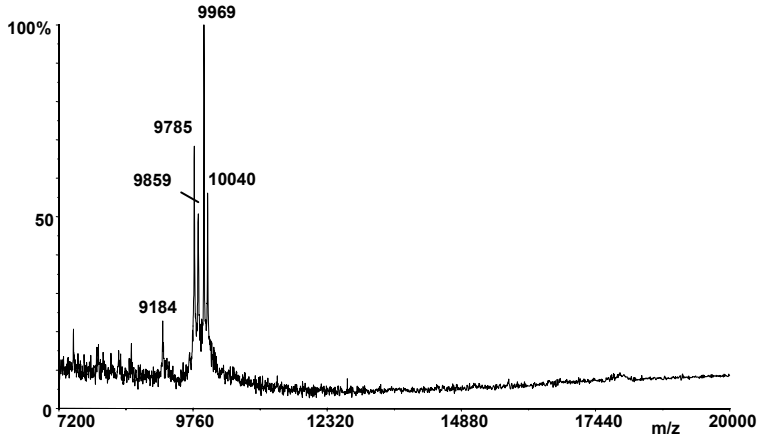
(A.1) Table 5_Part 2. ***Olive Pollen Allergens with Clinical Relevance Developed and Maintained by Allergen Nomenclature*** _ Subcommittee of the IUIS (www.allergen.org), Including Allergens Whose IgE Reactivity Has a Prevalence of >5%**. Legend: (1)** Apparent molecular mass in SDS-PAGE. (**2)** Theoretical molecular weight. **(3)** Sequence information obtained by C, cDNA; P, peptide sequence;N, nucleotide sequence. **(4)** Swissprot database. **(5)** Mass spectrometry determination.

Allergen name	Common Name	Variants	MW (Da)[2]	Accesion number [4]
Ole e 1	Major pollen allergen	**3-3**, I→V (in Ole e 1.0102, Ole e 1.0103 and Ole e1.0105) **23-23**, R→P (in Ole e1.0105) **24-25**, AG→SR (in Ole e1.0105) **39-39**, L→V (in Ole e 1.0102, Ole e 1.0103 and Ole e1.0105) **44-44**, K→R (in Ole e1.0105) **45-45,** D→E (in Ole e1.0105) **46-46**, K→I (in Ole e1.0105) **47-47**, E→K(in Ole e1.0105) **48-48**, N→K(in Ole e1.0105) **50-50**, D→S(in Ole e1.0105) **51-51**, V→I (in Ole e1.0105) **56-56**, V→I (in Ole e 1.0102, Ole e 1.0103 and Ole e1.0105) **69-69**, V→I (in Ole e1.0105) **87-87**, R→S (in Ole e1.0105) **91-91**, N→D (in Ole e1.0105) **95-95**, T→I (in Ole e1.0105) **106-106**, K→I (in Ole e 1.0102, Ole e 1.0103 and Ole e1.0105) **108-108**, N→S (in Ole e 1.0103) **111-111**, N→D (in Ole e 1.0102) **121-121**, G→R (in Ole e 1.0102) **123-123**, F→Y (in Ole e 1.0102, Ole e 1.0103)	16330	P19963

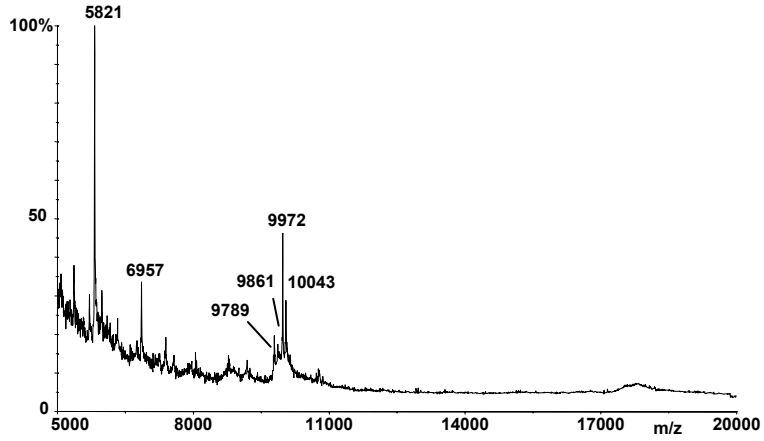
Ole e 2	Profilin-1		14489	O24169
	Profilin-2		14427	O24170
	Profilin-3		14399	O24171
Ole e 3	Polcalcin Ole e 3		9356	O81092
	Olee3 allergen(frag.)		5798	Q5DTB7
Ole e 4	Major pollen allergen Ole e 4 (frag.)		2711	P80741
Ole e 5	Superoxide dismutase [Cu-Zn] (frag.)		2973	P80740
Ole e 6	Pollen allergen Ole e 6		5833	O24172
	Ole e 6 allergen (frag.)		4986	Q84UC2
Ole e 7	Pollen allergen Ole e 7 (frag.)	**5-5**, S→G (in type B) **10-10**, L→K (in type B) **18-18**, I→K (in type B)	9905-10032[5] 2199	P81430
Ole e 8	Ca^{2+}-binding protein	**43-44**, GV→CA **58-58,** G→A **60-50,** I→M	18907	Q9M7R0
Ole e 9	B-1,3-glucanase		48830	Q94G86

(A.2) MALDI MS, first hydro-soluble fraction.

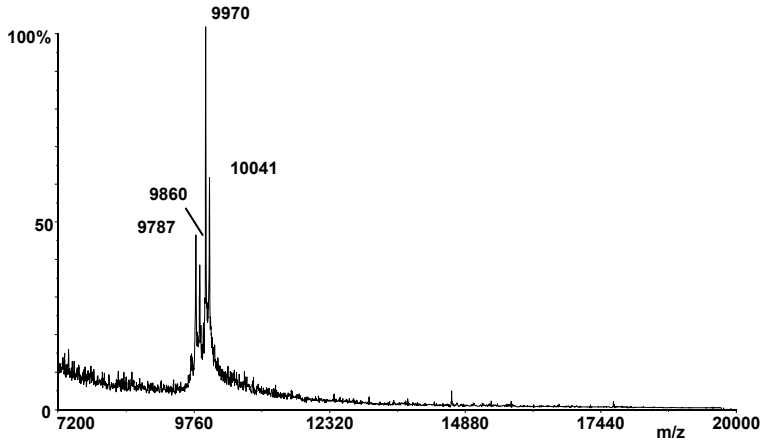
MALDI MS. First hydro-soluble fraction from pollen Dolce di Rossano (**3**).



