Università degli Studi di Padova

Centro di Ricerca Interdipartimentale
per le Biotecnologie Innovative
CRIBI

Dottorato di Ricerca
in
BIOCHIMICA E BIOTECNOLOGIE
Ciclo XX

Analysis of Allergenic Proteins by Mass Spectrometry

Coordinatore: Prof. Giuseppe ZANOTTI

Supervisore: Prof. Angelo FONTANA

Dottoranda: Francesca BATTAGLIA

31 Gennaio 2008

FOREWARD

The general aim of the activities conducted in the framework of my PhD were to learn modern techniques for analyzing proteins by mass spectrometry (MS) and then to apply these techniques and approaches to the analysis of allergenic proteins contained in foods. My research activities were conducted at the Laboratory of Protein Chemistry of CRIBI, University of Padua, where previously I have conducted the research for my Thesis for the Doctor degree in Pharmaceutical Biotechnologies. During the fist year of my PhD I have concluded the Thesis project on the amyloid aggregation of α-lactalbumin, a model protein utilized for investigating molecular aspects of protein amyloidogenesis. The results of this research were quite interesting and indeed they have been published in an international journal.

During the first two years I acquired a solid knowledge on several aspects of the MS methodology and I was able to learn the theory and practice of several modern techniques and approaches in this ambit. The specific aim was to analyze allergenic proteins contained in complex matrices as foods and to this aim several proteins were extracted and purified from several food samples. The research has been focused mostly on the allergenic proteins from milk and eggs, known to cause widespread allergies. The proteins of interest were analyzed by using several chromatographic and electrophoretic techniques and also by means of HPLC connected to a tandem MS electrospray instrument. I was able to show that MS techniques can be used to identify allergenic proteins even when contained in very complex mixtures. Therefore, these MS techniques perhaps can be used as an alternative to the immunochemical methods nowadays in use for detecting allergens. I have also analyzed the chemical modifications that allergenic proteins suffer during several industrial treatments of foods, including heat treatment.

During the third year of my PhD I spent a six months period at the Biochemistry Laboratory of the Imperial College in London, being involved in a project aimed to study in a large scale the proteins of the mosquito *Anopheles gambiae*. The MS analyses were focused on the proteins responsible of the mating behaviour of *A. gambiae*, hoping to identify a target for controlling the behaviour of this vector of the malaria disease.

Summing up, besides the publication dealing with amyloid aggregates of α-lactalbumin, this PhD Thesis is composed by a major part dealing with MS analysis of allergenic proteins and by a minor one dealing with MS analysis of proteins from *A. gambiae.*

Francesca Battaglia

PREMESSA

L'obiettivo generale del mio corso di dottorato è stato quello di apprendere le moderne tecniche di analisi di proteine mediante spettrometria di massa (MS) e di applicarle per l'identificazione ed analisi di proteine allergeniche contenute in alimenti. Le ricerche sono state svolte presso il Laboratorio di Chimica delle Proteine del CRIBI dell'Università di Padova, dove era stata svolta in precedenza la Tesi sperimentale per la laurea specialistica in Biotecnologie Farmaceutiche. Durante il primo anno di dottorato ho portato a termine anche il progetto della mia Tesi di Laurea riguardante l'aggregazione amiloide di α-lattalbumina, una proteina utilizzata come modello per lo studio dell'importante problema dell'aggregazione di tipo amiloide delle proteine. I risultati di questo lavoro sono stati particolarmente interessanti e sono stati anche pubblicati in una rivista internazionale.

Nei primi due anni di dottorato ho approfondito le mie conoscenze sulla spettrometria di massa ed ho acquisito una buona esperienza sia teorica che pratica in questo ambito. L'obiettivo specifico è stato poi quello di analizzare proteine allergeniche e per questo scopo sono state estratte e purificate da alimenti varie proteine allergeniche. La ricerca si è focalizzata su alcune proteine soprattutto del latte e dell'uovo, note per essere causa di allergie molto diffuse. Le proteine di interesse sono state analizzate con varie tecniche cromatografiche ed anche mediante un sistema HPLC accoppiato a tandem MS *electrospray* (LC-MS/MS ESI). Successivamente ho utilizzato tecniche MS per l'analisi delle modifiche chimiche che alcune proteine allergeniche subiscono durante vari trattamenti industriali, in particolare durante i processi termici.

Durante il terzo anno di dottorato e per sei mesi ho svolto attività di ricerca presso il Laboratorio di Biochimica dell'Imperial College di Londra, dove ho collaborato ad un progetto volto all'analisi su larga scala delle proteine della zanzara *Anopheles gambiae*, responsabile della malaria. Lo scopo di queste analisi è stato quello di identificare potenziali *target* potenzialmente utili per lo sviluppo di strategie di controllo del comportamento della specie *A. gambiae*.

In sintesi, oltre alla pubblicazione sugli aggregati amiloide d1 α-lattalbumina, questa Tesi di dottorato è composta da una parte principale riguardante l'analisi di proteine allergeniche mediante tecniche MS e da una parte di minori dimensioni riguardante l'analisi delle proteine della zanzara *A. gambiae*.

Francesca Battaglia

Analysis of Allergenic Proteins by Mass Spectrometry

ABBREVIATIONS

AGEs	Advance glycation products
BSA	Bovine serum albumin
BTP	Bis-tris propane
CEL	N-Carboxyethyl-lysine
CID	Collision induce dissociation
CML	N-Carboxymethyl-lysine
DTT	Dithiothreitol
E:S	Enzyme to substrate ratio
ELISA	Enzyme-Linked Immunosorbent Assay
ESI	Electrospray ionization
FPLC	Fast Protein Liquid Chromatography
FUR	Furosine
GA	Glyoxal
GalNAc	N-Acetyl galactosamine
GlcNAc	N-Acetylglucosamine
HMF	5-Hydroxymethyl-2-furfuraldehyde
HPLC	High-pressure liquid chromatography
IAA	Iodoacetamide
IEX	Ion exchange chromatography
IgE	Immunoglobulin E
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
Lys	Hen egg-white lysozyme
m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MG	Methylglyoxal
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
NaH_2PO_4	Sodium monobasic phosphate
NH_4HCO_3	Ammonium bicarbonate
NMR	Nuclear magnetic resonance

OA	Ovalbumin
OM	Ovomucoid
Q-Tof	Quadrupole-Tof
RP	Reverse-phase
HPLC	High-pressure liquid chromatography
Rpm	Revolutions per minute
RT	Retention time
SDS	Sodium dodecyl sulphate
PAGE	Polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	Tetra(methylethylene)diamine
TFA	Trifluoroacetic acid
Tic	Total ion current
Tof	Time-of-flight
Tris	Tris(hydroxymethyl)aminomethane
UV	ultraviolet
w/v	Weight/volume
α-LA	α-Lactalbumin
α_{s1}-CN	α_{s1}-Casein
α_{s2}-CN	α_{s2}-Casein
β-CN	β-Casein
β-LG	β-Lactoglobulin
κ-CN	κ -Casein

ABBREVIATIONS OF AMINO ACIDS

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cisteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glicine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lisine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

SUMMARY

Allergy is a widespread disease affecting millions of people all over the world and food is the main cause of this disease. The allergic reaction can be caused by an allergen that can be both of both natural and synthetic origin. However, proteins are the most common food allergens and often they can induce an immediate allergic reaction that can have severe consequences in sensitive subjects. The only way to escape allergy is by avoiding the specific food or some of its ingredients. Therefore, it is very important to know all the components present in a food, even if in traces, since also a minute amount of an allergenic protein can cause the symptoms of allergy. Proteins from cow's milk, chicken egg, soybean and peanuts are the most frequent food allergens contained in the complex foods prepared by industrial processes. Consequently, there is an urgent need of specific and reliable methods capable to detect protein allergens in foods. Presently, the immunochemical methods based on the ELISA test are being used to detect specific allergens, but theses techniques, however, suffer from several limitations and, therefore, there is a need for reliable alternative analytical methods.

The aim of this PhD project was to develop methods capable of detecting and analyzing allergenic proteins in foods by using modern techniques of mass spectrometry (MS). The main initial problem was to prepare protein samples suitable for a direct analysis by MS. Chromatographic and electrophoretic techniques were used to separate proteins contained in several food extracts, the mostly investigated being proteins from bovine milk and eggs. First, we used samples of pure homogeneous proteins in order to optimize the MS techniques and approaches, including accurate mass analysis, fingerprinting, sequencing by tandem MS and database searches. The possibility to detect protein allergens in a whole protein extract was demonstrated by identifying a specific protein allergen in extracts of both peanuts and soybean. It is shown that the knowledge of the amino acid sequence of a 10-12 residue peptide fragment of a protein, as obtained by its tryptic digestion, followed by search of databases of protein sequences allows the identification of that protein with high confidence.

Proteins in commercial foods are usually subject to a variety of chemical and physicochemical modifications due to thermal processing, cooking, mixing, colouring, emulsifying and homogenization. These procedures can cause modifications of the covalent and three dimensional structures of proteins and they can lead to the onset of antigenic properties in otherwise native, innocuous proteins. Thermal processes can have a dramatic

effect on protein allergy, but thermal processes are unavoidable in practice for food preparation. Nevertheless, it is known that heating can cause also a decrease of the allergenic potency of foods. Therefore, it is clear that the analysis of heat-mediated protein modifications is extremely important for studying protein allergy and for detecting protein allergens in foods. Here, we have investigated the effects of heating of model proteins in the presence of sugar and the modifications caused thereof. In particular, we analyzed the proteins contained in a home-made baked food and thus being composed of a variety of potential allergenic proteins of different origins. The MS analysis permitted an accurate study of the fate of allergenic proteins before and after cooking. We were able to identify proteins which are resistant to the heating process, as well as to identify the main chemical and physicochemical protein modifications caused by cooking.

Summing up, in this PhD Thesis we have established that MS techniques can be applied for identifying allergenic proteins contained in complex matrices of foods. The advantages of these techniques include, besides the requirement of minute amounts of protein sample, the possibility to identify proteins even after their damaging in terms of covalent structure and conformation, thus when immunochemical methods of analysis are expected to fail.

I. INTRODUCTION

1. Allergic reactions to foods

Adverse reactions to food represent a prominent and actual problem all over the world (Teufel *et al.*, 2007). A food allergy is an adverse reaction to a food or a food component, usually a protein, involving the body immune system (David, 2000). The intolerance or the allergy to food is a phenomenon that is increasing year by year, according to the significant changes in our feeding habits. The symptoms of allergy are usually skin reactions, respiratory problems, urticaria, bronchocostriction, as well as gastrointestinal and cardiovascular symptoms. The last and maximal manifestation is the anaphylactic shock, which can cause the death of a sensitive subject (Ring *et al.*, 2001). The diagnosis of food allergy is done by history, skin tests and *in vitro* allergy diagnosis. Even a trace amount of a food allergen can cause the adverse reactions. Patients who suffer from a food allergy can only avoid the respective allergen, even though it is not always possible, due to the fact that usually the ingredients in a food are not known in full (Businco *et al.*, 1999). Nowadays, the food industry is producing even more complex foods deriving from different foodstuffs not always listed in the label. Milk allergy is the most common food allergy (Skripak *et al.*, 2007) (Fig.1). It consists in an immunological mediated adverse reaction to one or more cow's milk proteins, which are normally harmless to a non-allergic individual (Wal 2004).

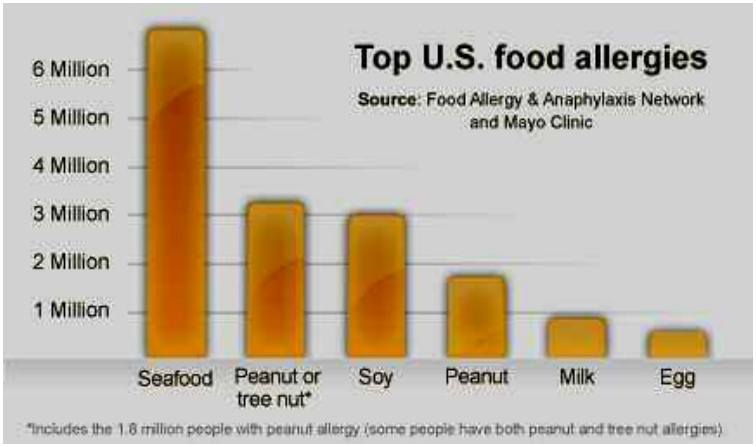


Fig. 1. Incidence of food allergy in USA.

Allergens and antibodies are specific for a food, but sometimes they can be involved in cases of cross-reactivity, where different foods have near identical allergens. The severity and location depend on the quantity of food eaten, form of the food, food processing and quantity of histamine and other chemicals released. The immediate reaction, called

sensitization, is usually IgE mediated and occurs when allergens cross the gastrointestinal barrier initiating a chain of reactions (Bischoff *et al.*, 2000). In order to shed light into the problem of food allergy, nowadays the research aims are to elucidate the physiopathology of food allergy and to improve allergen detection strategies. There are several allergy tests in current use for analyzing the specific allergy of a patient, but at present no cure is available for food allergies.

2. Allergenic proteins and their structural features

Food allergens are almost always proteins, but of course not all food proteins are allergens. All proteins adopt a specific three-dimensional (3D) structure that is of paramount importance for their function (Besler *et al.*, 2001). The primary structure is the amino acid sequence of the peptide chains, the secondary structure creates highly regular sub-structures (*alpha helix* and *strands of beta sheet*) by their patterns of hydrogen bonds between the main-chain peptide groups. The tertiary structure is the three-dimensional structure (3D) of a single protein molecule, the spatial arrangement of the secondary structures; the quaternary structure is the complex of several protein molecules or polypeptide chains.

Proteins are synthesized by ribosomes associated with the membranes of the endoplasmic reticulum (ER) and many of them undergo proteolytic cleavage following translation. A variety of proteins are synthesized as inactive precursors which are activated under proper physiological conditions by limited proteolysis or they can undergo different post-translational modifications by the action of specific enzymes. In many cases, the N-terminal methionine is hydrolyzed and an acetyl group is added to the new N-terminal amino acid. Glycosylation is a well known protein modification mediated by specific enzymes and the resulting glycoproteins consist of proteins covalently linked to carbohydrates. The predominant sugars in glycoproteins are glucose, galactose, mannose, fucose, GalNAc and GlcNAc. The N-glycosydic linkage is through the amide group of N, the O-glycosidic linkage is to the hydroxyl of Ser, Thr or hydroxylysine. Also some specific amino acids can be selectively modified. For example, methylation occurs at Lys residues in some proteins. Post-translational phosphorylation is one of the most common protein modifications that occur in animal cells. The vast majority of phosphorylations occurs as a mechanism to regulate the biological activity of a protein and as such is transient. Ser, Thr and Tyr are the amino acids subject to phosphorylation in animal cells.

Proteins which are allergens are often modified and food allergens have several biochemical characteristics in common (Takagi *et al.*, 2005). Often these include

glycosylation patterns, stability to proteases, as well as to heat and protein denaturants. Their structure can be globular and very compact, but also partially unfolded and with a low amount of structure. The most peculiar and significant characteristic of a protein allergen is stability to the proteolytic and acidic conditions of the digestive tract. This imparts to the protein an increased probability of reaching the intestinal mucosa, where absorption can occur. Indeed, several allergenic proteins have been tested for their stability to simulated gastric fluid (in the presence of pepsin at pH 1) and were shown to survive for extended periods of time when compared with other non allergenic food proteins. Even though protein stability has been demonstrated for a variety of protein allergens, the molecular mechanism underlying this enhanced stability is not known (Sharma *et al.*, 2001).

Often children suffer from food allergy and they may experience extreme allergic reactions including asthma attacks and hives (Fujita *et al.*, 2007). Many protein allergens are not destroyed even by cooking and can trigger an allergic reaction. Peanut proteins Ara h 1 and Ara h 2 are two of the strongest and most studied allergens. Studies have shown that higher order structures are largely responsible for their resistance to enzymatic digestion and therefore the persistence of their allergenicity (Kopper *et al.*, 2004). Egg, peanut and soybean allergies are based on hypersensitivity to dietary proteins from egg, peanuts and soybean (Plaza-Martín *et al.*, 2007; Sicherer et al. 2000). It causes an overreaction of the immune system which may lead to severe physical symptoms for millions of people.

3. Analysis of allergic proteins

Immunochemical methods. Allergen proteins can be detected by serological methods, like enzyme immunoassay or *in vivo* skin test procedures (Peterson *et al.*, 2007), but the problem of the diagnosis of a food allergy is further complicated by the cross-reactivity between allergens from different foods and between allergens in food and aeroallergens, like pollen or low molecular mass molecules (Vieths *et al.*, 2002). A method largely used nowadays is Enzyme-linked Immunosorbent Assays (ELISA) (L'Hocine *et al.*, 2007). ELISA tests are widely utilized to detect substances that have antigenic properties, primarily proteins, this technique combines the specificity of antibodies with the sensitivity of a simple enzyme assays. In the "indirect" ELISA, samples of known antigen concentrations are used to derive a standard curve in order to calculate antigen concentration in samples. Other surfaces are then coated with serum samples of unknown antigen concentration. A concentrated solution of non-interacting protein, like bovine serum albumin, is added to all plate wells to block the non-specific absorption of other proteins. A detection antibody, specific to the antigen of

interest is applied to all plate wells, which will only bind to immobilized antigen on the well surface, not to other serum proteins or the blocking proteins. Secondary antibodies, which will bind to any remaining detection antibodies, are added to the wells. These secondary antibodies are conjugated to the substrate-specific enzyme. A substrate, which is converted by the enzyme to elicit a chromogenic or fluorogenic signal is applied. The enzyme acts as an amplifier; even though only few enzyme-linked antibodies remain bound, enzyme molecules will produce many signal molecules. A problem of the indirect ELISA is that the method of antigen immobilization is non-specific. Any protein in the sample will stick to the well, so small concentrations of analyte in serum must compete with other serum proteins when binding to the well surface. The technique called "Sandwich" ELISA is used to improve the detection of antigen. The antigen is stick to the plate with the specific capture antibody bound on the surface. Then, the primary antibodies, which bind specifically to the antigen, are applied. Subsequently, the enzyme-linked secondary antibody, which is specific to the primary antibodies, is used. Afterwards, the binding is detected, with chemical converted by the enzyme into a color or fluorescent or electrochemical signal (Fig. 2). The presence of antibodies in the plate wells can be determined measuring the fluorescence or electrochemical signal.

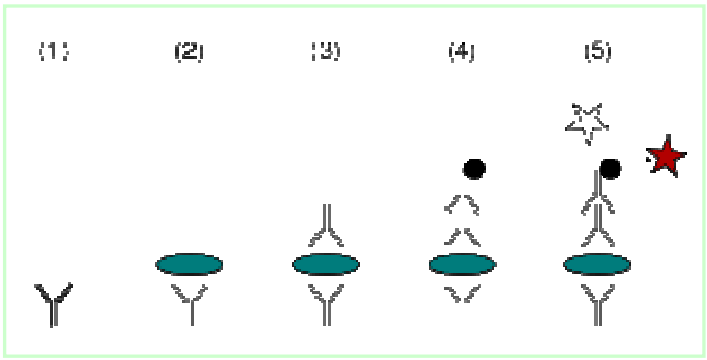


Fig. 2. A sandwich ELISA. (**1**) Plate is coated with a capture antibody; (**2**) a sample is added, and any antigen present binds to capture antibody; (**3**) a detecting antibody is added, and binds to antigen; (**4**) an enzyme-linked secondary antibody is added, and binds to detecting antibody; (**5**) a substrate is added, and is converted by the enzyme to a detectable form.

The major problems of the immunochemical techniques are the dependence to the structure and integrity of the antigen (Boccagni *et al.*, 1994). For example, in an analysis for the presence of an allergen in a food, it is not possible the use of fermented and hydrolyzed

proteins. Allergens which are used as a substrate for fermentation of microorganisms can't be analyzed by using the standard ELISA methods. This is because the microorganisms partially break down the allergenic proteins in the fermentation process, and this makes the proteins undetectable in the assays. However, there can still be active allergenic protein residues present that may pose a hazard to allergic individuals. Also protein modification during cooking or freezing the food can compromise the immunological analysis, inhibiting the antibodies binding to the food allergen. The ELISA technique is very sensitive, 10 ppm, and now is produced in several kits, easy to use (Gómez-Ollés *et al.*, 2006).

Analysis by mass spectrometry. The major aim of the research conducted in the ambit of this PhD Thesis was to identify allergenic proteins in a complex matrix composed of several hundreds of proteins. To solve the problem of identification of a specific protein is necessary the use modern proteomic tools, like mass spectrometry on line with a column for reverse phase chromatography (LC-MS/MS).