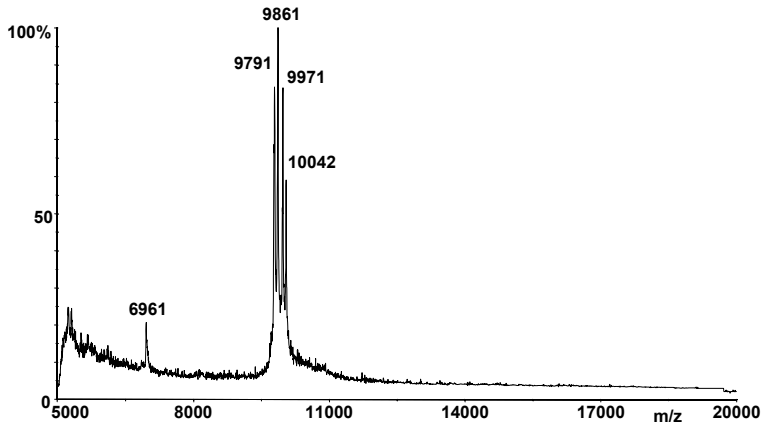
MALDI MS. First hydro-soluble fraction from pollen Cassanese (**4**).



MALDI MS. First hydro-soluble fraction from pollen Coratina (**5**).

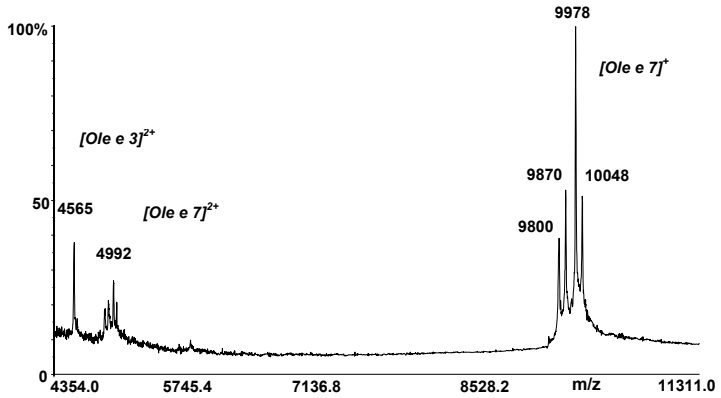


MALDI MS. First hydro-soluble fraction from pollen Sinopolese (**8**).

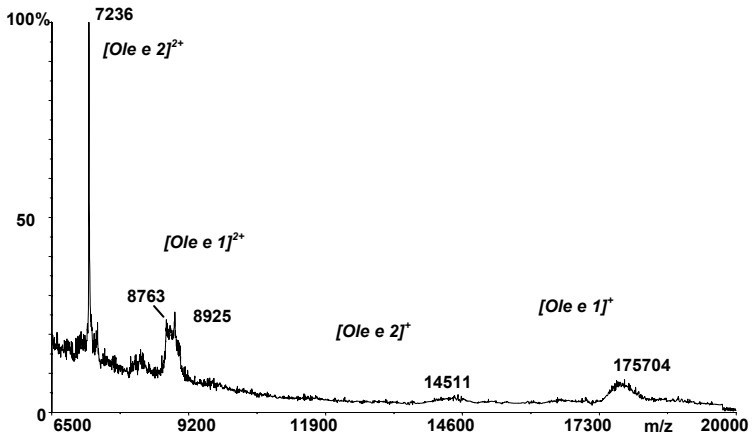


(A.3) MALDI MS, second hydro-soluble fraction.

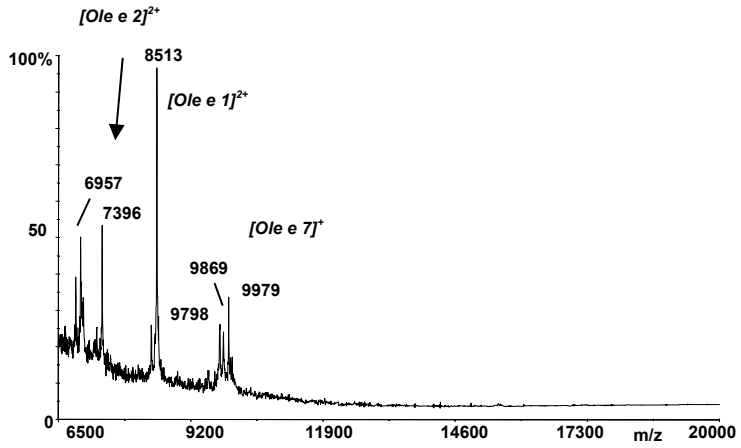
MALDI MS. Second hydro-soluble fraction from pollen Ottobratica (**1**).



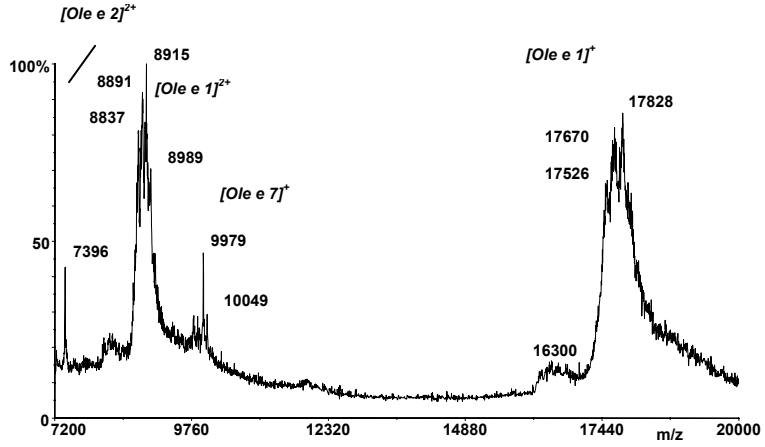
MALDI MS. Second hydro-soluble fraction from pollen Carolea (**2**).



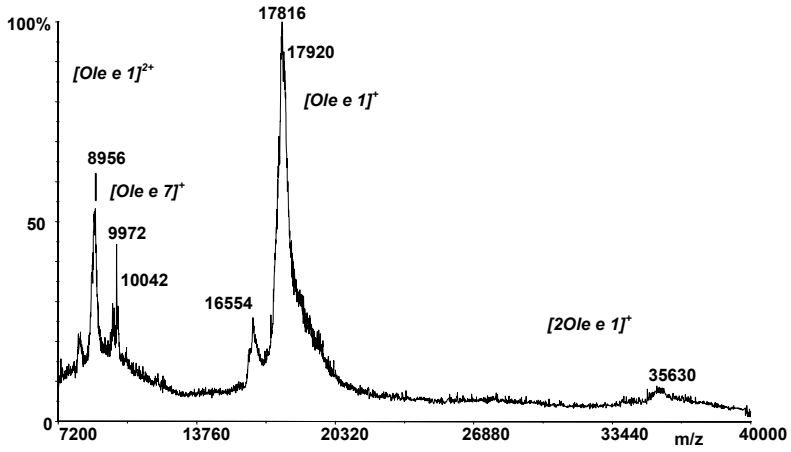
MALDI MS. Second hydro-soluble fraction from pollen Dolce di Rossano (**3**).



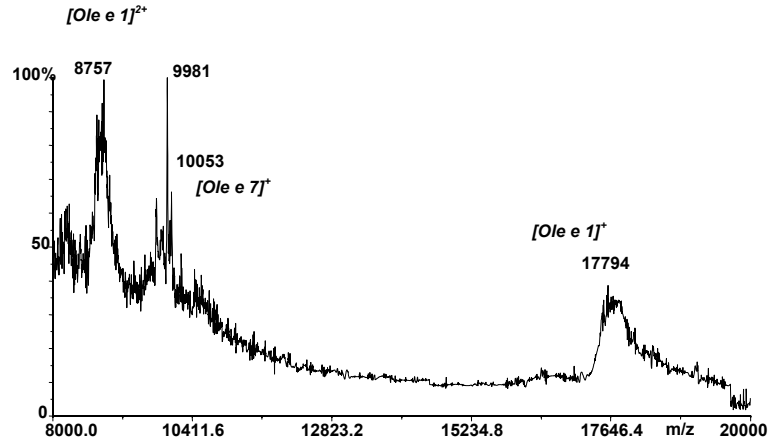
MALDI MS. Second hydro-soluble fraction from pollen Cassanese (**4**).



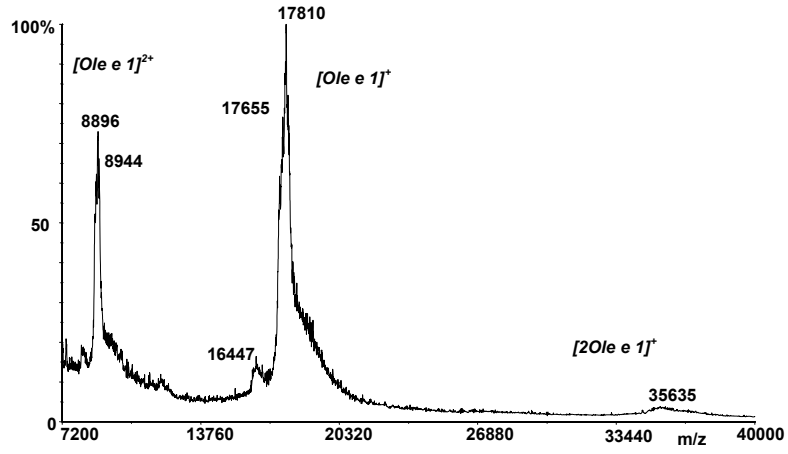
MALDI MS. Second hydro-soluble fraction from pollen Coratina (**5**).



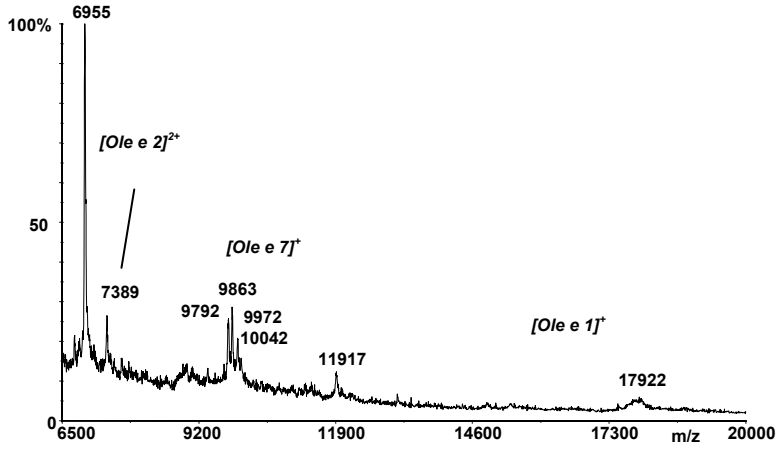
MALDI MS. Second hydro-soluble fraction from pollen Nocellara del Belice (**6**).



MALDI MS: second hydro-soluble fraction from pollen Villacidro (**7**).



MALDI MS. Second hydro-soluble fraction from pollen Sinopolese (**8**).