Proteomics research is based on the identification and characterization of all the expressed proteins in a cell, tissue or organism, including the protein isoforms, polymorphism and post-translational modifications. Usually, in a complex matrix all the proteins are not expressed at the same level, with the most abundant proteins signal that covers the low level ones. A single protein can be synthesized by a cell in different forms and with different modifications and, therefore, thousands genes can produce up to millions of proteins (Lane, 2005). The complexity of the proteomics research requires new technologies and new analytical protocols for sample preparation, protein detection and subsequent data analysis.

Mass spectrometry (MS) is the most important technique in proteomics in order to identify and characterize proteins (Hanash *et al.*, 2003; Phizicky *et al.*, 2003, Sali *et al.*, 2003; Tyers & Mann, 2003; Zhu *et al.*, 2003). It is important to say that mass spectrometry analysis is helped from other important discoveries and outputs in the genomic and software research. Actually, now it is possible to advice hundreds of genomes from lots of different organisms, and collect data from databases of information from all over the world. The technology for high-resolution two-dimensional electrophoresis has been considerably improved, which makes the method more reliable and reproducible. New methodologies for MS analysis are being developed, like novel ionization techniques and detectors, which allow the analysis of proteins and peptides at a higher sensitivity. Finally, new bioinformatics tools for combining and bundling the huge amount of data produced by new high throughput analysis methods have been developed.

The easiest and fastest way to identify proteins is schematically shown in Fig. 3 (Papping *et al.*, 1993). In the first step a complex protein sample is fractionated by mono-dimensional or bi-dimensional SDS-PAGE in order to purify and separate different proteins. The electrophoresis technique can be useful to separate proteins in different and sometimes extreme conditions. Proteins can be maintained in their native form or be denatured by different agents, like reducing reagents and detergents. Different protocols are applied in an acid or in a basic sample, to dissolve and separate all the proteins. Some information can be detected from a mono-dimensional SDS-PAGE experiment: for example, from the molecular weight of the analyzed protein it can be detected if the protein is aggregated or modified. Theoretically every protein expressed in a cell should be detectable, but with the actual state of electrophoretic technologies it seems impossible to separate up to million(s) of different protein entities.

After the SDS-PAGE separation, the gel spot containing the protein of interest is digested in situ with a proteolytic enzyme, mostly trypsin. The digestion of a protein into smaller peptides is used to enhance the solubility and malleability of the sample and to reduce the error in the mass analysis. Trypsin is very active and stable and its reaction consists on hydrolyzing the polypeptide chain selectively at the C-terminus of the amino acids Lys and Arg. This technique is called fingerprinting, because the peptides produced are really specific for the belonging protein. The peptides produced by this reaction are 15-20 residues long. The tryptic fragments are eluted from the gel and submitted to MS analysis. Since the cleavage sites of trypsin are known, theoretical tryptic peptide masses can be generated and compared with the experimental ones.

Although the molecular weight of a peptide can be correlated with the parent protein, there is the need of a more accurate analysis in order to avoid the possibility that post-translational modifications, like oxidation or deamidation, can alter the MS data. More specific information on a protein identity can be achieved by using peptide sequence analysis. This technique consists of the selection, during the MS analysis, of a peptide from the spectrum and fragment it inside the instrument. This process is called tandem mass spectrometry (MS/MS). The resulting fragment ion masses are indicative of the amino acid sequence and can be used to generate a sequence ladder. The data derived from this analysis can be used to search the sequence and the identity of the peptide in protein database.

In a proteomics experiment the mass spectrometer is usually interfaced with a liquid chromatography system, where the protein samples, or the tryptic fragments, are loaded in a RP-HPLC column. The peptides are separated with a gradient of organic solvent, usually

acetonitrile, following their hydrophobic scale. This system is miniaturized and specific to load minute amounts of sample. The analysis described herewith are made with this type of nanotechnology, with a column of diameter 75 μm and length 150 mm, eluted at a rate of 200 nanoliters per min. The peptide eluted from the column is injected in the electrospray (ESI) ion source through a very small capillary (Fig. 4). In electrospray ionization, a liquid is pushed though the capillary, with the analyte dissolved in a large amount of solvent, which is usually much more volatile than the analyte (Fenn *et al.*, 1989). The analyte exists as an ion in the solution either in its anion or cation form. Because of charges repel, the liquid pushes itself out of the capillary and forms an aerosol, a mixture of small droplets (about 10 μm diameter). An uncharged carrier gas such as nitrogen is sometimes used to help nebulize the liquid and evaporate the neutral solvent in droplets. As the solvent evaporates, the analyte molecules are forced closer together, repel each other and break up the droplets. This process is driven by repulsive Coulombic forces between charged molecules. When the analyte is an ion free of solvent, it moves to the mass analyzer.

In the following step the ion produced is analyzed, in the instrument, by a Q-Tof Micro (Micromass, Manchester), that connects an ESI source with two combined analyzers quadrupole Tof (*time of flight*). Mass analyzers separate ions according to their mass-to-charge ratio, following the dynamic properties of charged particles in electric and magnetic fields in vacuum. The quadrupole mass analyzer uses oscillating electrical fields to selectively stabilize or destabilize ions passing through a radio frequency (RF) quadrupole field, acting as a mass selective filter. The time-of-flight (Tof) analyzer uses an electric field to accelerate the ions through the same potential, and then measures the time they take to reach the detector. If the particles all have the same charge, then their kinetic energies will be identical, and their speed will depend only on their masses. Lighter ions will reach the detector first.

The data produced are represented in a mass chromatogram of total ion current (TIC), measured in the ion source during the RP-HPLC separation (Fig. 5A). The instrument acquires a mass spectrum of the peptides eluted from the column (Fig. 5B), based on an intensity vs. *m/z* (mass-to-charge ratio) plot. Afterwards, the instrument can determine the sequence of the peptides though a Tandem mass spectrometry analysis (Fig. 5C).

The Tandem mass spectrometry MS/MS experiment involves multiple steps of mass selection or analysis, usually separated by some form of fragmentation. The first mass analyzer isolates one peptide from many entering a mass spectrometer. Then, a second one stabilizes peptide ions while they collide with a gas (argon, helium or nitrogen), causing them to fragment by collision-induced dissociation (CID). Finally, a third mass analyzer registers

the fragments produced from the most abundant peptides. The ion of the initial peptide is called “precursor ion”, while the ions produced in the MS/MS spectrum are called product ions. Fig. 6 represents the typical fragmentation of a peptide, and the nomenclature of the product ions in a MS/MS experiment. A peptide sequence tag obtained by tandem mass spectrometry can be used to identify a peptide in a protein database. Peptide fragment ions are indicated by a, b, or c if the charge is retained on the N-terminus and by x, y or z if the charge is maintained on the C-terminus (Biemann, 1992) (Roepstorff & Fohlman 1984). The subscript indicates the number of amino acid residues in the fragment. Superscripts are sometimes used to indicate neutral losses in addition to the backbone fragmentation, for example for loss of ammonia and for loss of water. Although peptide backbone cleavage is the most useful method for sequencing and peptide identification, other fragment ions may be observed under certain conditions. These include the side chain loss ions and immonium ions.

The amino acid sequence is determined in the MS/MS spectra by the difference from a fragment with the immediately subsequent one, which is the molecular weight of a specific amino acid (Fig. 5C). There are two main drawbacks in the data analysis: the main problem is that some ions can be missed, or spectra can be complex for the presence of other ions not belonging to the specific series. Finally, in these cases, the peptide sequence can only be partially determined.

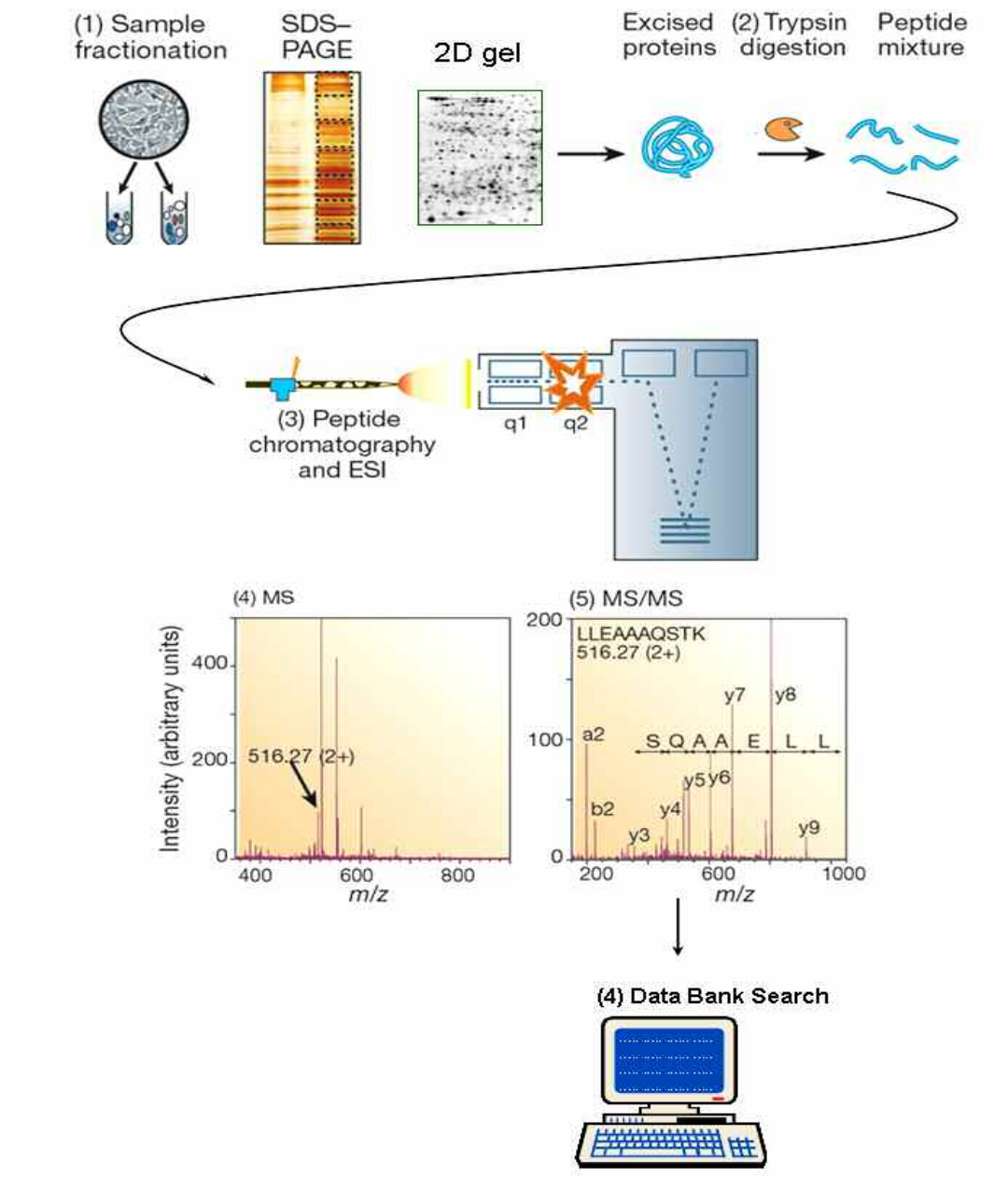


Fig. 3. Scheme of the usual process in a proteomic experiment.

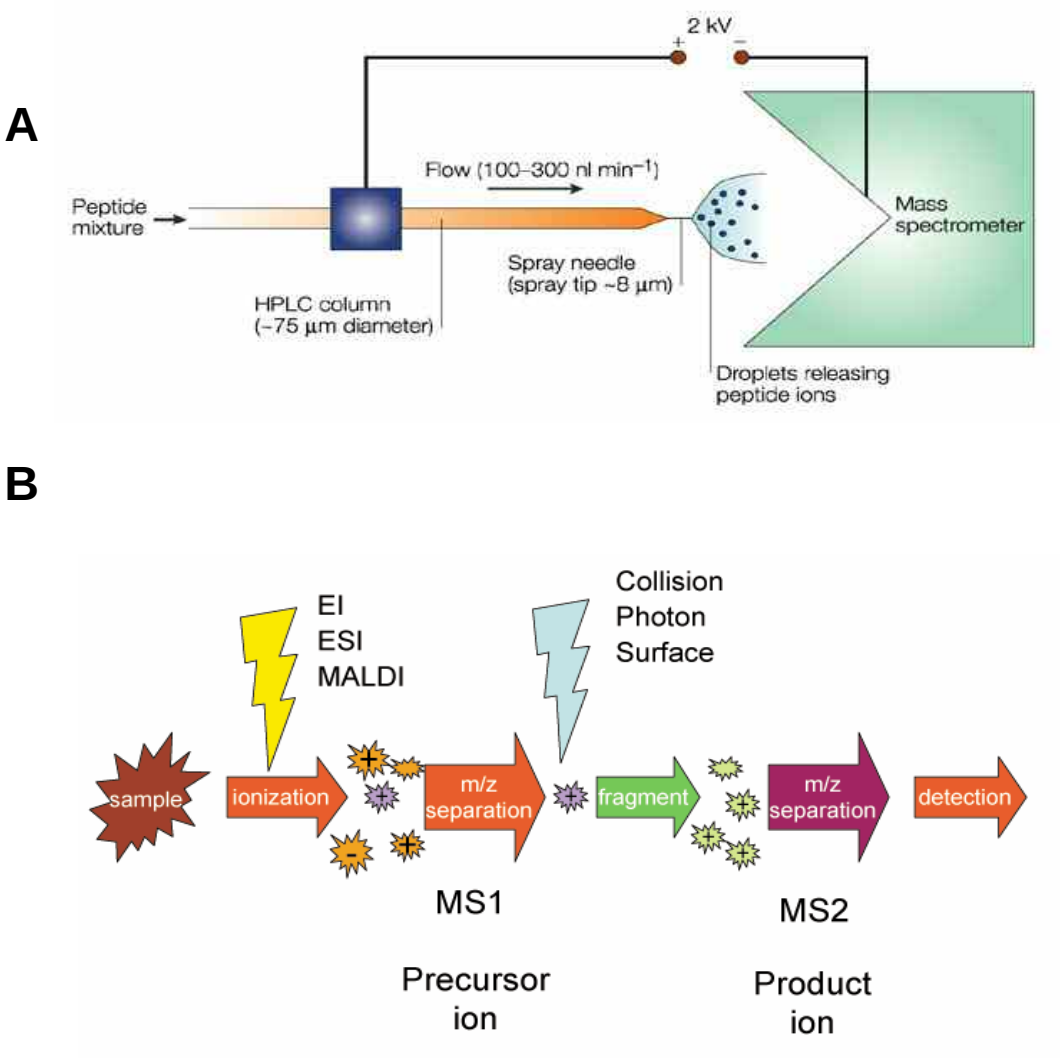


Fig. 4. Electrospray ion source interfaced with an RP-HPLC system for the LC-MS/MS analysis (**A**). Scheme of the Tandem mass spectrometry experiment (**B**).

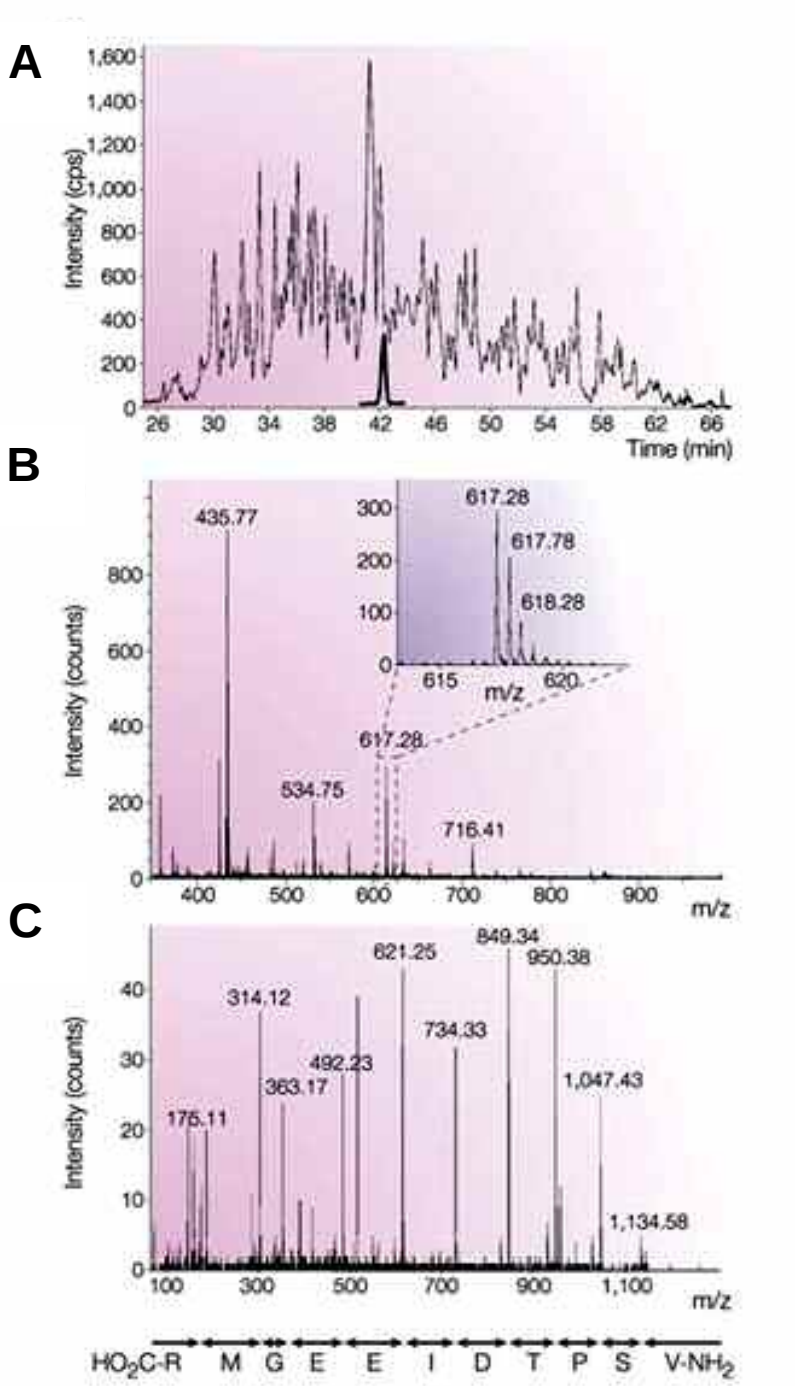


Fig. 5. Example of results spectra from a LC-MS/MS analysis. Total ion current (TIC), in function of the time, measured in the mass spectra during the experiment (**A**). Mass spectra of the peptides eluted in the same experiment, the ion of interest is highlighted (**B**). MS/MS spectra of the peptide of interest, the mass difference of the y ions series originated from the parent ion allows the reconstruction of the amino acid sequence (**C**).

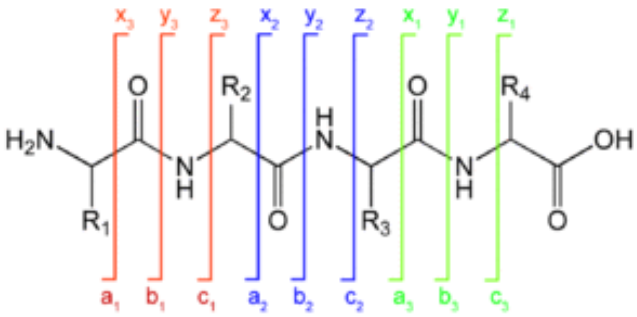


Fig. 6. Peptide fragmentation notation in a Tandem Mass spectrometry experiment.

The following step of a proteomic experiment is to use the sequence data derived from product ion MS/MS spectra from one or more peptides. The fragmentation of peptides within the mass spectrometry can yield stretches of sequence information. A number of search engines can correlate this information to protein sequence database. The most important sequence engines are: SEQUEST, MASCOT, SONAR. Sequence information is significantly more discriminating than molecular mass alone and as such, one peptide of quite long length is sufficient to identify a protein. A further tool for protein identification is the “expressed sequence tag” database (ESTs). ESTs are short nucleotide sequences, which peptide sequence data can be used to search against. If the database search is unsuccessful, it is possible also perform a manual peptide *de novo* sequenced. This is the case of an organism whose genome is not completely sequenced, or the characterization of genes splice variants.

During this research project, the sequence engine used is MASCOT (Perkins *et al.*, 1999) (Fig. 7). The algorithm is based on the comparison between the MS/MS experimental spectra results and all the spectra theoretically created in the database. The comparison starts from the ions signal with higher intensity, usually of the b or y series, until the complete sequence. Proteins are then identified on the basis of a probabilistic score, calculated by taking the number of identified fragments as casualty. The software, therefore, can transform the MS and MS/MS data in sequences of amino acids and then identify the analyzed protein.