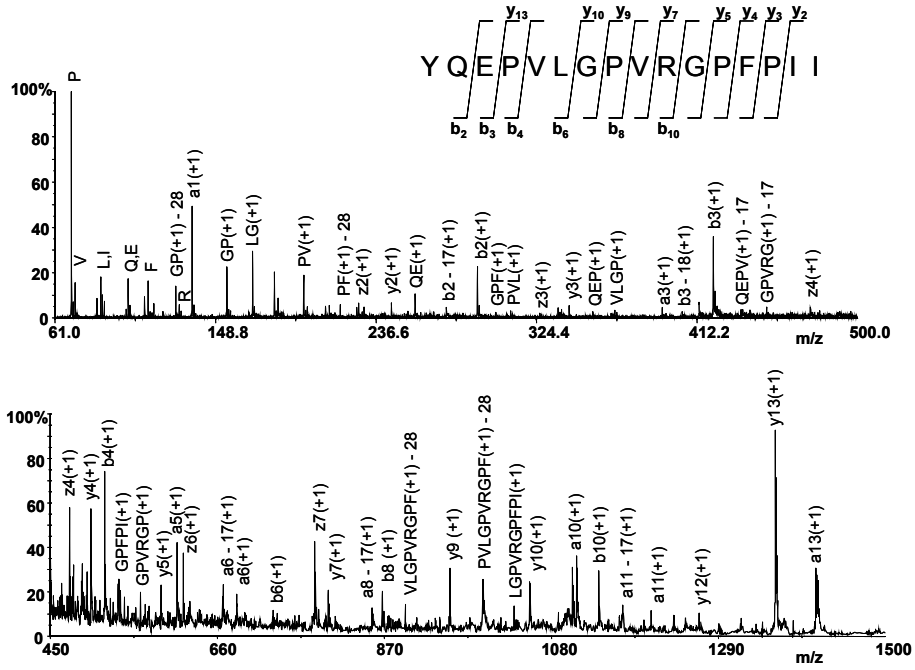


(A.4) Table 10.

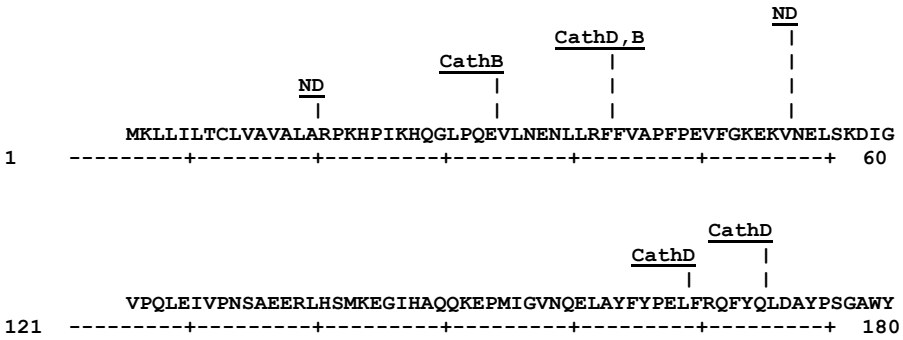
Peptides from fractions **Ls**, **Ha** and **Hb** (pH 8) identified by Database search.

		Protein name	Acc. No.	Peptides sequence	Span	$[M+H]^+$
1	**Ls**	β-casein precursor	P02666	WMHQPHQPLPPTVMFPPQSVL	158-178	2468.00
2	**Hb**	β-casein precursor	P02666	LLYQEPVLGPVRGPFPIIV	206-224	2106.92
3	**Hb**	β-casein precursor	P02666	LYQEPVLGPVRGPFPIIV	207-224	1993.85
4	**Ls**	β-casein precursor	P02666	YQEPVLGPVRGPFPII	208-223	1782.10
5	**Ls**	β-casein precursor	P02666	YQEPVLGPVRGPFPIIV	208-224	1881.18
6	**Ls**	β-casein precursor	P02666	VLGPVRGPFP	212-221	1038.68
7	**Ls**	β-casein precursor	P02666	SLSQSKVLPVPQKAVPYQRDMPIQAF	179-204	3024.79
8	**Ls**	α-S1-casein	P02662	RPKHPIKHQGLPQ	16-28	1535.66

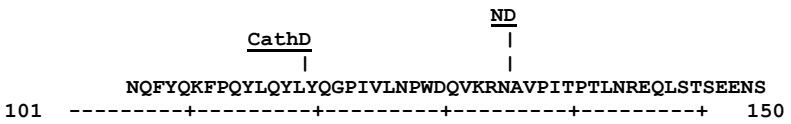
(A.5) Figure 22. MALDI MS/MS spectrum of ion peak at m/z 1781,97.



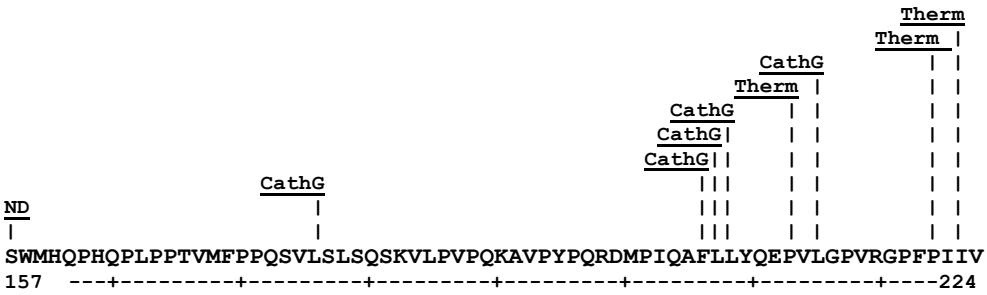
(A.6) Scheme 2. Cleavage sites identified on αs1-casein (P02662) at pH 6.5.



(A.7) Scheme 3. cleavage sites identified on αs2-casein (P02663) at pH 6.5.



(A.8) Scheme 4. Cleavage sites identified on β-casein (P02666) at pH 8.