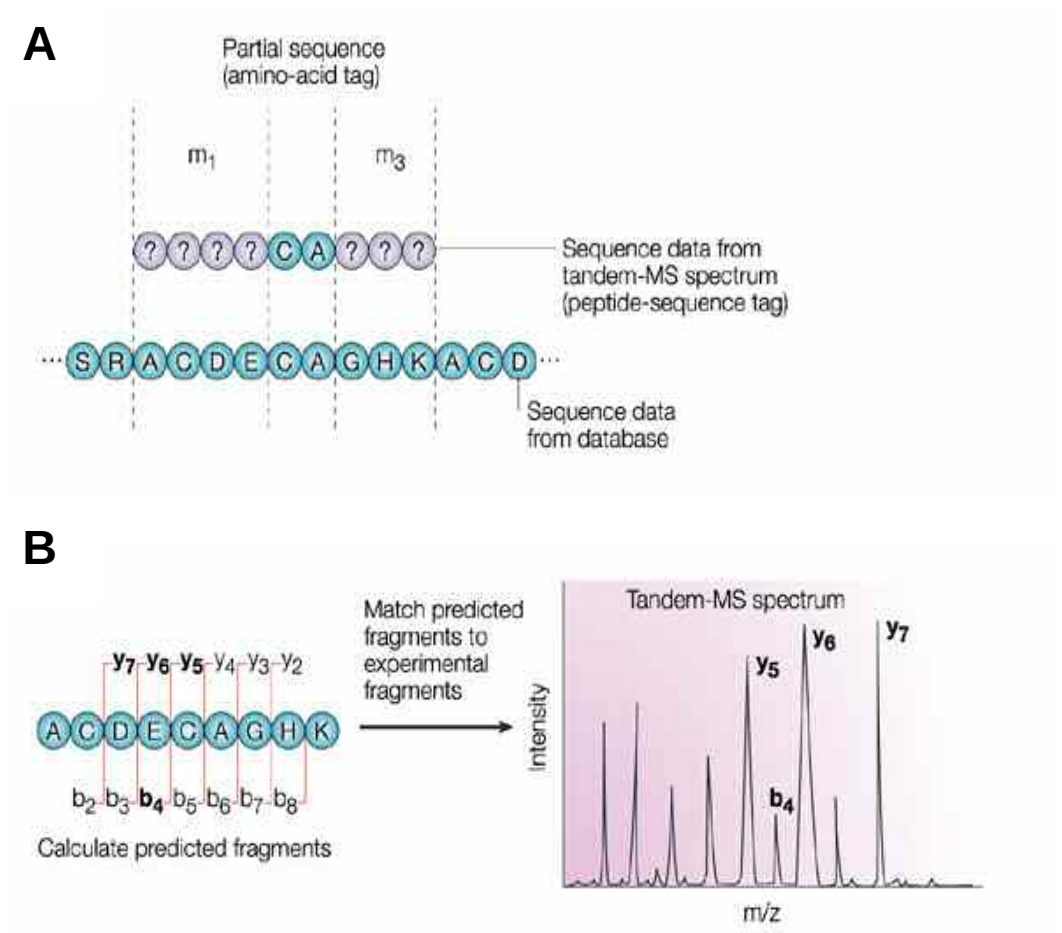


Fig. 7. Different ways for protein identification in a database. *Peptide Sequence Tag* (**A**). Protein identification by MASCOT database (**B**).

4. Milk and egg-white protein allergens

Purification of food allergens. Food allergies are a widespread disease and indeed several statistics reveal that almost 10% of the Europeans suffer from food allergy or intolerance. In order to avoid the risk of an allergic reaction, it is important for a sensitive subject to know all the ingredients in a food, but sometimes this is impossible. The actual method to detect the presence of a specific allergenic protein in a food is ELISA kit test, which reaches a sensitivity of 1 ppm in some kits. To improve this technique, we have collaborated with other research groups for the immunochemical analysis. The first aim of our research was the isolation of the allergen proteins from food with the highest purity;

afterwards, the other member of the team could develop a sensitive ELISA method to detect and quantify the allergens analyzed.

Milk and eggs are convenience goods, but also foods that trigger important allergic reactions (Urisu *et al.*, 1997). The most important allergens are proteins: β-Lactoglobulin (β-LG), α and β caseins (α-CN and β-CN) in milk, ovomucoid (OM), ovalbumin (OA) and lysozyme (Lys) in egg's white. These ingredients are very common, in bread, cakes, pasta and many others, even though in little amount. For this reason, it is very important to detect and study their protein content, also in complex foods where they are present only in traces.

The protein purification from milk and egg white was performed with an analytical method, studied to have the best purification of both the overall extract and every protein. Other foods, like peanuts and soybean, where analyzed in their total protein content. Gel chromatography and reverse phase chromatography are techniques of common use for protein extraction and analysis, purifying the sample of interest from the raw food. Proteins are separated by a physical characteristic, like dimension, or shape, or chemical characteristics, like idrophobicity. Gel electrophoresis SDS-PAGE is a useful technique for protein analysis, especially to have an overall picture of the total protein content in a food. These technique were used in this project, to analyze the proteins and to provide the sample for further and more accurate analysis like mass spectrometry.

The chemical and structural analysis of a protein is important to characterize its function and behavior in different physical conditions, where external agents can influence its structure (Czerwenka *et al.*, 2006). Protein allergens have very peculiar characteristics, about the secondary and tertiary structures, the post translational modifications, and the resistance to high temperature or to proteolysis (Luz Sanz *et al.*, 2007). Milk and egg allergens are the most common in our food, and, in this project, we studied the structure of some of them.

The first characterization of a protein of interest is the analysis of its primary structure, the amino acidic sequence, by using mass spectrometry. First of all, proteins analyzed were standard proteins available from a company, α-LA and β-LG from milk, and Lys, OM and OA from egg white (Fig. 9). This analysis provides a standard for the following experiments with proteins extracted from food. α-LA and β-LG were also extracted from pasteurized milk, and analyzed by fingerprint in solution and ESI mass spectrometry (Fogliano *et al.*, 1998). This experiment permits a direct comparison with the commercial standard, processed in the same way. The analysis gives information about the integrity and the modifications of proteins which has already been subject to a thermal process in milk (Siciliano et al. 2000).

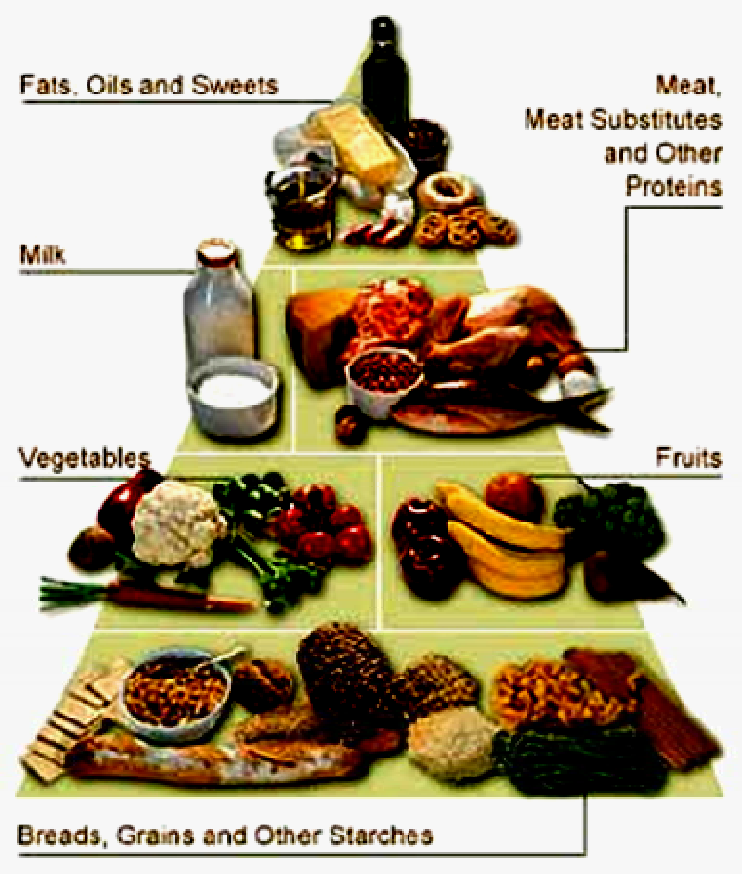


Fig. 8. Common food containing allergenic proteins in Europe

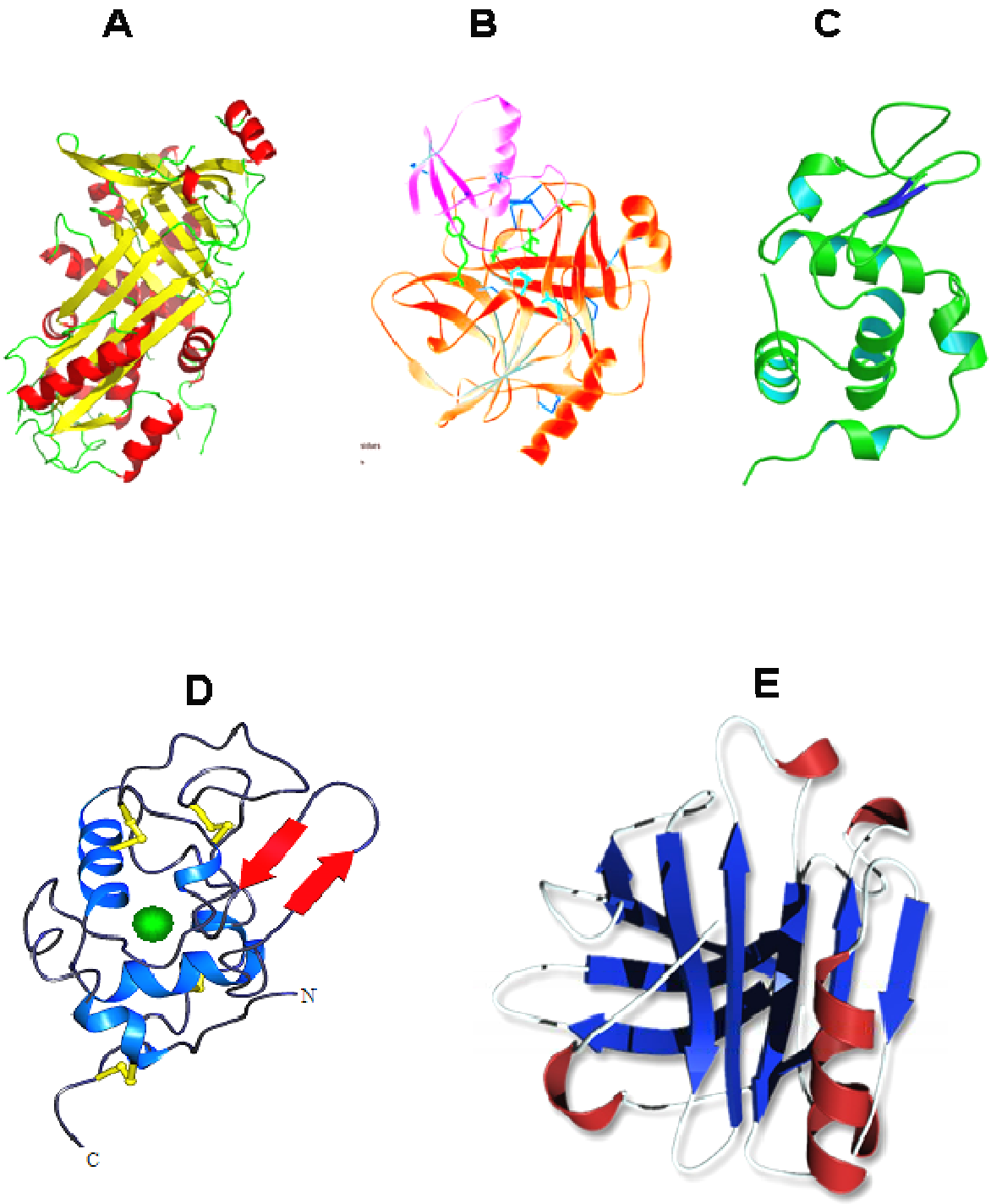


Fig. 9. Three-dimensional (3D) structure of the allergenic proteins herewith analyzed: Ovalbumin (PDB code 1p1z) (**A**), Ovomucoid (PDB code 1ple) (**B**), Lysozyme (PDB code 1ghl) (**C**), α-Lactalbumin (PDB code 1hfz) (**D**), β-Lactoglobulin (PDB code 1beb) (**E)**.

Analysis by mass spectrometry. The main purpose of this project, after the isolation and characterization of the most common allergens, is to apply the mass spectrometry technology to the allergy problematic. The detection and identification of allergens in food, nowadays is conducted by ELISA test, which is very specific and powerful, but with some problems in its application. The immunological analysis is strictly dependent on the allergen and on its integrity and quality. The mass spectrometry analysis, by LC-MS/MS, of proteins separated by SDS-PAGE electrophoresis, can provide an alternative technique to use in parallel (Natale *et al*., 2004). As a matter of fact, sometimes the immunological ELISA test can fail, especially when the allergen protein is modified. A protein is subject to enzymatic modifications inside the cell, like phosphorylation, and to non-enzymatic modifications as well. In a food, transformed by an industrial process, proteins can be oxidized and aggregated, in a way not always predictable. These chemical alterations can hide an epitope, the specific sequence where the antibody binds. It can be very difficult to make an antibody specific for every protein modification, and the mass spectrometry technique is a method that can recognize a protein modified, which is the modification and where it is situated in the amino acid sequence.

The allergens in protein mixtures extracted from foods were identified by LC-MS/MS (Fig. 10). Milk was the source of α-LA and β-LG and egg-white was used to analyze ovalbumin (OA), lysozyme (Lys) and ovomucoid (OM). Subsequently, these proteins were analyzed in the protein extract of a complex food (cake) made of a mixture of different ingredients and subject to cooking (Hasenkopf *et al.*, 2002). This experiment wants to elucidate the physico-chemical conditions of the allergens analyzed. Many alterations can happen during mixing or cooking, like hydrolysis, aggregation or chemical modifications. The high cooking temperature is a good test, both of protein resistance to thermal stress and the sensitivity of the new methodology. It is also important to improve the technique of extraction and separation of proteins from a food, in order to be able to analyse all proteins before and after the stress modifications. It is known that one protein should be harmless or cause an allergic reaction only for a single mutation, which can trigger an immunogenic response. For this reason, it is very important not only to detect the allergen, but also to identify the protein epitope that has caused the allergic reaction.

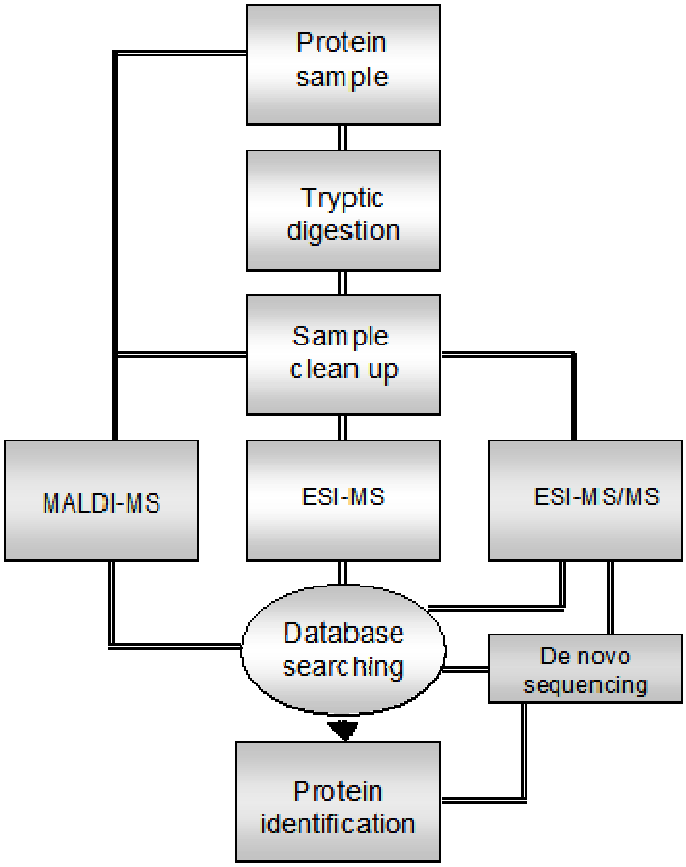


Fig. 10. Scheme of the method developed to identify the proteins in a complex mix, by a mass spectrometry analysis MALDI and ESI.

5. Case studies of mass spectrometry analysis of food allergens

Ara h1 from peanut and Gly m db 28K from soybean. Allergens not always are described in food labels, and it is important to analyze every food suspected of contamination, in order to protect people who suffer from food allergies. The mass spectrometry technique can be very useful in this analysis, to identify and analyze the protein content in a commercial food. During the last years this technique improved a lot, and now it allows faster and more sensitive analysis. The automation of the methodology LC-MS/MS can be applied with good results, either to find allergenic proteins in a mix food, or to identify a protein present in a small amount, but really allergenic. Peanut allergy is one of the most diffuse, trigging an anaphylactic shock in a sensitive patient. The percentage of children sensitive to peanuts is the 4-6%, in adults is the 3-4%. The study of peanut proteins is very important, to have an early ddiagnosis of the allergy and treat the patients. Peanuts contain many allergenic proteins, the most are storage proteins of different molecular weight. The most important, and studied, for their allergenicity are Ara h1, of 63.3 kDa, and Ara h2, of 17 kDa. (Koppelmann

et al., 1999). Ara h1, the protein studied in this project, has 626 amino acids and has no enzymatic activity (Fig. 11) (Viquez et al. 2003). This molecule is very resistant to high temperature, to denaturants and to degradation by stomach proteases and particularly to pepsin. Ara h1 has also an inhibiting action to trypsin, a useful characteristic to achieve more resistance to degradation.

Ara h1 is undoubtedly the principal protein allergen in peanuts: 90% of the patients allergic to peanuts, are reactive to this protein. Immunological studies show the presence of 23 independent epitopes, located all over the tertiary structure of the protein (Burks *et al.*, 1994, Shin *et al.*, 1998). Ara h1 is usually glycosylated and forms trimers under physiological conditions, structures that help the resistance to proteases and cover allergenic sites, difficult to find in an immunologic analysis.

The soybean is a very common legume with high protein content. Nevertheless, it is a strong allergenic food and in the last years the number of allergic people has raised from 1 to 6% (Cordle, 2004). It has been calculated that only 0.3 g of soybean flour is sufficient to trigger an allergic reaction in sensitive subjects (Sicherer *et al.*, 2007). About 70% of soybean proteins has storage function, are rich of nitrogen, sulphur and carbon and are often glycosylated. The most important allergens of the soybean are Gly m bd 30K, Gly m Bd 68K and Gly m Bd 28K (Roychaudhuri *et al.*, 2004, Kasai & Ikehara, 2005). Gly m Bd 28K is a glycosylated protein of the vicilin family, whose molecular weight is 26 kDa. It is present only in a low quantity in soybean and is totally absent in some varieties (Hiemori *et al.*, 2000, Bando et al., 1998). Gly m Bd 28K, even though in low quantity, is a strong allergen and is the cause of the 25% of allergic reactions to soybean. It is synthesized in the cell as a prepro-protein of 52.6 kDa, is composed of 473 amino acid residues, with a signal peptide and a C-terminal sequence called 23 kDa peptide (Xiang *et al.*, 2004). The tertiary structure of Gly m Bd 28K was solved by X-ray methods and shown to consist of a β-barrel structure in the centre of the molecule (Breiteneder & Radauer, 2004). Gly m Bd 28K has been modified in a mutant cell line, Tohou 124, deleting the allergic sequence fragment, and some ipoallergenic food is now made with this recombinant soybean, allowing the 80% of the soybean allergic patients to eat this food. Soybean proteins sometimes cross react with peanut allergens, like Gly m Bd 30K and Ara h1, which have the 70% of sequence homology.

The isolation of Ara h1 and Gly m Bd 28K, from the protein extract of the seeds, and the subsequent analysis, is a good application of mass spectrometry to the allergy problematic. In the case of Ara h1, this experiment provides a standard of the protein analysis, to find, in a subsequent experiment, this allergen also in complex samples or mix food. The

purification and analysis of Gly m Bd 28K give another example of the success of this technique, in this case also applied in searching for a low concentrate protein. Mass spectrometry analysis is a very sensitive tool, and permits to find proteins present less than a nanomolar concentration in a sample.

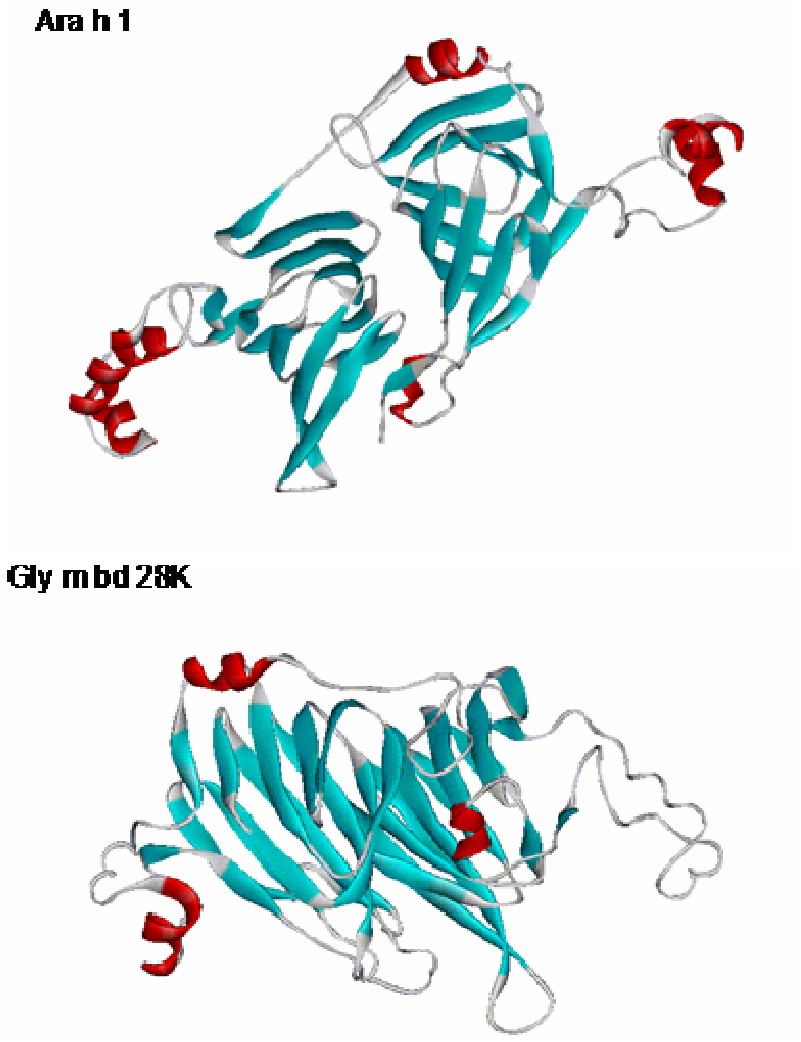


Fig. 11. Structure of the allergen proteins Ara h1 (PDB code 1od5) and Gly m Bd 28K (PDB code 2cav).

6. Analysis of allergenic proteins after heat treatment of food

***Maillard reactions in proteins*.** A complex food can cause an allergic reaction in a sensitive subject, which can be enhanced or avoided if the food is subject to a thermal process (Ehn et al., 2005). The allergenic potential of a protein is based on its capacity to bind IgE antibodies that, in some cases like for β-LG, is enhanced by food processing (Breiteneder & Mills, 2005). It is actually known that food antigenic proteins can change their properties with process dependent covalent modifications (Nielsen et al., 1985).

Food and food proteins are usually subject to thermal processing, like cooking and freezing; biochemical procedures like fermentation; isolation and purification processing, like melting, pressing, filtration; chemical preservation with alcohol, salt and pickling and other procedures, like mixing, colouring, emulsifying, homogenization (Mills & Breiteneder, 2005). These processes can alter the properties of foods and modify the proteins and lipids, creating new allergens or destroying them. The most studied food is milk, in particular infant milk formulae, for the importance that it has all over the world in the feeding of babies and its great consumption because of its presence in many foods. The reactions developed between milk proteins and other milk molecules after thermal treatment are well studied. For example, it is know that boiling decrease the IgE binding of whey proteins and fermentation decrease the 99% of the antigenicity of α-LA and β-LG. Egg is a convenience good studied for its allergenic proteins, OM and OA that decrease their allergenicity of 58% when boiled.

Basically, thermal processing should decrease food allergenicity because heating and cooking cause the disruption of protein structure, or should create neoantigens, changing the digestibility, solubility and resistance of proteins to stomach acid (Humeny *et al.*, 2002). Modified protein is more resistant to acid or enzymatic digestion and, if stable, can be absorbed intact. The chemical modification can be recognized by the immune system, forming a new B-cell epitope or new anti IgE epitope. Glycation, the reaction between a sugar and a free amino group of the protein, is the first and simplest reaction between a protein and a sugar. The study of food proteins is important to modulate the processing conditions of their manufacturing, in order to minimize the loss of nutritional value (Seiquer *et al.*, 2006).

Food proteins and lipids react with reducing sugar or sugar breakdown products at high temperature (Fig. 12) (Gerrard *et al.*, 2005). These reactions, called Maillard reactions, cause the glycation of a protein N-termini or a ε-amino group in the amino acid K. The Shiff base formed is in equilibrium with glucose in solution, until the irreversible rearrangement in other products, called Amadori products (Fay & Brevard, 2005). The advanced glycation end

products (AGEs) are responsible of the yellow or brown colour of foods and are often the crosslink agent of food proteins. There are several consequences of the cascade of reactions started by glycation: oxidized lipids and reactive oxygen intermediates can be formed, different proteins can be modified in their chemical and conformational structure (Ahmed & Thornalley, 2002). Food proteins can change completely for a disulphide bond scrambling or the deamidation and subsequent loss of asparagines. The series of rearrangements and dehydration cause the formation of a large variety of low molecular weight species, like furfural, furanose, hydroxyketones and several dicarbonyl compounds, which react with other reactive molecules, amines, amino acids and aldehydes (Ahmed, 2005). The products of these reactions are odour active molecules, responsible for the food aroma, while in the last stage of Maillard reactions there is the formation of a molecule at high molecular weight called Malanoidin polymer (Odani *et al.*, 1998). These big molecules are formed by condensation of cyclic subunits and are often individuated by immunohistochemistry (Miller & Gerrard, 2005).

Several techniques are employed to study Amadori products and other molecules formed during the Maillard reactions. The most important investigation techniques are mass spectrometry and NMR, while also capillary electrophoresis and specific chromatography are necessary for the isolation and quantification of Amadori compounds (Marvin *et al.*, 2002). Studies have been done on milk proteins, especially in serum proteins like β-LG, to study milk proteins changes in structure and their physical and chemical property after heat treatment (Sun Hong *et al.*, 2006). The main Maillard reaction leads to the formation of lactosylated protein species, enhancing its emulsifying and foaming properties. For example, the immunoreactivity of β-LG increases with thermal denaturation, because of the changes in the secondary and tertiary structure which alter its antigenic properties (Docena *et al.*, 2002). Caseins, which represent the 80% of milk proteins, are another important class of milk compounds. These proteins are often phosphorylated and glycosylated, and can cause an allergic reaction in a sensitive subject.

Maillard reactions cause a nutritional loss and other important consequences, like the missed intestinal uptake of an Amadori product, absorbed only by diffusion (Birlouez-Aragon *et al.*, 2001). The digestibility pathway of food proteins at this point looses control, with an increase to pepsin and pancreatin resistance from the crosslinked proteins. The reaction between albumin and globulin from wheat protein with glutaraldeyde, the milk covalent polymerization during food processing, are recurrent phenomena subject to an in depth study (Kato *et al.*, 2000). The AGEs molecules can not be recognized by the immune system, so

normal food proteins became epitope for an allergic reaction (Rytkönen *et al.*, 2006). An epitope formed by primary or secondary structure of a protein can also be deleted by the formation of an Amadori product, because of the modification of an amino acid or the disruption of a secondary structure, with the consequence that the protein is no more allergenic (Davis *et al*., 2001).

The technique HPLC-MS is used to quantify several compounds like lactuloselysine and furanine. Mass spectrometry, both ESI and MALDI, is very useful to detect Maillard reactions products and marker molecules of the degraded proteins in a complex sample. The fluorescent properties of the pentosidine at a specific wavelength can be used for the characterization of the same Amadori products, derived from the modification of K and R.

Glioxal (GA), Methyl glyoxal (MG) and Deoxiglucosone are some of the dicarbonyl compounds formed in a non-enzymatic way, by amine-catalyzed sugar fragmentation reactions, or, in the human body, from a spontaneous decomposition of triose phosphate intermediates in glycolysis (Fig. 13A) (Uchida *et al*., 1997). The most reactive is MG, which irreversibly modifies proteins under certain physiological conditions and concentration (Ikeda *et al*., 1998). The products of the conjugation of an amino group of a protein with these active molecules are fluorescent. The modified protein can be toxic for the cell, and cause the inhibition of cellular respiration (Ankrah & Appiah-Opong, 1999).

The reaction between MG and the amino acid R gives different isomeric products: N^{δ}-(5-methyl-4-imidazonolon-2-yl)-L-ornitine (5-methylimidazolone), N^{δ}-(5-hydroxy-5-methyl-4-imidazonolon-2-yl)-L-ornitine (5-hydro-5-methylimidazolone) and N^{δ}-(5-hydroxy-4,6-dimethylpyrimidine-2-yl)-L-ornitine (argpyrimidine), the most fluorescent product (Fig. 13B) (Oya *et al.*, 1999). The reaction between MG and the amino acid K gives MG-derived K-K crosslink (imidazolise), and N^{ε}-carboxy-ethyllysine (CEL) (Thornalley, 2005). Glyoxal is less active than MG, but often reacts with K, forming N^{ε}-ethyllysine. All these Amadori products have a specific molecular weight, and can be detected by mass spectrometry. Also ELISA antibody test and fluorescence spectroscopy can be employed to verify the presence of these molecules, which are allergenic and fluorescent. Fluorescence measurements are used to define the presence of a fluorescent modification, excitating at a wavelength of 330 nm, the maximum for the MG modified products, and monitoring the emission maximum between 400 nm and 440 nm. The increase in fluorescence is a proof of an increase of protein modification, even though this is not a quantitative measure (Lo *et al.*, 1994). Protein fluorescence is measured by its W and Y intrinsic fluorescence, with the excitation

wavelength of 280 nm and measuring the emission at 320, 350 nm. The MG modified proteins are subject to a strong shift in fluorescence, the intrinsic protein fluorescence decreases and there is an increase of the signal between 395-440 nm, the fluorescence wavelength typical of an AGEs modified protein. Also GA and glucose modified proteins can be subject to the same fluorescence decrease at 320 nm and increase at a wavelength specific for their fluorescent advanced modifications.

Other modified molecules have allergenic and mutagenic properties, as heterocyclic aromatic amines and the acrylammide produced by β-elimination of Amadori products. So, the study of their composition and structure is important to detect and quantify them. AGEs occur in vivo and in vitro: a sample is usually a complex mixture of different structures derived from different reactions of a great variety of molecules (Coussons *et al*., 1997). The differences between AGEs samples are determined by the buffer or the matrix in which the reagents react, the incubation time, the temperature of the process and the concentration of modifiers. All these variables make the comparison between different samples difficult (Baisier & Labuza, 1992). Maillard reactions cause the deterioration of functional properties of proteins, either in vitro, like in foods or medical preparation, and in vivo (Newkirk *et al.*, 2003). As a matter of facts, glycation reactions occur endogenously in all tissues and body fluids under physiological conditions. In the diabetes disease, modifications of proteins are not caused from sugar per se, present in a huge concentration, but from dicarbonyl compounds derived from Maillard reactions (Bucala *et al.*, 1995; Fu *et al.*, 1996). Methylglyoxal is originated in vivo by various mechanisms: by dephosphorylation of the glycolytic intermediated, polyol pathway metabolites and acetone. The best precursors in vivo of MG are glyceraldehyde, the sugar ribose and L-Threose. MG, present in vivo only in nanomolar amounts, reacts rapidly, and is more effective than glucose with a quicker kinetic. The human body is already used to this molecule and it has specific enzymes to modulate MG-mediated reactions. On the other hand Maillard reactions can be intentionally used to produce stable proteins. It is known that the stability of a protein at higher temperature can be increase by some chemical modifications, and an improvement of emulsifying activity after Maillard reaction with a sugar is also observed.

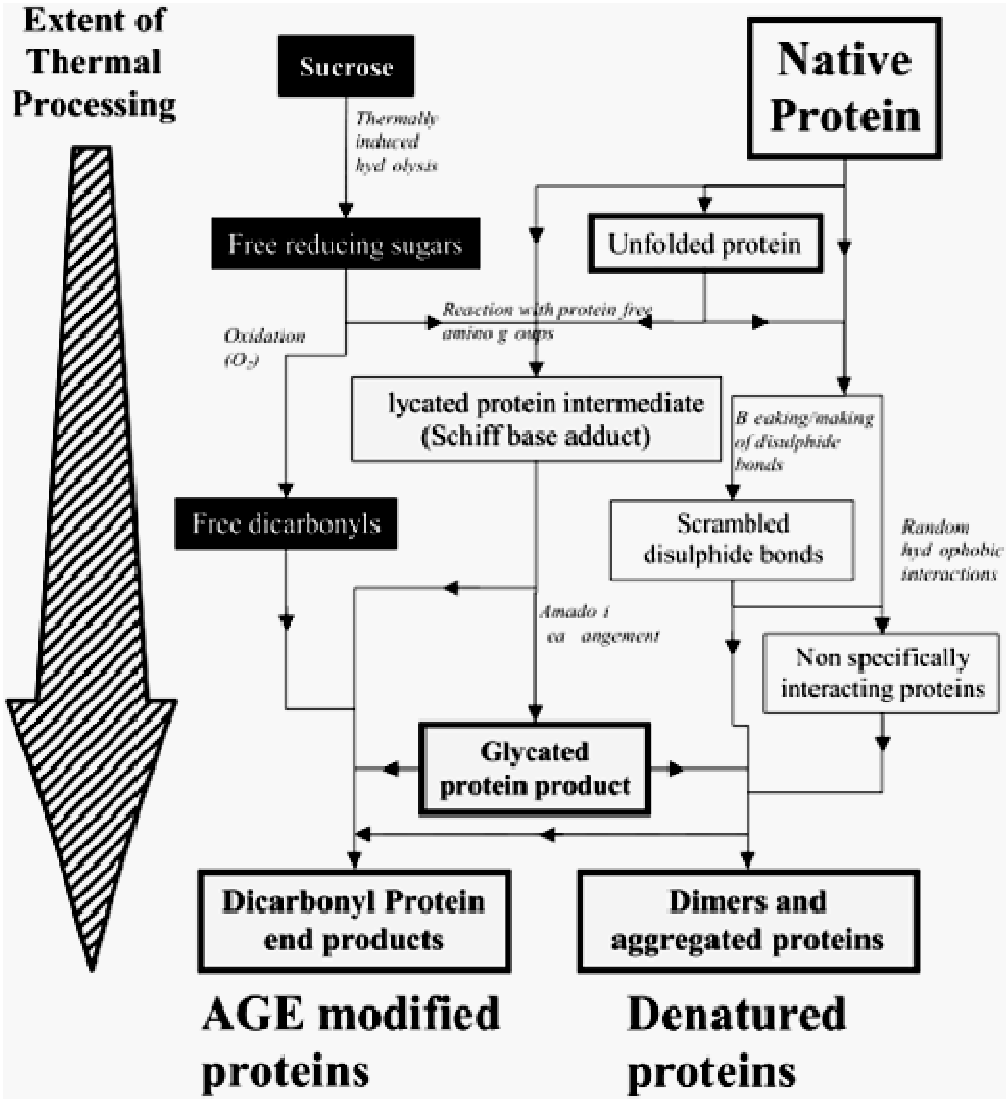


Fig. 12. Diagram of the chemical modifications in proteins, caused by heating in presence of a sugar. The black box shows sugar derived molecules, in the white box are shown the proteins. The covalent modified proteins are in grey boxes, the grey boxes delimited by a thick line are covalent modified proteins with an increased allergenicity.

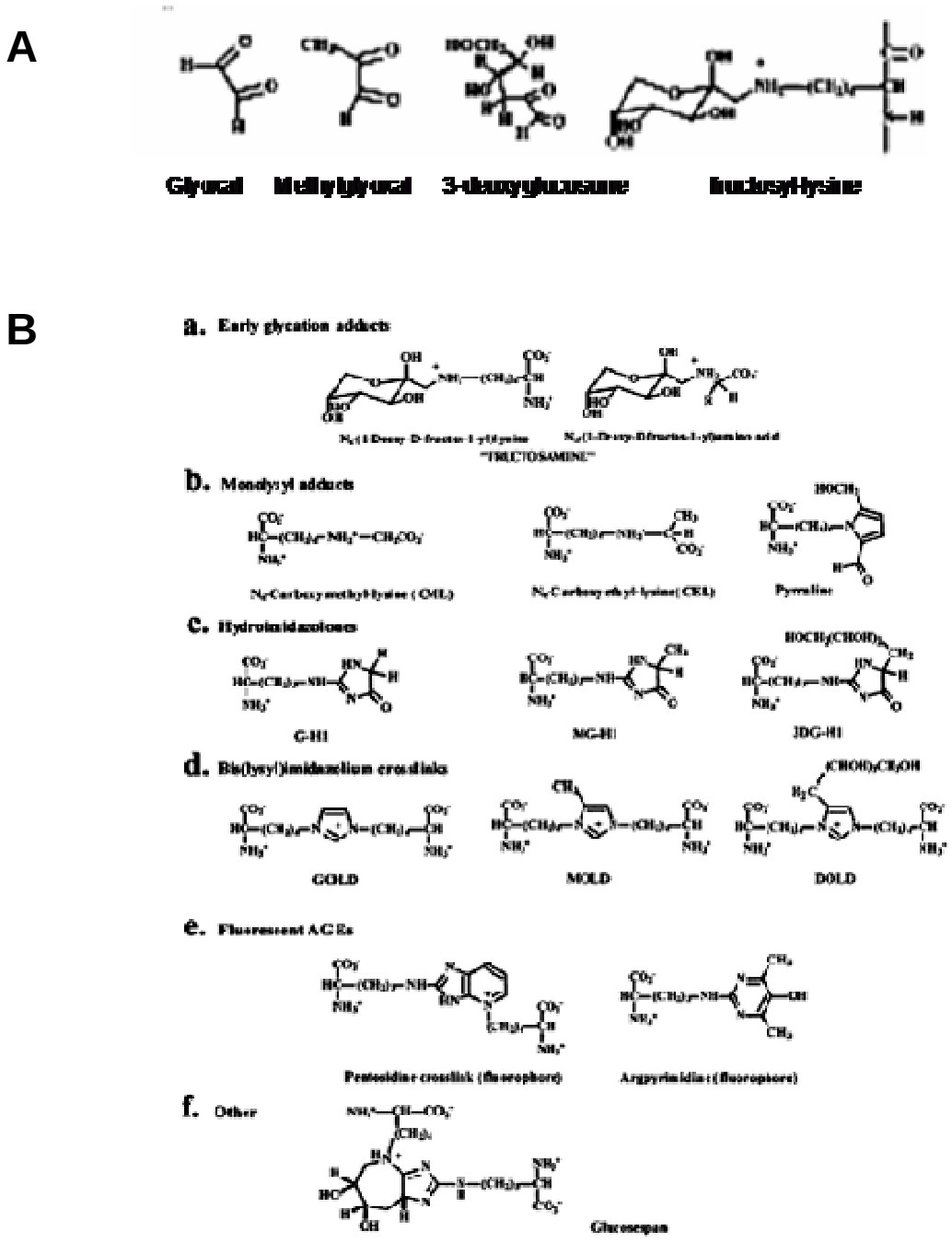


Fig. 13. Structure of the most reactive dicarbonyl compounds derived from sugar and lipid oxidation (**A**). Structure of the most frequent Amadori products derived from interaction of dicarbonyl compound and K or R (**B**).

Analysis of caseins in foods. Caseins are the most abundant proteins in milk and in all the milk food derivates. The principal casein fractions are α_{s1}-CN and α_{s2}-CN, β-CN, and κ-CN. The distinguishing property of all the caseins is their low solubility at pH 4.6. The common compositional factor is that caseins are conjugated proteins, mostly with phosphate groups esterified to serine residues. These phosphate groups are important for the structure of the casein micelle, to explain the function of calcium binding. The conformation of caseins is similar to denatured globular proteins: the high number of proline residues causes particular bending of the protein chain and inhibits the formation of close-packed, ordered secondary structures. Caseins contain no disulfide bonds and, as well, the lack of tertiary structure matters in the stability of caseins against heat denaturation. The low level of secondary and tertiary structure associated with the high exposition of hydrophobic residues results in a strong association reaction of the caseins and renders them insoluble in water. Caseins are present in different quantity in milk, the respective concentration is 4(α_{s1}-CN):1(α_{s2}-CN):4(β-CN):1(κ-CN) (Malin *et al.*, 2005).