References

1. Wilkins, M. R. et al. *Biotechnology*, **1996**, *14***,** 61–65.
2. Kvasnicka, F. *Journal of Chromatography B*, **2003**, *787*, 77-89.
3. Tyers, M.; Mann, M. *Nature*, **2003**, *422*, 193-197.
4. Sriyam, S.; Sinchaikul, S. Tantipaiboonwong, P.; Tzao, C.; Phutrakul, S.; Chen, S. T. *Journal of Chromatography B*, **2007**, *849*, 91-104.
5. Pandey, A.; Mann, M. *Nature*, **2000**, *405***,** 837–846.
6. Wilkins, M. R.; Sànchez, J. C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. *Biotechnol. Genet. Eng. Rev.*, **1996**, *13*, 19–50.
7. Xiao, Z.; Prieto, D.; Conrads, T.P.; Veenstra, T.D.; Issaq, H. J. *Mol. Cell. Endocrinol.*, **2005**, *230*, 95-106.
8. Chen, R.; Pan, S.; Brentnall, T.A.; Aebersold, R. *Mol. Cell. Proteomics*, **2005**, *4*, 523-533.
9. Challapalli, K.K.; Zabel, C.; Schuchhardt, J.; Kaindl, A.M.; Klose, J.; Herzel, H. *Electrophoresis*, **2004**, *25*, 3040- 3047.
10. Paweletz, C.P.; Wiener, M.C.; Sachs, J.R.; Meurer, R.; Wu, M.S.; Wong, K.K.; Yates, N.A.; Hendrickson, R.C. *Proteomics*, **2006**, *6*, 2101-2107.
11. Lin, S.M.; Haney, R.P.; Campa, M.J.; Fitzgerald, M.C.; Patz, E.F. *Cancer Inform.*, **2005**, *1*, 32-40.
12. Beer, I.; Barnea, E.; Ziv, T.; Admon, A. *Proteomics*, **2004**, *4*, 950-960.
13. Veraart, J.R.; Lingeman, H.; Brinkman, U. A. Th. *J. Chromatogr.A*, **1999**, *856*, 483-514.
14. Nielsen, M.L.; Savitski, M. M.; Zubarev, R. A. *Mol. Cell. Proteomics*, **2005**, *4*, 835-845.
15. Fields, S.; *Science*, **2001**, *291*, 1221–1224.
16. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C. *Nature*, **2001**, *409***,** 860–921.
17. Carbonaro, M. *Trends in Food Science & Technology*, **2004**, *15*, 209–216.
18. Ogorzalek-Loo, R.R.; Loo, J.A.; Andrews, P.C. *Methods Mol. Biol.*, **1999**, *112*, 473-485.
19. Ogorzalek-Loo, R.R.; Cavalcoli, J.D.; VanBogelen, R.A.; Mitchell, C.; Loo, J.A.; Moldover, B.; Andrews, P.C. *Anal. Chem.*, **2001**, *73* (17), 4063-4070.

20. Jensen, P.K.; Pasa-Toli, L.; Anderson, G.A; Horner, J.A.; Lipton, M.S.; Bruce, J.E.; Smith, R.D. *Anal. Chem.*, **1999**, *71* (11), 2076-2084.
21. Pappin, D.J.C.; Hojrup, P.; Bleasby, A. *Curr. Opin. Biotechnol.*, **1993**, *3*, 327-332.
22. Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.*, **1993**, *22* (6), 338-345.
23. Jensen, O.N.; Podtelejnikov, A.; Mann, M. *Rapid Commun. Mass Spectrom.*, **1996**, *10* (11), 1371-1378.
24. Zellner, M.; Winkler, W.; Hayden, H.; Diestinger, M.; Eliasen, M.; Gesslbauer, B.; Miller, I.; Chang, M.; Kungl, A.; Roth, E.; Oehler, R. *Electrophoresis*, **2005**, *26*, 2481-2489.
25. Jiang, L.; He, L.; Fountoulakis, M. *J. Chromatogr. A*, **2004**, *1023*, 317-320.
26. Manza, L.L.; Stamer, S.L.; Ham, A. J. L.; Codreanu, S.G.; Liebler, D. C. *Proteomics*, **2005**, *5*, 1742.
27. Tirumalai, R.S.; Chen, K. C.; Prieto, D.A.; Issaq, H.J.; Conrads, T.P.; Veenstra, T.D. *Mol. Cell. Proteomics*, **2003**, *2*, 1096-1103.
28. Haper, R. G.; Workman, S. R.; Schuetzner, S.; Timperman, A. T.; Sutton, J. N. *Electrophoresis*, **2004**, *25*, 1299-1306.
29. Echan, L. A.; Tang, H. Y.; Ali-Khan, N.; Lee, K. B.; Speicher, D.W. *Proteomics*, **2005**, *5*, 3292-3303.
30. Mehta, A. I.; Ross, S.; Lowenthal, M. S.; Fusaro, V.; Fishman, D. A.; Petricoin, E.; Liotta, F. *Dis. Mark.*, **2003**, *19*, 1-20.
31. Huang, H.L.; Stasyk, T.; Morandell, S.; Mogg, M.; Schreiber, M.; Feuerstein, I.; Huck, C.W.; Stecher, G.; Bonn, G. K.; Huber, L. A. *Electrophoresis*, **2005**, *26*, 2843-2849.
32. Steel, L. F.; Trotter, M. G.; Nakajima, P. B.; Mattu, T. S.; Gonye, G.; Block, T. *Mol. Cell. Proteomics,* **2003**, *2*, 262-270.
33. Gaberc-Porekar, V.; Menart, V. *J. Biochem. Biophys. Methods*, **2001**, *49*, 335-360.
34. Kange, R.; Selditz, U.; Granberg, M.; Lindberg, U.; Ekstrand, G.; Ek, B.; Gustafsson, M. *J. Biomol. Tech.*, **2005**, *16*, 91-103.
35. Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Process*., **1987**, *78*, 53-68.
36. Karas, M.; Hillenkamp, F. *Anal. Chem.*, **1988**, *60*, 2291-2301.
37. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science*, **1989**, *246*, 64-67.
38. Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. N. J. *Chem. Soc. Commun.,* **1981**, 325-327.
39. Thiede, B.; Hohenwarter, W.; Krah, A.; Mattow, J.; Schmid, M.; Schmidt, F.; Jungblut, P. R. *Methods*, **2005**, *35*, 237-247.
40. Shevchenko, A.; Wilm, M.; Vorm, O.; Jensen, O. N.; Podtelejnikov, A. V.; Neubauer, G.; Mortensen, P.; Mann, M. *Biochem. Soc. Trans.*, **1996**, *24*, 893-896.
41. Hewick, R. M.; Hunkapiller, M. W.; Hood, L. E.; Dreyer, W. J. *J. Biol. Chem.* **1981**, *256*, 7990-7997.
42. Pappin, D. J. *Methods Mol. Biol.* **1997**, *64*, 165-173.

43. Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 1791-1814.
44. Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741-4750.
45. Eriksson, J.; Chait, B. T.; Fenyo, D. *Anal. Chem.* **2000**, *72*, 999-1005.
46. Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044-1050.
47. Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 1998-2003.
48. Zubarev, R. A.; Hakansson, P.; Sundqvist, B. *Anal. Chem*. **1996**, *68*, 4060-4063.
49. Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem*. **1993**, *65*, 425-438..
50. Harrison, A. G.; Csizmadia, I. G.; Tang, T. H. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 427-436.
51. Yates, J. R., III; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem.* **1995**, *67*, 1426-1436.
52. Rodrìguez, R.; Villalba, M.; Batanero, E.; Gónzales, E. M.; Monsalve, R. I.; Huecas S., Tejera M.L., Ledesma A., *Allergy*, **2002**,*57:Suppl:71*, 6-16.
53. Mann M., Hojrup P., Roepstorff P.**,** *Biol. Mass Spectrom*., **1993**, *22*, 338-345.
54. Pappin D.J.C., Hojrup P., Bleasby A., *J. Curr. Biol*., **1993**, *3*, 327-332.
55. Hillenkamp F., Karas M., Beavis R.C., Chait B.T., *Analytical Chemistry*, **1991**, *63*,1193A-1203A.
56. Aiello D., Di Donna L., Napoli A., Sajjad A., Perri E., Sindona G. *Anal. Chem.* **2006**, *78* (10), 3434-3443.
57. Aiello D., Di Donna L., Prendushi H., Napoli A., Sindona G. *Ana. Chem.*, **2007,** *79*, 5941-5948.
58. Thiellement H., Zivy M., Plomion C., *Journal of Chromatography B*, **2002**, *782*, 137-149.
59. Tsugita A., Kamo M., Kawakami T., Ohki Y., *Electrophoresis*, **1996**, *17*, 855-865.
60. Bahrmann N., Petit R., *J. Mol. Evol.,* **1995**, *41*, 231-237.
61. Friedho L.R., Ehrlich-Kautzky E., *J. Allergy Clin. Immunol.*, **1986**, *78*, 1190–1201.
62. Florido JF, Delgado PG, de San Pedro BS, Quiralte J, de Saavedra JM, Peralta, *Int Arch Allergy Immunol*, **1999**;*119*, 133–137.
63. D'Amato G. Lobefalo G., *J Allergy Clin Immunol*, **1989**, *83*, 116–122.
64. Gioulekas D, Papakosta D, Damialis A, Spieksma F, Giouleka P, Patakas D., *Allergy* **2004**, *59*, 174–184.
65. Kirmaz C, Yuksel H, Bayrak P, Yilmaz O., *J Invest Allergol Clin Immunol*, **2005**, *15*, 140–145.
66. Liccardi G, D'Amato M, D'Amato G., *Int Arch Allergy Immunol*, **1996**, *111*, 210–217.
67. Kwon, S. W. *J. Agric. Food Chem.* **2004**, *52*, 7258-7263.
68. Li, Q., Li, L.; Rejtar, T.; Karger, B. L.; Ferry, J. G. *J. Proteome Res*. **2005**, *4*, 129-135.
69. Di Donna L.; Maiuolo, L.; Mazzotti, F.; De Luca, D.; Sindona, G. *Anal. Chem.* **2004**, *17*, 5104-5108.