α_{s1}-CN is a protein of molecular weight 23.000 Da, 199 residues, of which 17 amino acids are proline residues. The protein is formed by two hydrophobic regions, containing all the proline residues, separated by a polar region, which contains all but one of eight phosphate groups. It can be precipitated at very low levels of calcium. α_{s2}-CN is a protein of molecular weight 25.000 Da and 207 residues, with 10 prolines. This protein concentrates the negative charges near the N-terminal and the positive charges near C-terminus. It can also be precipitated at very low levels of calcium. β-CN has a molecular weight of 24.000 Da, contains 209 residues, and 35 amino acids are prolines. The N-terminal region is highly charged and the C-terminal region is hydrophobic, so β-CN is a very amphiphilic protein and it acts as a detergent molecule (Farrell *et al.*, 2002). The property of self association is temperature dependant, the consequence is that, basically, it will form a large polymer at 20°C but not at 4°C. This protein is less sensitive to calcium precipitation than the other caseins. κ-CN is a protein of molecular weight 19.000 Da, 169 residues, of which 20 amino acid are prolines. This protein is very resistant to calcium precipitation, with the function of stabilizing other caseins (Kumosinski *et al.*, 1991). The protease rennet cleavage at the F105-M106 bond of κ-CN eliminates the stabilizing ability, leaving a hydrophobic portion, called para-κ-casein, and a hydrophilic portion, called kappa-casein glycomacropeptide (GMP), or more accurately, caseinomacropeptide (CMP) (Minikiewicz *et al.*, 1996).

Many casein proteins exist in a colloidal particle known as the casein micelle (Fig. 14). Its biological function is to carry large amounts of highly insoluble Calcium phosphate (CaP) to mammalian babies in liquid form and to form a clot in the stomach for more efficient nutrition (Smyth *et al.*, 2004). Besides casein proteins, calcium and phosphate, the micelle also contains citrate, minor ions, lipase and plasmin enzymes, and entrapped milk serum. These micelles are rather porous structures, occupying about 4 ml/g and 6-12% of the total volume fraction of milk. In the submicelle model, it is thought that there are small aggregates of whole caseins, containing 10 to 100 casein molecules, called submicelles**.** It is thought that there are two different kinds of submicelle; with and without κ-CN. These submicelles contain a hydrophobic core and are covered by a hydrophilic coat which is at least partly comprised of the polar moieties of κ-CN. The hydrophilic CMP of the κ-CN exists as a flexible hair. The resulting hairy layer, at least 7 nm thick, acts to prohibit further aggregation of submicelles by steric repulsion. The stability of the casein micelle plays an important role in the protein function and milk nutritional value: more than 90% of the calcium ions content of skim milk are associated in some way or another with the casein micelle. The removal of calcium leads to reversible dissociation of β-CN without micellular disintegration. Otherwise, the addition of calcium leads to aggregation of casein micelles.

Caseins are very sensitive to the pH: lowering the pH leads to dissolution of calcium phosphate until, at the isoelectric point (pH 4.6), all phosphate is dissolved and caseins precipitate. Also the temperature affects the behaviour of these proteins: at 4° C, the dissociation of β-CN from the micelle begins, at 0° C, there is no micellar aggregation; freezing produces a precipitate called cryo-casein. At high temperature casein micelles are modified by other milk proteins, the whey proteins become adsorbed to them, altering the behaviour of the micelle. Casein micelles have a strong tendency to aggregate because of hydrophobic interactions. α_{s1}-CN, α_{s2}-CN, β-CN and κ-CN are present in milk in different isoforms and with a different grade of phosphorylation, which influences their secondary structure. The sites of phosphorylation are both S and T, which can be also close or even consecutives in the amino acid sequence (Galvani *et al*., 2001).

Caseins have been defined as the major cow's milk allergens by some studied during the past years, especially when they are present as high molecular weight aggregate of 60 or 70 kDa, and when new epitopes not present in the monomeric form can be shown to the immune system (Heddleson *et al.*, 1997). The antigenic determinant sites have been individuated by ELISA test for determinant of specific IgE and IgG: 6 for α_{s1}-CN, 6 for β-CN

and 4 for κ-CN (Docena *et al.*, 1996; Elsayed *et al*., 2004). The epitopes are mostly based in the primary sequence rather than in conformation, because these proteins lack an ordered secondary and tertiary structure (Scaloni *et al.*, 2002). α_{s1}-CN has a major antigenic region immunogenic also after protein denaturation (Fig. 15A). It has been determined by specifically IgE recognition, and is situated in the N-terminal and the C-terminal of the protein. The most allergenic region in β-CN is the N-terminal 1-25, rich of E in a polyGlu motif. This peptide is heat stable, but affected by the presence of phosphate, necessary for the protein function of calcium binding protein (Fig. 15B). The study of caseins, extracted from food, is very important to understand milk allergy. These proteins are very resistant to high temperature, so they can constitute the main part of the protein content in a cooked food. Studies conducted during this PhD have shown the particular resistance of caseins, especially α-CN and β-CN, at food processing like milk pasteurization, the mixing with different ingredients and food cooking. Caseins don't degraded with heating at high temperature or with reactions with other food molecules, and can be isolated in a complex and baked food (Parker *et al.*, 2005). The study of their physical and chemical characteristics during the whole cooking process reveals some changes in their structure, which can determine either the loss or the gain of allergenicity by a food (Morales et al. 1996).

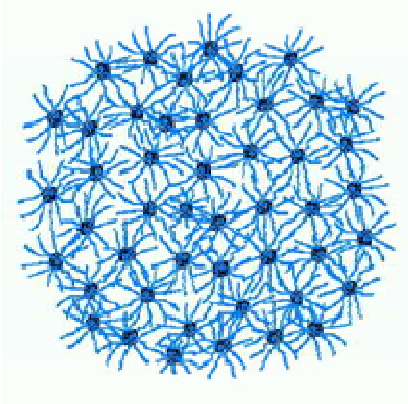


Fig. 14. Representation of a micellar structure formed by caseins of different types and CaPhosphate.

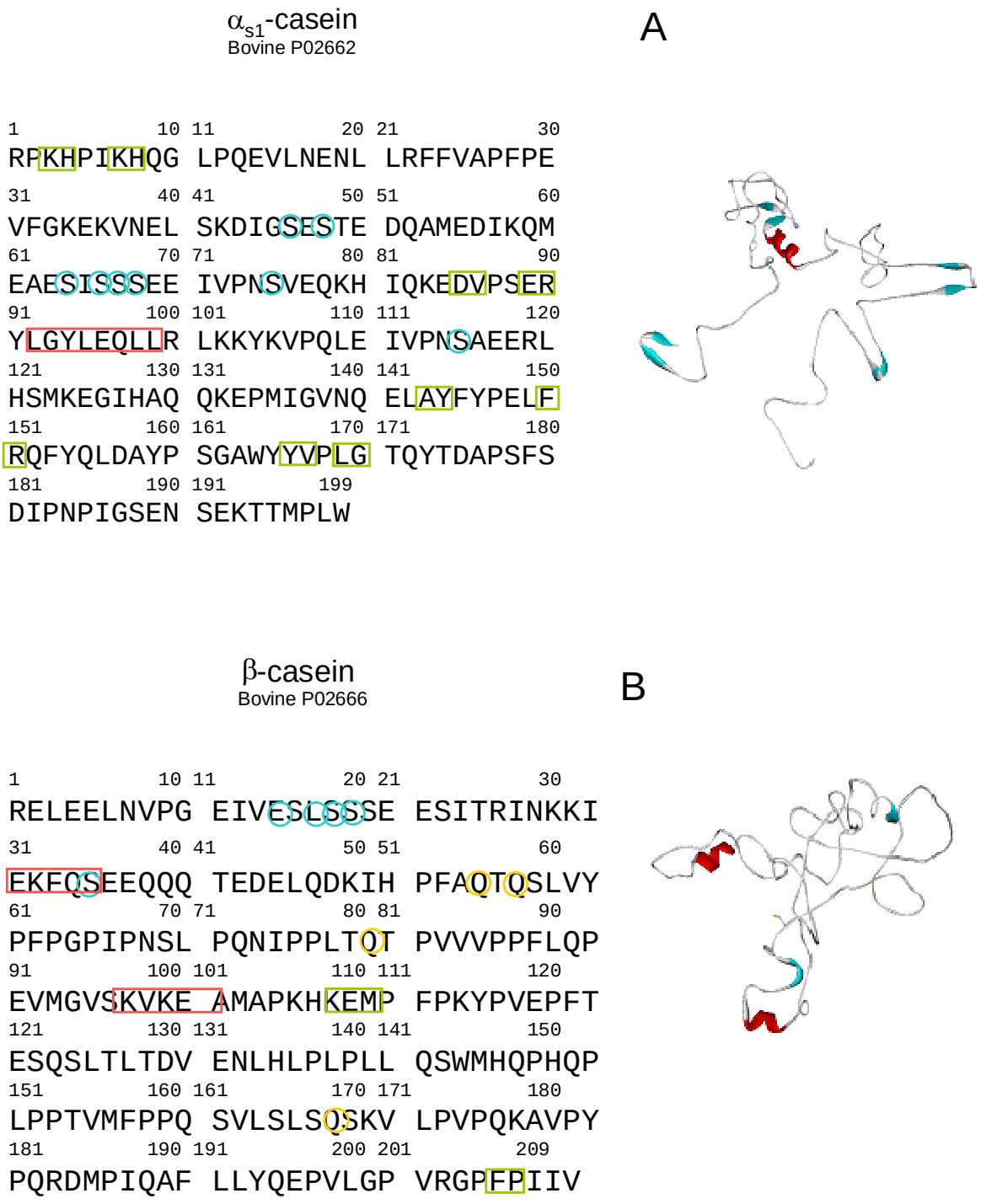


Fig. 15. Amino acids sequence and structure of α_{s1}-CN (**A**). Amino acids sequence structure of β-CN (**B**). The α-helixes are highlighted in red squares, the β-sheet structures in green squares, the phosphorylation in blue circles and the glycosylations in yellow circles. The structure models are generated by Sybyl, a product TRIPOS.

II. MATERIALS AND METHODS

1. Materials

Reagents and solvents of analytical reagent grade for chromatographic analysis and mass spectrometry were obtained from Fluka (Buchs, Switzerland) and from Sigma-Aldrich (Milwaukee, WI, USA) and those for SDS-PAGE electrophoresis from BioRad (Richmond, IL, USA). Proteins from milk and egg-white were purchased from Sigma-Aldrich. Porcine trypsin was obtained from Promega (Madison, WI, USA), furosine (FUR) from Neosystem (Strasbourg, France) and 5-hydroxymethyl-2-furfuraldehyde (HMF) from Fluka. Milk Analysis of milk were conducted using Granarolo and Ala-Zignago products. Eggs, peanuts, soybean and other foods were commercial market products obtaine from local market.

2. Protein extraction and purification

Protein extraction from milk under acidic conditions. A sample of 50 mL of milk are acidified with 1 M HCl, until pH reaches the value of 4.6. Under acidic conditions caseins precipitate and are separated from the skimmed milk by centrifugation at 3000 rpm. The soluble fraction is neutralized at pH 7.0 with 1 M NH_4HCO_3. The acidification and centrifugation are repeated to additionally precipitate caseins eventually still present in the sample. Afterwards, the sample is centrifuged at 15,000 rpm, at 4°C for 40 minutes, in order to eliminate milk fatty acids. To purify the sample from salt, it is dialyzed with a 8000 cut-off membrane, against deionised water, at 4°C overnight. The dialyzed sample is then filtered with a 1.2 μm membrane and the protein concentration of the sample is then measured.

Protein extraction from milk in neutral conditions. A sample of 50 mL of milk are diluted 1:10 with NaCl 0.9% (w/v) and stirred for 2 hours at room temperature. Afterwards, the sample is centrifuged at 15,000 rpm, at 4°C for 40 min in order to eliminate milk fatty acids. To purify the sample from salt, it is dialyzed with a 8000 Da cut-off membrane, against water, at 4°C overnight. The dialyzed sample is filtered with a 1.2 μm membrane and then the protein concentration of the sample is measured.

Protein extraction from egg- white and egg yolk. Egg-white and yolk from two eggs are manually separated and are both suspended in water milliQ at 1:3 ratio. Each suspension is stirred for 2 hours at room temperature. Afterwards, the samples are centrifuged at 3000 rpm for 15 minutes in order to separate the supernatant from the precipitate. The solutions thus obtained are then separated and centrifuged at 15000 rpm at 4°C for 40 min in order to eliminate the remaining precipitate and fatty acids. To purify the samples from salt, they are

dialyzed with a 8000 Da cut-off membrane against water at 4°C overnight. Dialyzed samples are finally filtered with a 1.2 μm membrane and then the protein concentration of the samples is measured.

Protein extraction from row and cooked cake. Cake is made by mixing 250 g of flour, 3 eggs, 600 g of sugar, 200 mL of milk, baking powder, 50 g of butter. It was cooked for 40 min at 180°C. Two aliquots of 30 g of cooked cake are reduced into powder with a mixer and then suspended in 25 mL of acetone. Two aliquots of 30 g of row cake, instead, are directly suspended in 25 mL of acetone to extract the lipids (two times). Once removed the acetone, samples are suspended in 120 mL of 20 mM Tris·HCl for the protein extraction at pH 7.2, or 120 ml of ammonia acetate pH 4.5 for the acid extraction. Samples are stirred for 2 hours at room temperature. Afterwards, the samples are centrifuged at 3000 rpm for 15 minutes in order to separate the supernatant from the precipitate. The solutions are separated and centrifuged at 15,000 rpm at 4°C for 40 min in order to eliminate some precipitate and fatty acids. The samples are then dialyzed with a 8000 Da cut-off membrane against water at 4° C overnight. Dialyzed samples are finally filtered with a 1.2 μm membrane and the protein concentration of the samples is calculated.

Protein extraction from peanuts and soybean. Samples of 30 g of peanuts and 30 g of soybean are reduced into powder with a mixer and suspended in 25 mL of acetone to extract the lipids (two times). Once removed the acetone, samples are suspended in 200 ml of 20 mM Tris·HCl, pH 7.2, stirred for 2 hours at room temperature. Then, samples are centrifuged at 3000 rpm for 15 minutes, to separate the surnatant from the precipitate. The surnatants are separated and centrifuged at 15000 rpm, at 4°C for 40 minutes, in order to eliminate the left precipitate and fatty acids. The samples, in order to be purified from salt, are then dialyzed with a 8000 cut-off membrane, against deionising water, at 4° C overnight. Dialyzed samples are finally filtered with a 1.2 μm membrane and the protein concentration of the samples is analysed.

Protein quantification. The Bradford protein assay is one of several simple methods commonly used to determine the total protein concentration in a sample (Bradford, 1976; Sedmark and Grossberg, 1977). This method is based on the fact the dye Coomassie is capable to bind to proteins. Within the linear range of the assay (~5-25 mcg/mL), the more protein present, the more Coomassie binds. Furthermore, Coomassie absorbs at 595 nm, so the assay is colorimetric and the colour of the test sample becomes darker as the protein concentration increases. The protein concentration of a test sample is determined by

comparison of its absorbance to protein standards known to exhibit a reproducibly a linear absorbance profile in this assay. Although different protein standards can be used, we have chosen bovine serum albumin (BSA) as our standard, because it is the most widely used.

Our procedure implies the preparation of 2 mg/mL BSA stock solution. Different aliquots of BSA are diluted adding 200 ml of Bradford reagent (BioRad) and the quantity of buffer (20 mM Tris·HCl, pH 8.2) necessary to reach 1 mL of the final sample solution. After 10 min of incubation at room temperature, the absorbance of each sample is measured at 595 nm by a spectrophotometer Lambda-25, Perkin Elmer (Norwalk, CA, USA). The reference line is made by plotting the absorbance of each BSA standard as a function of its concentration. The equation of the regression line is used to calculate the concentration of the protein sample based on the measured absorbance.

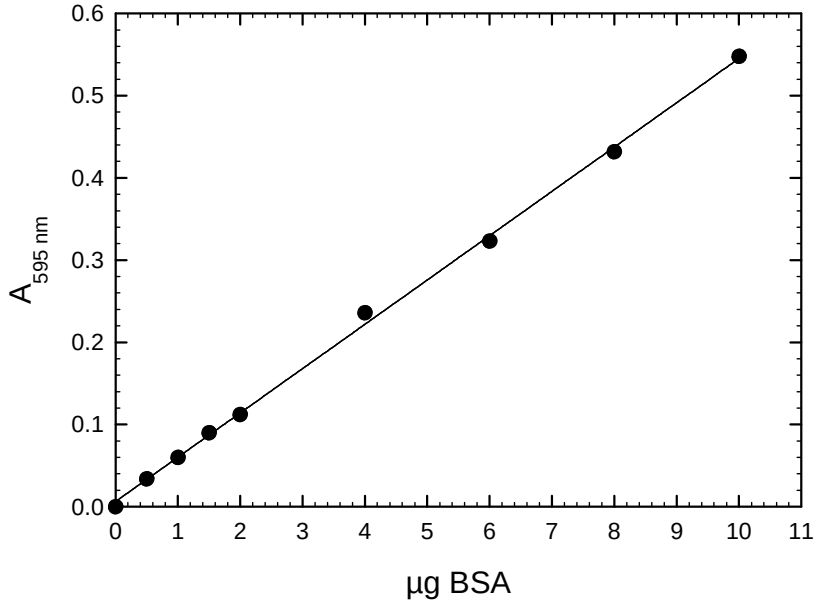


Fig. 20. Reference line obtained plotting the absorbance data at 595 nm in function of the µg of protein.

3. Size-exclusion chromatography

Size exclusion chromatography (SEC) is a chromatographic method in which proteins are separated on the basis of their hydrodynamic volume. To perform the size exclusion chromatography analysis with a AKTA-FPLC (Amersham, Freiburg, Germany) instrument, a Superdex 75 coluns (Pharmacia, Uppsala, Sweden) is used. Proteins elution was conducted with 20 mM Na_2HPO_4, pH 7.5, with a flow rate of 0.4 mL/min. The absorbance of the effluent was monitored at 280 nm. In order to determine the sizes of protein molecules of interest,

standards of a known molecular mass have been run on the same Superdex column and a calibration curve has been obtained.

4. Ion-exchange chromatography

Ion-exchange chromatography is a process that allows the separation of polar molecules based on their charge properties. This type of chromatography is further subdivided into cation and anion exchange chromatography. We used the anion exchange column Source-Q (Pharmacia, Uppsala, Sweden) with the liquid chromatographer FPLC-AKTA. The column was equilibrated with 20 mM Tris·HCl, pH 7.2, and the column was eluted with 20 mM BTP, 1 M NaCl, pH 7.2. The linear gradient applied was from 0 to 100% in 60 min and the flow rate 1 mL/min. The absorbance of the effluent from the column was monitored at 280 nm.

5. RP-HPLC

Reversed phase HPLC (RP-HPLC) consists of a non-polar stationary phase and an aqueous, moderately polar mobile phase, in a high pressure system. The RP-HPLC analysis has been performed with a liquid chromatograph instrument at high pressure purchased from Agilent Technologies, model 1100 (Waldbronn, Germany). The columns used were a Vydac C4 (The Separations Group, Hesperia, CA, USA) and a Jupiter C18 (Phenomenex, Torrance, CA, USA), whose dimensions are 4.6 x 150 mm with particles of diameter 5 μm. Elutions were performed with an acetonitrile gradient containing 0.1% trifluoroacetic acid (TFA). The absorbance of the effluent from the RP-HPLC column was monitored at 226 nm or 280 nm.