70. De Nino, A.; Di Donna, L.; Mazzotti, F.; Muzzalupo, E.; Perri, E.; Sindona, G.; Tagarelli, A. *Anal. Chem.* **2005**, *77*, 5961-5964.
71. Bousquet J.; Cour, P.; Guerin, B.; Michel, F. B. *Clin. Allergy* **1985**, *14*, 249-258.
72. Bousquet, J.; Guerin, B.; Hewitt, B.; Lim, S.; Michel, F. B. *Clin. Allergy* **1985**, *15*, 439-448.
73. Villalba , M.; Lope´ z-Otý`, C.; Martin-Orozco, E.; Monsalve, R. I.; Palomino, P.; Lahoz, C.; Rodrìguez, R. *Biochem. Biophys. Res. Commun.* **1990**, *172*, 523 528.
74. Blanca, M.; Boulton, P.; Brostoff, J.; Gónzalez-Reguera, I. *Clin. Allergy* **1983**, *13*, 473-478.
75. Ledesma, A.; Rodrìguez, R.; Villalba, M. J. *Allergy Clin. Immunol*. **1996**, *98*, 805-815.
76. Rubio, N.; Brieva, A.; Alca`zar, B. *J. Chromatogr.* **1987**, *403*, 312-318.
77. Lauzurica, P.; Giurbino, C.; Maruri, N. *Mol. Immunol*. **1988**, *25*, 329-335.
78. Boluda, L.; Sastre, J.; Casanovas, M.; Ferna`ndez-Caldas, E. J. *Immunol. Methods* **1999**, *223*, 17-26.
79. Martìnez, A.; Asturiaa, J. A.; Placios, R.; Sanz, M. L.; Sa`nches, G.; Oehling, A.; Martìnez, J. *Allergy* **1999**, *54*, 584-592.
80. Baldo, B. A.; Panzani, R. C.; Bass, D.; Zerboni, R. *Mol. Immunol.* **1992**, *29*, 1209-1218.
81. Lombardero, M.; Quirce, S.; Duffort, O.; Barber, D.; Carpizo, J.; Chamorro, M. J.; Lezaun, A.; Carreira, J. *J. Allergy Clin. Immunol*. **1992**, *89*, 884-887.
82. Barber, D.; Carpizo, J.; Garcìa-Rumbao, M. C.; Polo, F.; Juan, F. *Ann. Allergy* **1990**, *64*, 43-46.
83. Waisel, Y.; Geller-Bernstein, C.; Keynan, N.; Arad, G. *Allergy* **1996**, *51*, 819-825.
84. Conde Hernàndez, J.; Conde Hernàndez, P.; Gónzalez Quevedo Tejerína, M. T.; Conde Alcaniz, M. A.; Conde Alcaniz, E. M.; Crespo Moreira, P. *Allergy*, **2002**, *71*, 60–65.
85. Vrtala, S.; Grote, M.; Duchene, M. *Int. Arch. Allergy Immunol.* **1995**, *107*, 160-169.
86. Carne`s, J.; Fernandez-Caldas, E.; Boluda, L.; Casanovas, M.; Sastre, J.; Lluch Bernal, M.; Blanca, M. *Allergy* **2002**, *57*, 798-804.
87. Rajnarayanan, R. V.; Wang, K. *J. Mass Spectrom*. **2004**, *39*, 79-85.
88. Tejera, M. L.; Villalba, M.; Batanero, E.; Rodrìguez, R. *J. Allergy Clin. Immunol*. **1999**, *104*, 797-802.
89. Batanero, E.; Ledesma, A.; Villlalba, M.; Rodrìguez, R. *FEBS Lett.* **1997**, *410*, 293-296.
90. Barral, P.; Tejera, M. L.; Trevino, M. A.; Batanero, E.; Villalba, M.; Bruix, M.; Rodrìguez, R. J. *Protein Expression Purif.* **2004**, *37*, 336-343.
91. Trevino, M. A.; Garcìa-Mayoral, M. F.; Barral, P.; Villalba, M.; Santoro, J.; Rico, M.; Rodrìguez, R.; Bruix, M. *J. Biol. Chem.* **2004**, *279*, 39035-39041.
92. Martìnez, A.; Asturias, J. A.; Monteseirìn, J.; Moreno, V.; Garcìa-Cubillana, A.; Hernàndez, M.; de la Calle, A.; Sànchez-Hernàndez, C.; Pérez-Formoso, J. L.; Conde, J. *Allergy* **2002**, *57*, 17-23.

93. Asturias, J. A.; Arilla, M. C.; Gomez-Bayon, N.; Martìnez, J.; Martìnez, A.; Palacios, R. *J. Allergy Clin. Immunol*. **1997**, *100*, 365-372.
94. Ledesma, A.; Villalba, M.; Batanero, E.; Rodrìguez, R. *Eur. J. Biochem*. **1998**, *258*, 454-459.
95. Villalba, M.; Batanero, E.; Lc¸pez-Otý`n, C.; Sa`nchez, L. M.; Monsalve, R. I.; Gónzalez de la Pena, M. A.; Lahoz, C.; Rodrìguez, R. *Eur. J. Biochem.* **1993**, *216*, 863-869.
96. Lombardero, M.; Barbas, J. A.; Moscoso del Prado, J.; Carreira, J. *Clin. Exp. Allergy* **1994**, *24*, 765-770.
97. Villalba, M.; Batanero Monsalve, R. I.; Gónzalez de la Pena, M. A.; Lahoz, C.; Rodrìguez, R. *J. Biol. Chem.* **1994**, *269*, 15217-15222.
98. Lombardero, M.; Obispo, T.; Calabozo, B.; Lezaúan, A.; Polo, F.; Barber, D. *Allergy* **2002**, *57*, 29-34.
99. Manso, M. A.; Léonil, J.; Jan, G.; Gagnaire, V.; *International Dairy Journal*, **2005**, *15*, 845–855.
100. Washburn, M. P.; Yates, J. R. III *Current Opinion in Microbiology*, **2000**, *3*, 292–297.
101. Aslam, M.; Jiménez-Flores, R.; Kim, H. Y.; Hurley, W. L. *Journal of Dairy Science*, **1994**, *77*, 1529–1536.
102. Roncada, P.; Gaviraghi, A.; Liberatori, S.; Canas, B.; Bini, L.; Greppi, G. F. *Proteomics*, **2002**, *2*, 723–726.
103. Marino, R.; Considine, T.; Sevi, A.; Kelly, A. L.; McSweeney, P. L. H. *Int. Dairy J.* **2005**, *15*, 1026-1033.
104. Eriksson, A; Persson Wallerb, K.; Svennersten-Sjaunja, K.; Haugen, J.-E.; Lundby, F.; Lind, O. *Int. Dairy J.* **2005**, *15*, 1193-1201.
105. Bastian, E. D.; Brown, R. J. *Int. Dairy J.* **1996**, *6*, 435-457.
106. Nielsen, S. S. *J. Agric. Food Chem.* **2002**, *22*, 6628-6634.
107. Kaminogawa, S.; Yamauchi, K. *Agric. Biol. Chem.* **1972**, *36*, 2351-2356.
108. Larsen, L. B.; Benfeldt, C.; Rasmussen, L. K.; Petersen, T. E. *J. Dairy Res.,* **1996**, *63*, 119-130.
109. Dulley, J. R. *J. Dairy Res.* **1972**, *39*, 1-9.
110. Farkye, N. Y.; Fox, P. F. *J. Dairy Res.* **1990**, *57*, 413-418.
111. Farkye, N. Y.; Fox, P. F. *J. Agric. Food Chem.* **1991**, *39*, 786-788.
112. Farkye, N. Y.; Fox, P. F. *J. Agric. Food Chem.* **1992**, *59*, 209-224.
113. Larsen, L. B.; Wium, H.; Benfeldt, C.; Heegaard, C.; Ardo, Y.; Qvist, K. B.; Petersen, T. E. *International Dairy Journal* **2000**, *10*, 67-74.
114. Munro, G. L.; Grieve, P. A.; Kitchen, B. J. *Australian J. Dairy Technol.* **1984**, *39*, 7-16.
115. Auldist, M. J.; Hubble, I. B. *Australian J. Dairy Technol.* **1998**, *53*, 28-36
116. Grohn, Y. T.; Eicker, S. W.; Ducrocq, V.; Hearti , J. A. *J. Dairy Sci* **1988**, *81*, 966-978.