6. Protein identification

SDS-PAGE electrophoresis. In SDS polyacrylamide gel electrophoresis (SDS-PAGE) charged proteins migrate in response to an electric field. Their rate of migration depends on the strength of the field and on the net charge, size and shape of the protein molecules. Sodium dodecyl sulphate (SDS) is an anionic detergent which denatures proteins by "wrapping around" the polypeptide backbone. In this project, electrophoresis experiments were conducted with an electrophoresis system Mini-Protean II purchased from BioRad using plates of 10.2 x 8.2 cm and spacer of 0.75 mm thickness. For the separation of proteins the gel was composed by 10, 12 or a 15% T (% total acrylamide and bis-acrylamide) and 0.4% C (% of bis-acrylamide). The running and stacking gel was made by a solution of acrylamide bis-acrylamide Protogel, 30% T and 0.8% C in running and 5%T in stacking gel. The buffer used

was 1.5 M Tris·HCl, pH 8.5, with 0.3% SDS. The acrylamide polymerization was obtained by adding TEMED and a solution 10% of ammonium persulfate. The electrophoresis was conducted with a running buffer 25 Tris·HCl, 250 mM glycine, 0.1% SDS, pH 8.5. Samples were dissolved in 0.3 M Tris·HCl, pH 6.8, with 0.1% SDS, 2.5% β-mercaptoethanol, 50% glycerol and traces of bromophenol blue. Afterwards, the protein samples were denatured heating at 100°C for 15 min. The electrophoresis was performed at room temperature with a power of 18 mA. The resulting spots were stained with Coomassie Brilliant Blue R 250.

7. Fluorescence spectroscopy

The fluorescence emission of a protein is given by the fluorescence of individual aromatic residues. Proteins contain three aromatic amino acid residues (tryptophan, tyrosine, and phenylalanine) which may contribute to their intrinsic fluorescence that is generally measured exciting proteins at 280 nm or at longer wavelengths, usually at 295 nm. The three residues have distinct absorption and emission wavelengths. They differ greatly in their quantum yields and lifetimes. Due to these differences and to resonance energy transfer from proximal phenylalanine to tyrosine and from tyrosine to tryptophan, the fluorescence spectrum of a protein containing the three residues usually resembles that of tryptophan and most of the fluorescence emission is due to excitation of this residue. The intensity of fluorescence is not so informative in itself. The magnitude of intensity, however, can serve as a probe of perturbations of the folded state. The wavelength of the emitted light is a better indication of the environment of the fluorophore. Tryptophan residues that are exposed to water have maximal fluorescence at a wavelength of 345-355 nm, whereas totally buried residues fluoresce at 320-330 nm.

The fluorescence experiments were conducted with a spectrofluorimeter Perkin-Elmer LS-50 using quartz cells of pathlengths of 1 cm at 25°C. Spectra were obtained exciting at various wavelengths and registering the emission spectra in a following wavelength interval. Samples were analyzed in 20 mM Na_2HPO_4, pH 7.5, buffer.

8. Acid hydrolysis of food proteins and analysis of furosine and furfural

Furosine (FUR) and free fufural compounds can be considered indicators of the extent of the Maillard reaction (Erbersdobler and Somoza, 2007). FUR is the indicator of earlier reactions and gives an estimate of the blocked, not reactive lysine residues. Furfurals (HMF) are intermediary compounds in the most advanced stages of the Maillard reactions. FUR and HMF concentrations were measured using RP-HPLC with UV detection at the absorbing

wavelength of each compound. In order to quantify the sample content of FUR and furfurals, the protein samples were acid hydrolyzed to free amino acids with acid. 0.5 mL of protein sample (1 mg/ml) was hydrolyzed in a glass tube in the presence of 6 N HCl. The tube with The sample was then analyzed by RP-HPLC, detecting FUR at 280 nm and HMF (hydroxymethylfurfural) at 284 nm. Calibration curves (0.02-0.2 μg samples) were carried out by plotting the absorbance of standard samples as expressed in area units *vs*. μg of FUR or HMF.

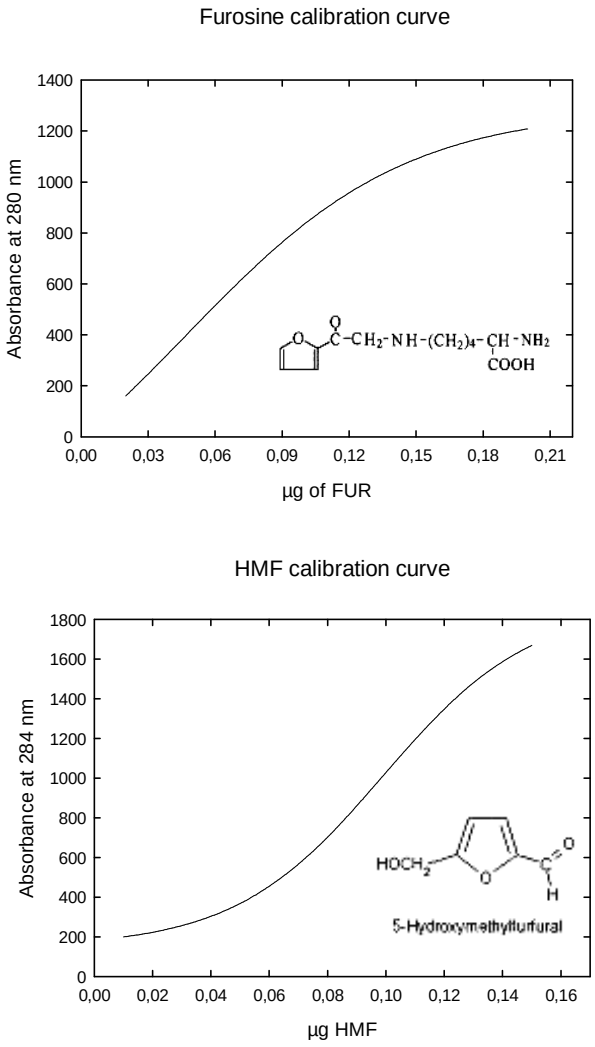


Fig. 21. (**A**) Reference curve obtained plotting the absorbance data at 280 nm in function of the μg of FUR. (**B**) Reference curve obtained plotting the absorbance data at 284 nm in function of the μg of HMF.

9. Fingerprintig analysis

Fingerprinting with trypsin is a most used method to anlyse a protein. The procedure implies first reduction of disulfide bridges of the protein to free thiols and subsequent S-alkylation of the free thiols. Then the S-alkylated protein is reacted with trypsin to produce a tryptic digest and then the identity of peptide fragments is determined by mass spectrometry. The fingerprinting methodology can be carried out with aprotein sample in solution or a protein in a gel spot.

In the protocol used by us for the fingerprinting analysis in solution, Tris(2-carboxyethyl)fosfine (TCEP) was used to reduce the disulfide bridges. The reagent was at a molar ratio 1:10 (reagent:Cys in the protein substrate) in 50 mM Tris·HCl, pH 8.9, containing 6 M Gdn·HCl. After incubation of the protein at 37°C for 30 min, cysteines were blocked with iodoacetamide (IAA). The IAA reaction was conducted in the same buffer at room temperature and in the dark for 30 min at a molar ratio of 1:10 (reagent:cysteine). After the desalting the sample by RP-HPLC, trypsin was added in a 1:25 (enzyme:substrate) ratio. The reaction was conducted in 250 μL of 0.1 M NH_4HCO_3, pH 8.3, at 37°C overnight. The trypsin digestion was stopped by adding 5 μL of 50% TFA.

The protocol of the in gel trypsin digestion was conducted according to the procedure described by Shevchenko *et al.* (1996). The gel spot was excised from the gel and furtherttreate as follows. The gel pieces were washed repeatedly in an Epperdof tube with milliQ water and acetonitrile until the Coomassie Brilliant Blue was removed and then dried. The gel pieces were rehydrated with 0.1 M NH_4HCO_3 containing 10 mM DTT in order to achieve reduction of disulfibe bridges and then incubated at 56°C for 30 min. After an acetonitrile wash, a solution of 55 mM IAA in 0.1 M NH_4HCO_3 was added to the sample in order to block the cysteines. The sample was incubated at room temperature for 20 min in the dark and then the gel pieces were washed with milliQ water and acetonitrile. The sample was reacted with 50 ng/mL of trypsin solution in 50 mM NH_4HCO_3, and left at 4°C for 1 hour. Trypsin can enter into in the gel pieces and perform the hydrolysis of the protein sample overnight at 37°C. The resulting peptide mixture was collected, washed with acetonitrile containing 5% formic acid and then dried.

10. Mass spectrometry ESI Q-Tof and LC-MS/MS

A mass spectrometer determines the mass of a molecule by measuring the mass-to-charge ratio (*m/z*) of its ion. Ions are generated by inducing either the loss or gain of a charge from a neutral species. Once formed, ions are electrostatically directed into a mass analyzer

where they are separated according to *m/z* and finally detected. The result of the molecular ionization, ion separation and ion detection is a spectrum that can provides an estimate of molecular During this project, the mass spectrometry analysis was conducted with a Micromass mass spectrometer QTof-micro (Manchester, UK), connected with a micro-HPLC CapLC (Waters, Milford, MA, USA). Tryptic digests were dissolved in 0.1% formic acid and analyzed with a column Symmetry C_{18} nano-ESI, 75 μm x 150 mm (Waters). Elution of the column was performed with milliQ water containing 5% acetonitrile and 0.1% formic acid (solvent A) and acetonitrile containing 5% water and 0.1% formic acid (solvent B). The linear gradient of solvent B was from 5 to 30% in 30 min at a flow rate of 200 nL/min.

11. Database search

The MS/MS spectra were analyzed with the MASCOT software (http://www.matrixscience.com) for the database research, selecting trypsin as proteolytic enzyme and S-carbamidomethylation (-S-CH_2CONH_2) of cysteines. We have not used a restriction of the database for a specific organism and some possible modifications of the amino acid side chains were considered, such as oxidation of Met nad phosphorylation of Ser, Thr and Tyr.

III. RESULTS AND DISCUSSION

1. Extraction and purification of allergenic proteins

The extraction of proteins from milk has been performed under neutral or acidic conditions in order to study the total protein content or the serum proteins devoid of caseins. The total casein fraction represents 80% of the milk proteins. The extraction of proteins from egg white, yolk, peanut and soybean was conducted following the procedures described under Materials and Methods: The protein concentration in the samples was determined by the Bradford assay and then protein samples were analyzed by SDS-PAGE electrophoresis (Fig. 18).

The milk protein extract under acid conditions was analyzed by size exclusion chromatography (SEC) (Fig. 19) and the analysis of the protein material contained in the chromatographic peaks was made by RP-HPLC and mass spectrometry (MS) (Table 1). The first component eluted from the chromatographic column at 20 min was bovine serum albumin (BSA). The shoulder of this peak at 18 min was shown to be aggregated β-lactoglobulin (β-LG), while monomeric β-LG eluted at 24 min and α-lactalbumin (α-LA) 26 min. β-LG has five cysteines (Cys) residues, four of them forming two disulphide bridges. The single free Cys usually makes inter-molecular disulfide bridges with other molecules of β-LG, causing protein aggregation. The mass values calculated of β-LG and α-LA were 18367.5 Da and 14180.0 Da, respectively, values in good agreement with the theoretical masses. The BSA mass value of 69212 Da measured by MS was different from the theoretical one, partly because the bigger the molecule, more imprecise the measure, and for the post-translational modifications of the protein. Another minor peak in the SEC chromatogram was at 42 min, due to the presence of salts and low molecular weight compounds in the sample. The second sample analyzed was a standard mix of milk proteins including serum proteins and caseins, as a protein extracted from milk at neutral conditions (Fig. 20). The chromatogram showed that the proteins were not completely separated, and the protein identification by RP-HPLC and mass spectrometry revealed that BSA, α-CN, β-CN and κ-CN eluted together in the first and second peaks (Fig.23, Table 2). These peaks corresponded to BSA, but considering the upper limit of the column all proteins with a MW higher than 66 kDa would elute in this peak. At physiological concentrations, caseins are aggregated in the micellar form, so it was expected that they elute at higher molecular weight. The MW of caseins determined by MS were heterogeneous and higher than expected, likely due to post-

translational modifications as phosphorylation. β-LG and α-LA eluted at the same retention time of the previous experiment and the figures for their molecular weight were close to the theoretical ones. SEC is a suitable technique to purify milk serum proteins, but it can not be used for separating caseins due to their chemical and physical characteristics.

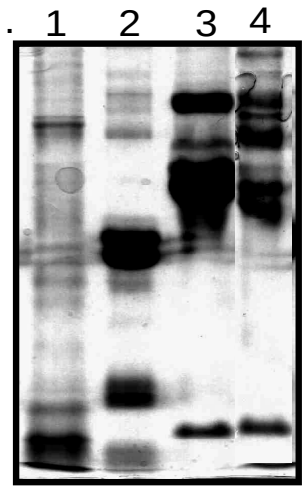


Fig. 18. SDS PAGE electrophoresis analyses of the protein extract from flour (1), milk (2), egg white (3) and yolk (4).

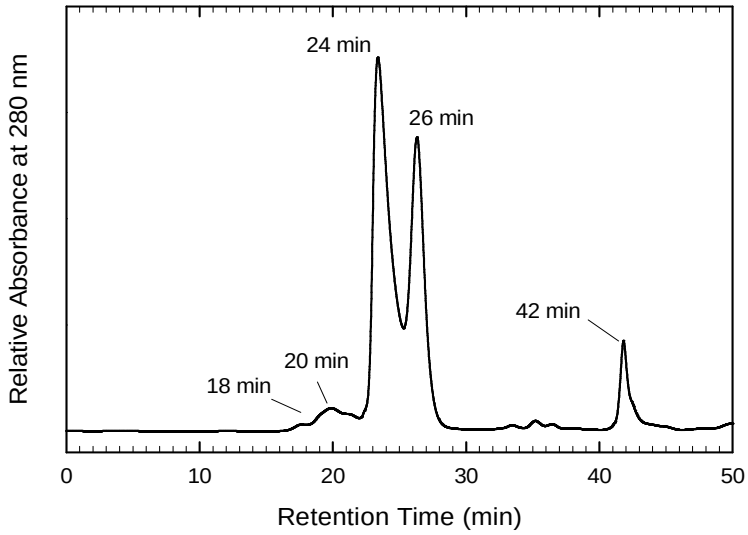


Fig. 19. Gel filtration chromatography (SEC) of milk serum proteins extracted from milk. The column used was a Superdex 75, eluted with 20 mM Na_2HPO_4, at a flow rate of 0.4 mL/min.

Table 1. MW determination by mass spectrometry of the proteins correspondent to the chromatographic peaks analyzed by SEC (Fig. 20).

RT-HPLC (min)	MW Found[a](Da)	MW Theoretical[b](Da)	Protein
30.2	69212.1	66398.9	BSA
30.5	18367.5	18363.3	β-LG
31.5	14180.0	14178.0	α-LA
34.1	18369.3	18363.0	β-LG

[a] Determined by mass spectrometry
[b] Calculated ion mass values, obtained by applying cleavage rules to the entries in a primary sequence database

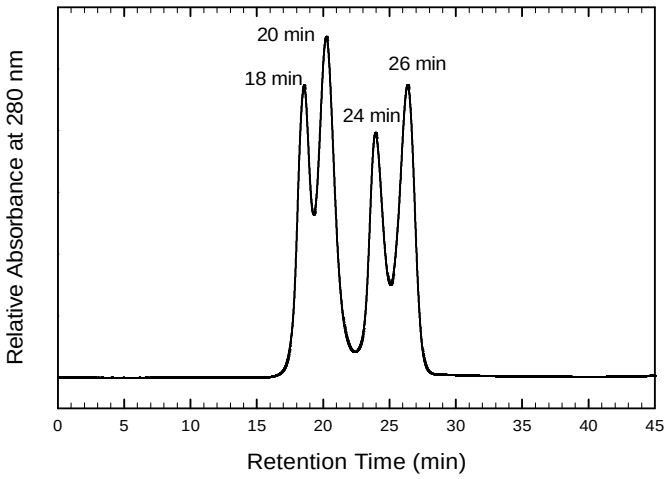


Fig. 20. SEC of a protein mixture (1 mg/mL) of milk standard proteins α-LA, β-LG, BSA, α-CN, β-CN and κ-CN. The column used was a Superdex 75 eluted with 20 mM Na_2HPO_4 at a flow rate of 0.4 mL/min.

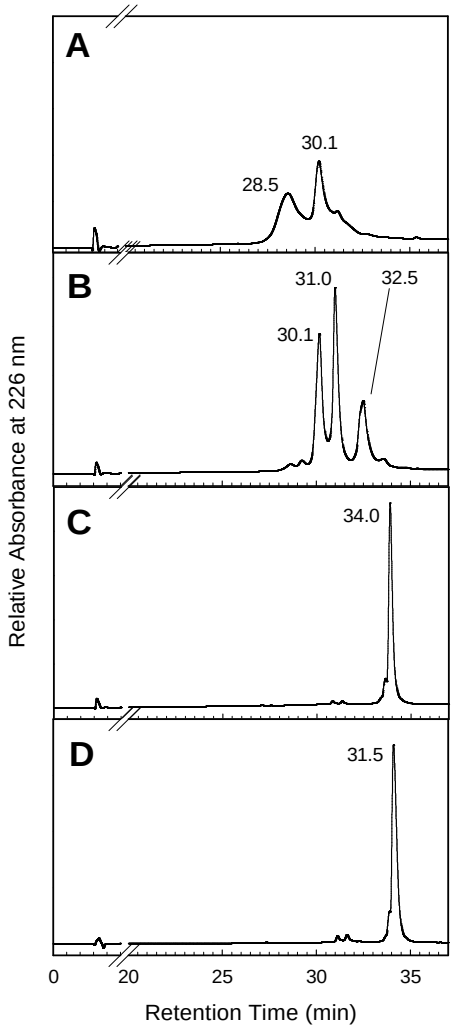


Fig. 21. RP-HPLC chromatography of the protein samples corresponding to the eluted peaks on Fig. 20. The peaks analyzed were RT 18 min **(A)**, 20 min **(B)**, 24 min **(C)** and 26 min **(D)**. The column Vydac C4 has been eluted with water and acetonitrile, with 0.1% (TFA), with an acetonitrile gradient from 5 to 34% in 15 min, from 34 to 60% in 23 min, at a flow rate of 0.6 mL/min.

Table 2. MW determination by mass spectrometry of the proteins correspondent to the chromatographic peaks analyzed by SEC and RP-HPLC (Fig. 21).

RT-HPLC (min)	MW found[a](Da)	MW theoretical[b](Da)	Protein
28.5	21060.5	18972.4	κ-CN
30.1	66561.6	66398.9	BSA
31.0	24029.2	24346.5	α-CN
31.5	14182.0	14178.0	α-LA
32.5	23989.9	23583.2	β-CN
34.0	18371.3	18363.0	β-LG

[a] Determined by mass spectrometry.
[b] Calculated ion mass values, obtained by applying cleavage rules to the entries in a protein sequence database.

The extraction of proteins from peanuts, including the major protein allergen Ara h1, was performed according to the extraction protocol described under Materials and Methods and then the protein extract was purified by ion exchange chromatography. The column used was a Source-Q anionic exchange and the column was eluted with a linear gradient of 1 M NaCl (Fig. 22). This chromatographic step allowed only a partial protein separation, because of the large quantity of protein applied to the column and heterogeneity of the sample. The chromatographic peaks were analyzed by SDS-PAGE electrophoresis. This analysis of the chromatographic peaks showed a band of 66 kDa in all the fractions (Fig. 23). The spot at 66 kDa could be ascribed to Ara h1, which has a MW of about 70 kDa. The identity of the protein was confirmed by the LC-MS/MS experiment, conducted on the major band at 66 kDa from the peak at 41 min (Fig. 24). The search in the database after the MS analysis detected three tryptic peptides, with an overall 5% of the protein sequenced. The protein was clearly identified as Ara h1. The relatively low sequence coverage by fingerprinting could be due to several post-translational modifications of Ara h1 or to modifications made by cooking (Maleki *et al.*, 2000). Other SDS-PAGE electrophoresis experiments allowed us to identify additional peanut proteins of lower MW, like Ara h2 of 17 kDa, as the band at low MW in the peak eluted at 34 min (data not shown).

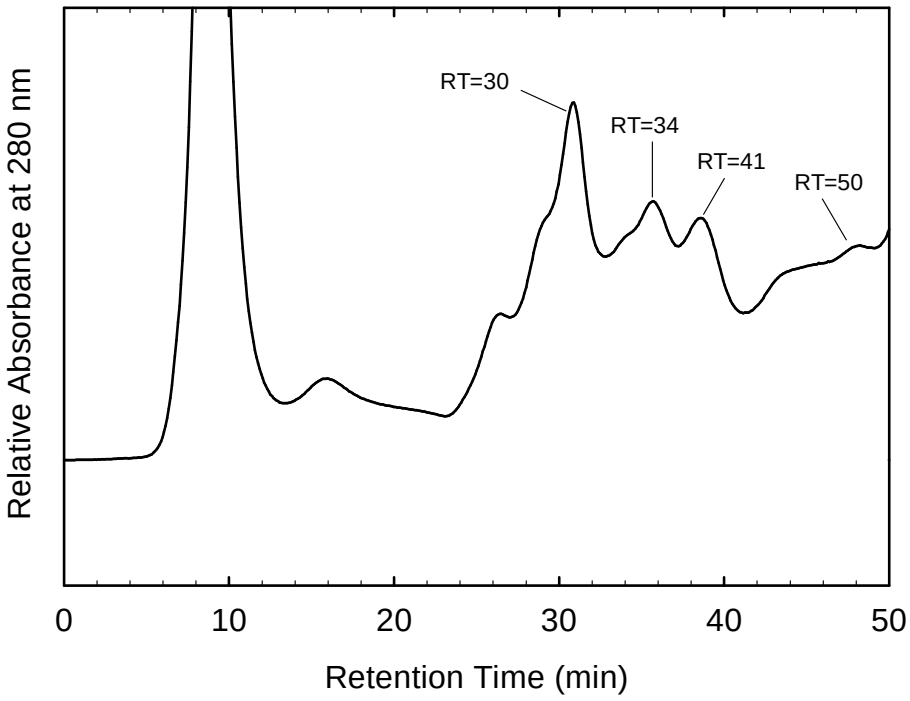


Fig. 22. Ion exchange chromatography (IEX) of a protein mix extracted from peanut. For the separation, a column Source Q was used, equilibrated with 10 mM BTP, pH 7.2. The elution of the proteins has been obtained with an NaCl gradient, from 0 to 10% in 5 min, from 10 to 40% in 15 min, with a flow rate of 0.5 mL/min.

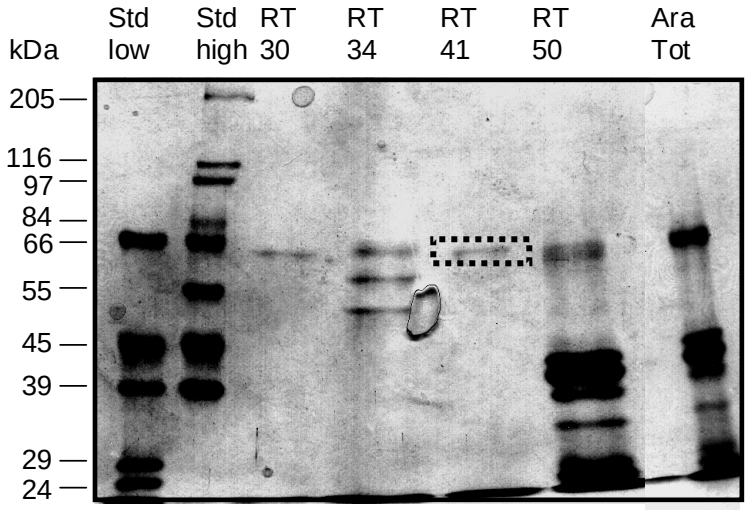


Fig. 23. SDS PAGE electrophoresis analysis of the sample corresponding to the peaks eluted from IEX in Fig. 23. In Ara Tot a sample from the entire protein extract was loaded; in Std low and St high, standard proteins of known MW were loaded. The spot analyzed by LC-MS/MS is highlighted with a broken line.

gi|1168390 **Mass:** 70639 **Score:** 100 **Queries matched:** 3
Allergen Ara h 1, clone P17 precursor (Ara h I)
☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	22	564.8226	1127.6306	1127.6298	0.0008	0	61	0.00048	1	K.GTGHLELVAVR.K
☑	24	571.2639	1140.5132	1140.5352	-0.0220	0	19	8.6	1	R.HNPFYFPSR.R
☑	52	688.8356	1375.6566	1375.6619	-0.0052	0	20	8.3	1	K.DLAFPGSGEQVEK.L

Fig. 24. MASCOT result obtained from the databank NCBI search, for the gel spot selected at 66 kDa in lane RT41. The score 100 is bigger than 41, the lower limit, indicating the reliability of the data. The peptides found, with the theoretical and found MW and the sequence, are reported.

Soybeans proteins were extracted as reported under Material and methods. The SDS-PAGE analysis showed the presence of different proteins in a wide range of MW (Fig. 25). The main aim of our study was the identification of the major soybean allergen Gly m Bd 28K, a protein of 26 kDa. Therefore, the region of interest in the SDS-PAGE gel was that corresponding to 24-30 kDa fragments. The gel bands in this region were cut and analyzed by fingerprint and LC-MS/MS. The proteins identified by MS were soybean proteins, like trypsin inhibitor A, KTi1 and KTi2, the seed maturation proteins PM30 and PM22, the allergenic proteins glycinin G1 and soybean agglutinin (Table 3). Proteins Gly m Bd 30K and Gly m Bd 28K, the major allergens, were not found. The possibility was that these proteins could be aggregated or contained in very small amounts in the soybean seed, thus not permitting their identification.

In order to enrich the sample in Gly m Bd 28K, the whole soybean protein extract was fractionated by size exclusion chromatography (SEC) (Fig. 26). The resulting SEC chromatogram showed the partitioning of the protein extract in four main fractions, each of them, by comparison with protein standards, had a MW as 70, 50, 45 and 30 kDa. The fractions collected were then analyzed by SDS-PAGE (Fig. 27). Two groups of proteins of MW 28.0 kDa were observed by this analysis, one in the first fraction and the other in the last. The gel spots were analyzed by fingerprinting and LC-MS/MS and also other gel bands of other MW were analyzed in order to have a more complete analysis of the protein extract. The results of database search showed the presence of different soybean proteins in the 30 kDa chromatographic fraction, but the protein of interest Gly m Bd 28K was not present. Consequently, the 30 kDa spots of the 70 kDa chromatographic fraction were analyzed according the same protocol and several allergenic proteins were identified: glycinin G2 and Glycinin G1, beta-conglycinin alpha chain and Gly m Bd 28K (Table 4). Other proteins were identified in the same band and in the spot at 70 kDa. This analysis showed that the majority of soybean proteins were aggregated and eluted at high MW in a size exclusion chromatography separation. Only under the reductive and denaturing conditions of SDS-PAGE electrophoresis the aggregates disaggregated. It was possible to identify the soybean allergen Gly m Bd 28K only after a purification step, which appears therefore to be crucial in order to improve the methodology of protein allergen detection.

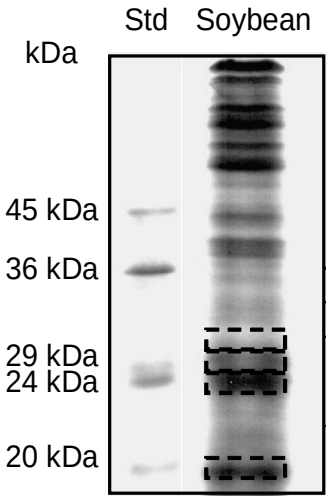


Fig. 25. SDS PAGE electrophoresis analysis of the proteins extract from soybean. In Std standard proteins of known MW were loaded. The spots analyzed by LC-MS/MS are highlighted with a broken line.

Table 3. LC-MS/MS of tryptic peptides from soybean bands, highlighted in Fig. 26.

Sample	Protein	Theoretical MW[a]	MW in gel (kDa)	Peptides sequenced	Sequence coverage (%)[b]
Soybean	Glycinin G1	32.6	30	2	4
	Trypsin inhibitor A	24.3	30	46	28
			29	34	39
			26	2	10
	Trypsin inhibitor KTi2	23.0	30	3	16
	Trypsin inhibitor C	20.4	29	18	44
	Trypsin inhibitor KTi1	22.8	29	2	10
	PR10-like protein	13.6	26	3	17
	Seed maturation protein PM30	15.1	26	2	15
	Seed maturation protein PM22	16.7	26	3	17
	2S albumin	19.0	20	6	27
	Soybean agglutinin	12.8	20	1	17

[a]Calculated ion mass values, obtained by applying cleavage rules to the entries in a primary sequence database.
[b] Sequence coverage calculated on protein precursor.

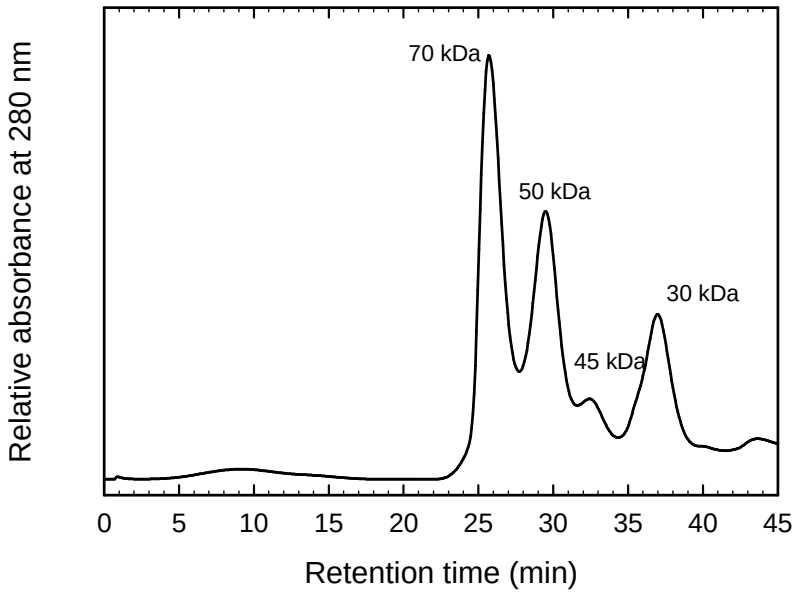


Fig. 26. SEC of a protein extract from soybean. The column used was a Superdex 75, eluted with 20 mM Na_2HPO_4, at a flow rate of 0.3 mL/min.

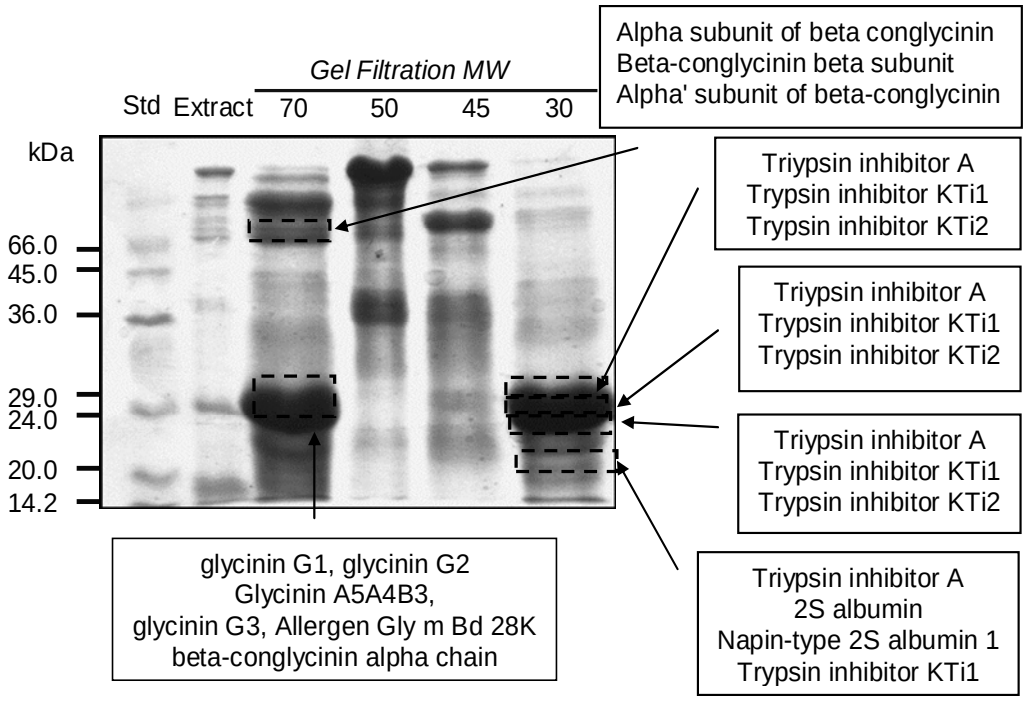


Fig. 27. SDS PAGE electrophoresis analysis of the sample corresponding to the peaks eluted from SEC in Fig. 26. In Extract a sample from the entire protein extract was loaded. in Std standard proteins of known MW were loaded. The spots analyzed by LC-MS/MS are highlighted with a broken line, in the box beside are listed the proteins identified.

Table 4. LC-MS/MS of tryptic peptides from the soybean bands highlighted in Fig. 27.

Sample	Protein	Theoretical MW (kDa)[a]	MW in gel (kDa)	Peptides sequenced	Sequence coverage (%)[b]
Soybean GF 70	Glycinin G1	32.6	30	32	32
	Glycinin G2	52.4	30	12	19
	Glycinin G3	54.9	30	17	12
	Glycinin A5A4B3	64.2	30	7	9
	Allergen Gly m Bd 28K	52.6	30	5	16
	beta-conglycinin alpha chain	70.5	30	2	2
	beta-conglycinin alpha chain	63.1	66	52	26
	beta-conglycinin beta chain	48.3	66	24	29
	beta-conglycinin alpha chain	65.1	66	25	11
Soybean GF 30	Trypsin inhibitor A	24.3	30	41	37
			29	39	47
			26	38	39
			20	7	27
	Trypsin inhibitor KTi1	22.8	30	12	20
			29	11	31
			26	8	20
	Trypsin inhibitor KTi2	23.0	30	4	16
			29	6	19
			26	5	16
	Trypsin inhibitor C	20.4	29	25	51
			26	24	44
	2S albumin	19.0	20	11	27
	Napin-type 2S albumin 1	18.4	20	3	21

[a] Calculated ion mass values, obtained by applying cleavage rules to the entries in a primary sequence database.

[b] Sequence coverage calculated on protein precursor

2. Analysis of allergenic proteins isolated from milk and egg white

α-Lactalbumin and β-Lactoglobulin. During this part of the project, we analyzed allergen proteins by fingerprint, to characterize the commercial proteins in order to have a standard reference for the analysis of proteins extracted from foods. The fingerprint technique is based on the trypsin digestion of proteins and the subsequent analysis by mass spectrometry. The standards analyzed were: α-LA, β-LG, Lys, OM and OA, which are all well known allergens in common food. α-LA and β-LG were also isolated from milk by the protocol reported previously, then analyzed by fingerprinting.

The commercial β-LG was reduced and carbamidomethylated, following the reaction by RP-HPLC (Fig. 28A, B, C). The starting sample was a mixture of two variants A and B. Only the reduced and blocked form of β-LG was collected and submitted to fingerprint, the peptides were separated by RP-HPLC and analyzed by electrospray mass spectrometry. Almost all the peptides theoretically derived from the digestion were found by mass spectrometry analysis (Table 6), and the sequence coverage was very high, 97.3% (with 158 amino acids sequenced in an overall of 162). The analysis permitted also the identification of the two forms, A and B in the peptide T5, from the amino acid 61 to 69, where an Asp in the A form is substituted with a Gly in the B form. The same protocol was used for β-LG from bovine milk (Fig. 28 D, E, F). As the standard one, both A and B variants were present, and the ESI mass spectra of the protein showed the presence of a glycated form of β-LG A coeluting with the native one, as the product of the reaction between the protein and the sugar lactose during the pasteurization process (Fig 29). The molecular mass found identified the variant β-LǴA and β-LǴA plus 324 Da, the exact molecular mass of a dimeric sugar like lactose. The fingerprint result, analyzed by RP-HPLC, was very similar to the commercial protein, showing a total sequence identity. The mass spectrometry analysis confirmed these observations (Tab. 6) and the sequence coverage was the 82.7%, which is lower than the standard, because of the modification of the protein caused by the thermal pasteurization. The peptide T5 permitted the identifications of the β-LG isoforms, as for the standard protein.

The α-LA standard was reduced and carbamidomethylated, then digested with trypsin. The reaction was followed by RP-HPLC chromatography (Fig. 30A, B, C). The fingerprint peptides were analyzed by ESI mass spectrometry and the MW were compared with the theoretical values (Table 7). All the peptides were found and the sequence coverage was of 100 %. The same analysis was performed on α-LA extracted from bovine milk with the

protocol described before (Fig 30 D, E, F). The ESI mass spectrometry analysis of the peptides gave complete sequence coverage of the protein (Table 7).

L*ysozyme, Ovomucoid and Ovalbumin.* Other allergen proteins, basically commercial standards, were analyzed, Lys, OM and OA from egg white were processed with a fingerprint analysis in solution, to have a reference for the subsequent experiments. The proteins were reduced and carbamidomethylated before the enzymatic digestion. The peptides resulted from the fingerprint experiment were analyzed by RP-HPLC chromatography and ESI mass spectrometry. The Lys tryptic digestion was good, as showed in the RP-HPLC chromatogram of the resulted peptides (Fig. 31A) and 81.4 % of sequence was determined, with 105 amino acid above the 129 in the sequence (Tab. 8). OM and OA proteolysis were influenced by the fact that both of proteins were trypsin inhibitors; therefore, even if denatured, can be partially resistant to proteolysis (Van der Plancken *et al.*, 2004). The experimental conditions were slightly modified for OM and OA, because of their natural function of trypsin inhibitors. As a matter of facts, these egg white proteins needed stronger reaction conditions, and the enzyme: substrate ratio was elevated to 1:20. The proteins reduction and alkylation were optimized, especially for OM, which has 18 cysteines in its sequence. The peptides produced by OM proteolysis were separated by RP-HPLC (Fig. 31B), then analyzed by ESI mass spectrometry. 8 peptides were identified and the OM sequence coverage was the 65 % (Tab. 9). This result was enough to determine the identity of the protein. OA has the peculiar characteristic of four free cysteines in its sequence, and therefore the protein reduction and alkylation have a mild effect on its trypsin inhibitory activity and, only with a high enzyme concentration, was finally possible a partially protein digestion (Tatsumi & Hirose, 1997; Rupa & Mine, 2003). The peptides were separated by RP-HPLC chromatography, where some non digested protein was found (Fig. 31C). 9 peptides were identified by ESI mass spectrometry, with sequence coverage of 25.7 % (Tab. 10).

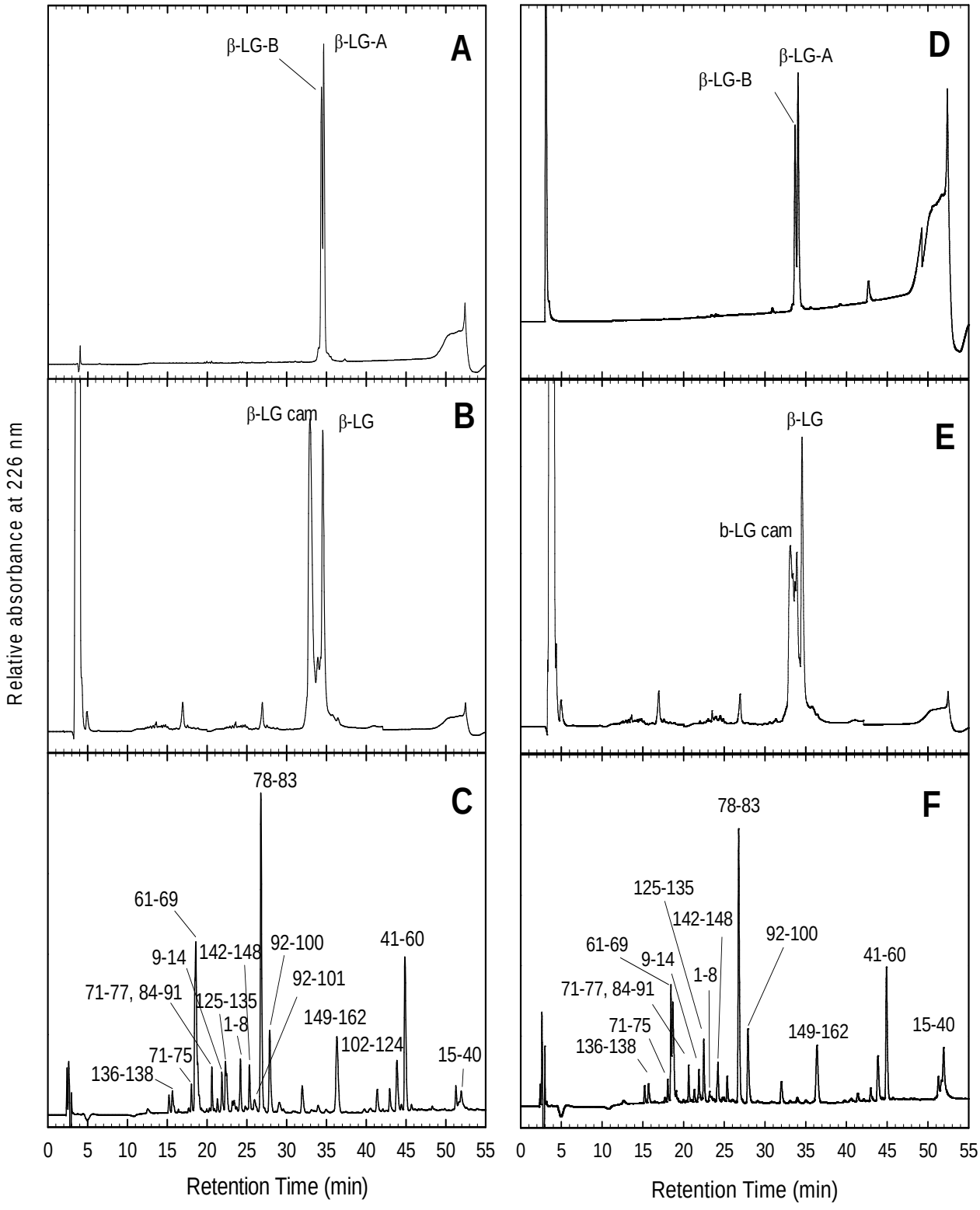


Fig. 28. RP-HPLC chromatography of β-LG standard: native **(A)**, reduced and carbamidomethylated **(B)**, and after trypsin digestion **(C)**. RP-HPLC chromatography of β-LG extracted from milk serum, native **(D)**, reduced and carbamidomethylated **(E)**, and after trypsin digestion **(F)**. The column Jupiter C18 has been eluted with water and acetonitrile, with 0.1% (TFA), with an acetonitrile gradient from 5 to 34% in 15 min, from 34 to 60% in 23 min, at a flow rate of 0.6 mL/min.