117. Sordillo, L. M.; Streicher, K. L. *J. Mammary Gland Biol. Neoplasia*, **2002**, *7*, 370-380.
118. Rainard, P. *Nature Biotechnology*, **2005**, *23*, 430-432.
119. Mehrzad, J.; Dosogne, H.; Meyer, E.; Burvenich. C. *Vet. Res.*, **2001**, *32*, 131–144.
120. O'Driscoll, B. M.; Rattray, F. P.; McSweeney, P. L. H.; Kelly, A. L. *Journal of Food Science*, **1999**, *64*, 606–611.
121. Hurley, M. J.; Larsen, L. B.; Kelly, A. L.; McSweeney, P. L. H. *International Dairy Journal*, **2000**, *10*, 673–681.
122. Considine, T.; Geary, S.; Kelly, A. L.; McSweeney, P. L. H. *Food Chem*. **2002**, 76, 59–67.
123. Considine, T.; Healy, A.; Kelly, A. L.; McSweeney, P. L. H. *Food Chem.* **2000**, *69*, 19–26.
124. McSweeney, P. L. H.; Fox, P. F *International Dairy Journal* **1995**, *5*, 321–336.
125. Magboul, A. A. A.; Larsen, L. B.; Kelly, A. L.; McSweeney, P. L. H. *International Dairy Journal*, **2001**, *11*, 865–872.
126. Considine, T.; Healy, A.; Kelly, A. L.; McSweeney, P. L. H. *International Dairy Journal*, **2004**, *14*, 117–124.
127. Morimura, S.; Nagata, H.; Uemura, Y.; Fahmi, A.; Shigematsu, T.; Kida, K. *Process Biochemistry* **2002**, *37* 1403–1412.
128. Mank, M.; Stahl, B.; Boehm, G. *Anal. Chem.* **2004**, *76*, 2938-2950.
129. De Dios Alche, J.; Castro, A. J.; Olmedilla, A.; Fernandez, M., C.; Rodrìguez, R.; Villalba, M.; Rodrìguez-Garcìa, M. I. *Jounal of Cell Science* **1999**, *112*, 2501-2509.
130. Boluda, L.; Sastre, J.; Casanovas, M.; Fernàndez-Caldas, E. *J. Immunol. Method*, **1999**, *223*, 17-26.
131. Van Ree, R.; Cabanes-Macheteau, M.; Akkerdaas, J.; Milazzo, J. P.; Loutelier-Bourhis, C.; Rayon, C.; Villalba, M.; Koppelman, S.; Aalberse, R.; Rodrìguez, R.; Faye L.; Lerouge P *J. Biol. Chem.* **2000**; *275*, 11451 – 11458.
132. Wilson, I. B. H. D.; Altman, F *Glycoconjugate J.*, **1998**, *15*, 1055-1070.
133. Lomardero, M.; Barbas, J. A.; Moscoso del Prado, J.; Carreira, J. *Clin. Exp. Allergy* **1994**, *24*, 765-770.
134. Jensen, O. N . *Curr. Opin. Chem. Biol.* **2004**, *8*, 33-41.
135. Gorg, A.; Weiss, W.; Deeth, H. C.; Alewood, P. F. *Proteomics* **2004**, *4*, 743-752.
136. Lauber, W. M.; Carroll, J. A.; Dufield, D. R.; Kiesel, J. R.; Radabaugh, M. R.; Malone, J.P. *Electrophoresis* **2001**, *22*, 906-918.
137. Gygi, S.P.; Corthals, G. L.; Zhang, Y.; Aebersold, R. *Proc. Natl Acad Sci* **2000**, *97*, 9390-9395.
138. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat Biotecnol.* **1999**, *17*, 676-682
139. Carnès, J.; Fernandez- Caldas, E.; Boluda, L.; Casanovas, M.; Sastre, J.; Lluch Bernal, M.; Blanca, M. *Allergy* **2002**, *57*, 798-804.
140. Dell, A.; Morris, H.R. *Science* **2001**, *291*, 2351-2356.
141. Holland, J. W.; Deeth, H. C.; Alewood, P.F. *Proteomics* **2006**, *6*, 3087-3095.

142. Kolarich, D.; Weber, A.; Turecek, P.; Schwarz, H. P.; Altmann, F. *Proteomics* **2006**, *6*, 3369-3380.
143. Rayon, C.; Lerouge, P.; Faye, L. *J. Experimental Botany* **1998**, *49*, 1463-1473.
144. Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. *Journal of Chromatography B* 200, *7849*, 115–128.
145. Roth, K. D. W.; Huang, Z. H.; Sadagopan, N.; Watson, J. T. *Mass Spectrometry Rev.* **1998**, *17*, 255-274.
146. Park, S.J.; Song, J. S.; Kim, H. J. *Rapid Commun. Mass Spectrom*. **2005**, *19*, 3089-3096.
147. Altmann, F. *Glycoconj. J.* **1998**, *15*, 79-82.
148. Mechref, Y.; Novotny, M. V.; Krishnan, C. *Anal. Chem* **2007** ,00,0000 (web available).
149. Carne´s, J.; Ferna`ndez-Caldas, E. *Allergy* **2002**, *57* (71), 24-28.