Table 5. LC-MS/MS molecular weight determination of the peptides derived from the tryptic digestion in Fig. 28.

Fragment	RT[a] (min)	Found[a] MW (Da) standard protein	Found[a] MW (Da) food protein	Theoretical[b] MW (Da)	Sequence
T15	15.7	408.24	408.23	408.20	136-138
T7	18.1	572.41	572.41	572.35	71-75
T5	18.5	1121.00 1178.50	1121.49 1178.51	1120.46 1178.47	61-69 form B 61-69 form A
T7-8 T10	20.6	802.44 915.53	802.44 915.53	801.50 915.47	71-77 84-91
T2	21.8	672.40	672.42	672.38	9-14
T14	22..4	1244.56	1244.59	1244.58	125-135
T1	24.3	932.54	932.59	932.54	1-8
T17	25.3	836.54	836.52	836.47	142-148
T11-12	26.1	1193.78		1192.67	92-101
T9	26.7	673.48	673.44	673.42	78-83
T11	27.9	1063.62	1064.62	1064.58	92-100
T18	36.3	1714.80	1714.80	1714.80	149-162
T13	44.5	2818.29		2818.26	102-124 form B
T4	44.8	2312.76	2312.82	2313.67	41-60
T3	51.9	2707.10	2707.65	2708.09	15-40

[a] Determined by mass spectrometry.
[b] Calculated ion mass values obtained by applying cleavage rules to the entries in protein sequence database.

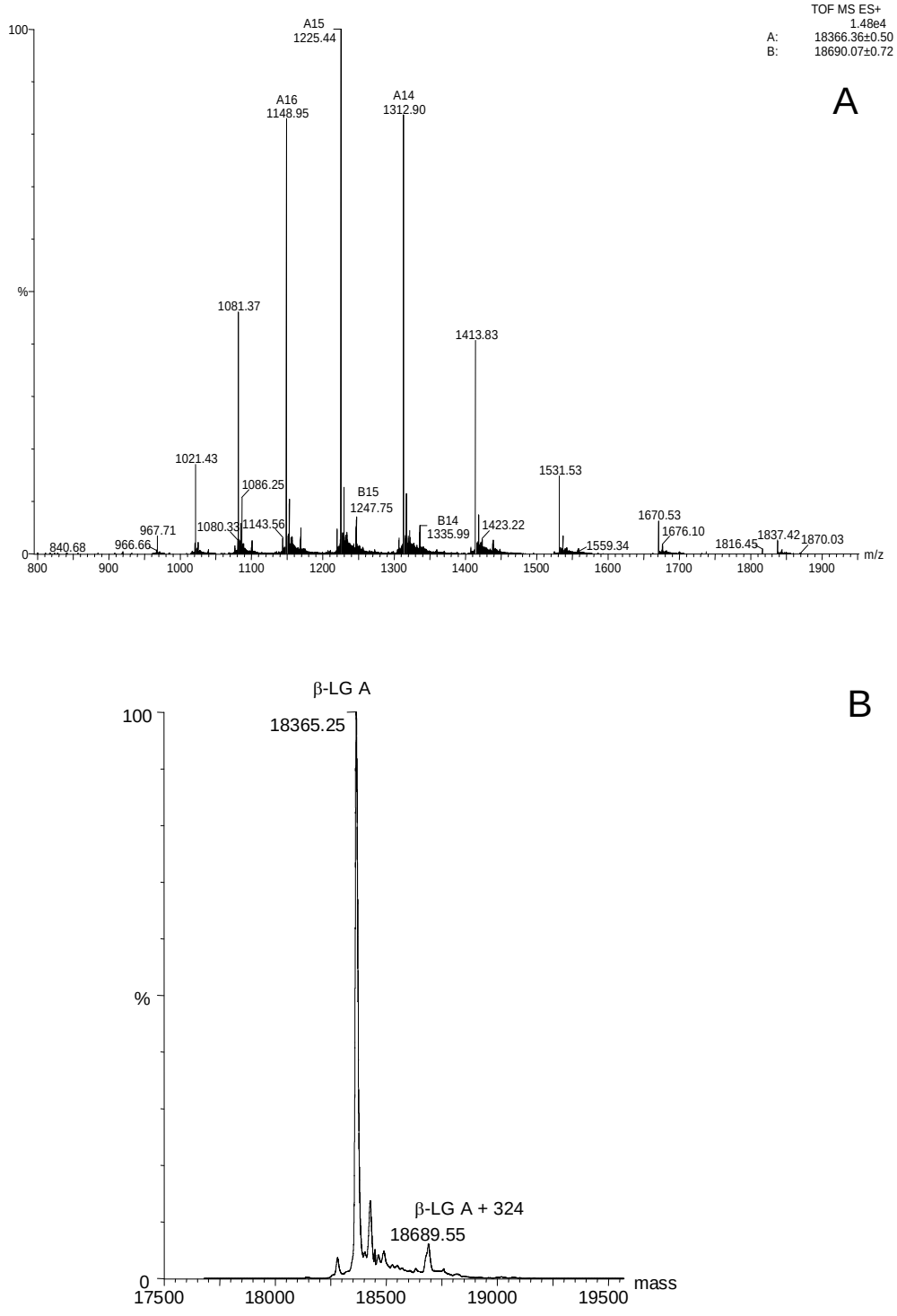


Fig. 29. ESI mass spectrometry analysis of β-LG extracted from milk. MS spectra of the sample at different charge, expressed as m/z **(A)**. Deconvoluted spectra of the peaks in Fig 30A **(B)**.

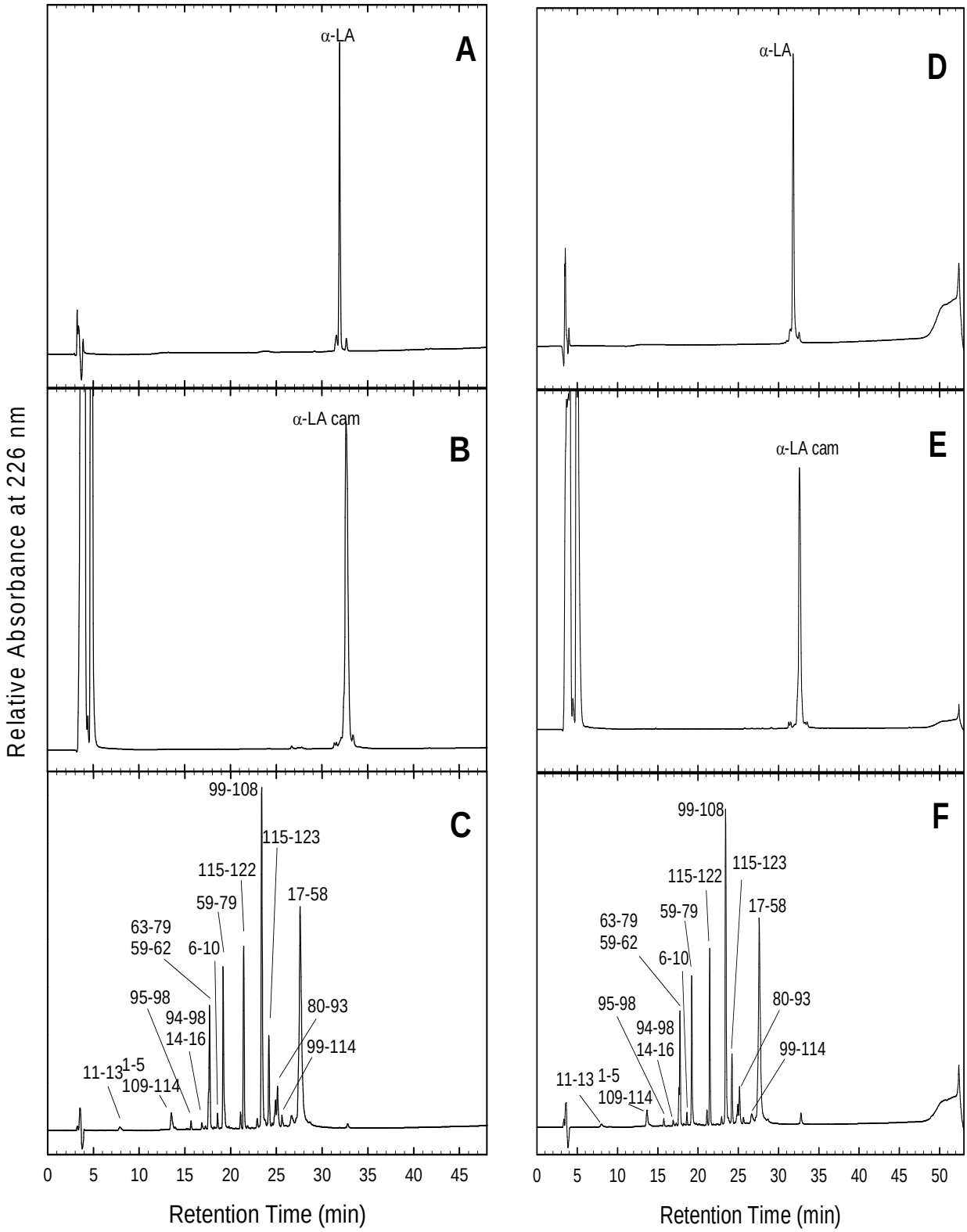


Fig. 30. RP-HPLC chromatography of α-LA standard: native **(A)**, reduced and carbamidomethylated **(B)**, and after trypsin digestion **(C)**. RP-HPLC chromatography of α-LA extracted from milk serum, native **(D)**, reduced and carbamidomethylated **(E)**, and after trypsin digestion **(F)**. The column Jupiter C18 has been eluted with water and acetonitrile, with 0.1% (TFA), with an acetonitrile gradient from 5 to 34% in 15 min, from 34 to 60% in 23 min, at a flow rate of 0.6 mL/min.

Table 6. LC-MS/MS molecular weight determination of the peptides derived from the tryptic digestion in Fig. 30.

Fragment	RT[a] (min)	Found[a] MW (Da) standard protein	Found[a] MW (Da) food protein	Theoretical[b] MW (Da)	Sequence
T3	8.1	388.23	388.21	388.23	11-13
T1	13.5	617.32	617.34	617.34	1-5
T12		706.32	706.33	707.32	109-114
T10	15.6	487.29	487.31	487.30	95-98
T4	16.8	374.21	374.22	374.22	14-16
T9-10		615.39	615.40	615.40	94-98
T6	17.7	605.31	605.31	606.28	59-62
T7		2005.33	2005.63	2006.07	63-79
T2	18.5	710.31	709.33	710.31	6-10
T6-7	19.2	2591.24	2591.08	2594.80	59-79
T13	21.4	1090.58	1090.55	1091.50	115-122
T11	23.4	1199.68	1199.64	1199.65	99-108
T13-14	24.2	1204.57	1204.57	1204.58	115-123
T8	24.9-25.1	1699.28	1699.75	1699.73	80-93
T11-12	25.6	1892.90	1892.47	1890.19	99-114
T5	27.6	4713.23	4713.55	4714.07	17-58

[a] Determined by mass spectrometry
[b] Calculated ion mass values, obtained by applying cleavage rules to the entries in a protein sequence database

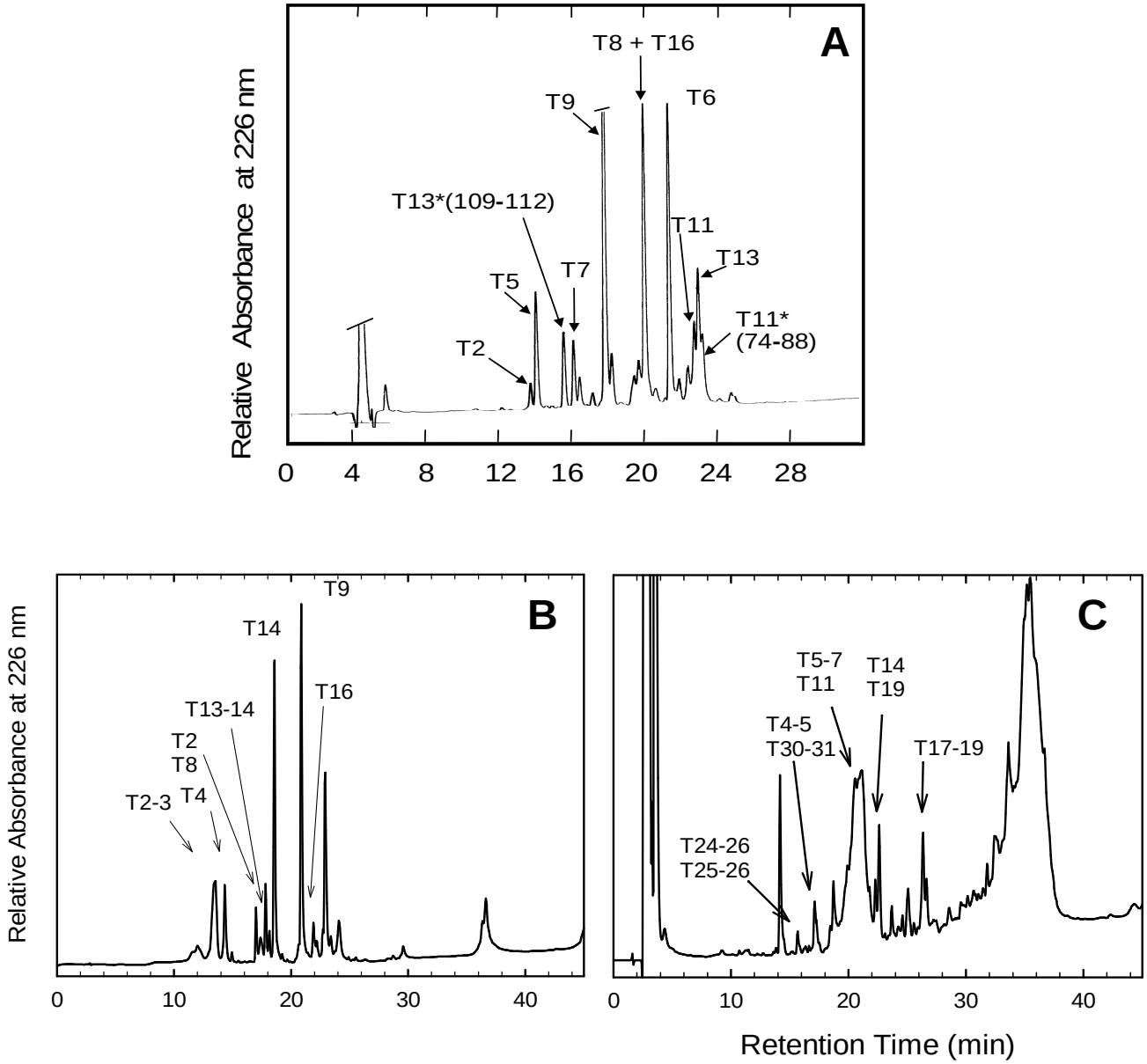


Fig. 31. RP-HPLC chromatography of standard Lys **(A)**, OM **(B)** and OA (**C**) after tryptic digestion. The column Jupiter C18 has been eluted with water and acetonitrile, with 0.1% (TFA), with an acetonitrile gradient from 5 to 34% in 15 min, from 34 to 60% in 23 min, at a flow rate of 0.6 mL/min.

Table 7. LC-MS/MS molecular weight determination of the peptides derived from the tryptic digestion of Lys in Fig. 31A.

Fragment	found[a] MW (Da) standard protein	theoretical[b] MW (Da)	Sequence
T2	477.27	477.32	2-5
T5	873.41	873.48	15-21
T6	1324.62	1324.75	22-33
T7	1427.64	1427.77	34-45
T8	1752.83	1752.99	46-61
T9	992.39	992.50	62-68
T11	2507.18	2507.15	74-88
T13	1674.79	1674.97	98-112
T16	1044.54	1044.63	117-125

Table 8. LC-MS/MS molecular weight determination of the peptides derived from the tryptic digestion of OM in Fig. 31B.

Fragment	RT[a] (min)	found[a] MW (Da) standard protein	theoretical[b] MW (Da)	Sequence
T2-3	13.5	1105.61	1105.54	8-17
T4	14.3	846.43	846.48	18-24
T2	16.9	791.46	791.38	8-14
T8	16.9	890.53	890.45	83-89
T13-14	18.1	3512.93	3512.86	129-159
T14	18.5	3384.54	3382.48	130-159
T4	20.9	846.74	846.43	18-24
T9	20.9	2641.53	2640.14	90-112
T16	21.9	2355.41	2354.08	165-185

Table 9. LC-MS/MS molecular weight determination of the peptides derived from the tryptic digestion of OA in Fig. 31C.

Fragment	RT[a] (min)	found[a] MW (Da) standard protein	theoretical[b] MW (Da)	Sequence
T24-26	17.4	1015.62	1015.65	277-284
T25-26	17.4	887.56	887.56	278-284
T4-5	18.7	1058.54	1061.55	47-55
T30-31	18.7	3768.86	3765.03	323-359
T11	21.0	1521.43	1521.79	111-122
T5-7	21.0	1344.71	1346.77	51-61
T14	22.5	1858.97	1857.96	143-158
T19	22.5	2283.82	2283.14	200-218
T17-19	26.3	3821.75	3822.4	187-218

[a] determined by mass spectrometry

[b] calculated ion mass values, obtained by applying cleavage rules to the entries in a primary sequence database

3. Identification of protein allergens in a complex matrix

The majority of foods are subjected to some type of processing either at home or by the manufacturer. Food allergenic potency could be altered by several procedures, such as mechanical, thermal, biochemical and chemical processes. A suitable method for the identification of allergenic proteins in food is the combined approach of SDS-PAGE and mass spectrometry analysis, with the double goal to find hidden or low represented proteins as well as to identify their chemical and physical-chemical modifications induced by manufacturing procedures. Moreover, the level of sensitivity of this technique in the identification of allergens is very high, also in a very complex matrix as a cooked food (Monaci & Hengel, 2007).

We have analyzed the protein extract of a hand-made cake, prepared by mixing eggs, milk, flour, yeast, sugar and butter and then cooked in the oven at 180°C for 45 min. The protein extraction from the cake was performed as reported in Materials and Methods. The protein extract, corresponding to 5.3 mg of cooked cake, was directly applied on a SDS-PAGE (Fig. 32). The resulted bands were cut and submitted to fingerprint digestion, then the peptides identified by LC-MS/MS mass spectrometry, after tryptic digestion of the protein material (Table 11).

Allergenic proteins from milk, egg white and other ingredients present in the cake were resolved by SDS-PAGE and identified by LC/MS-MS analysis, both in milk and egg white protein extracts and in the complex mix. The milk proteins found in the cake extract were α_{s1}-CN, α_{s2}-CN, β-CN and κ-CN, the MS/MS sequencing determined also two phosphorylation sites in the α_{s1}-CN sequence. The allergen proteins from egg white identified in the complex matrix were OA and Lys. While OA was found in the gel band correspondent to its MW, Lys was identified at double MW, as it was aggregated. The protein aggregation could be a consequence of the high temperature treatment in the cooking. Serpin is a flour protein identified in the protein extract. Others milk proteins as α-LA and β-LG were identified only in the milk extract, such as OM and OT from egg white. These proteins could be denaturized, hydrolyzed and aggregated by the physical-chemical process in the cake preparation.

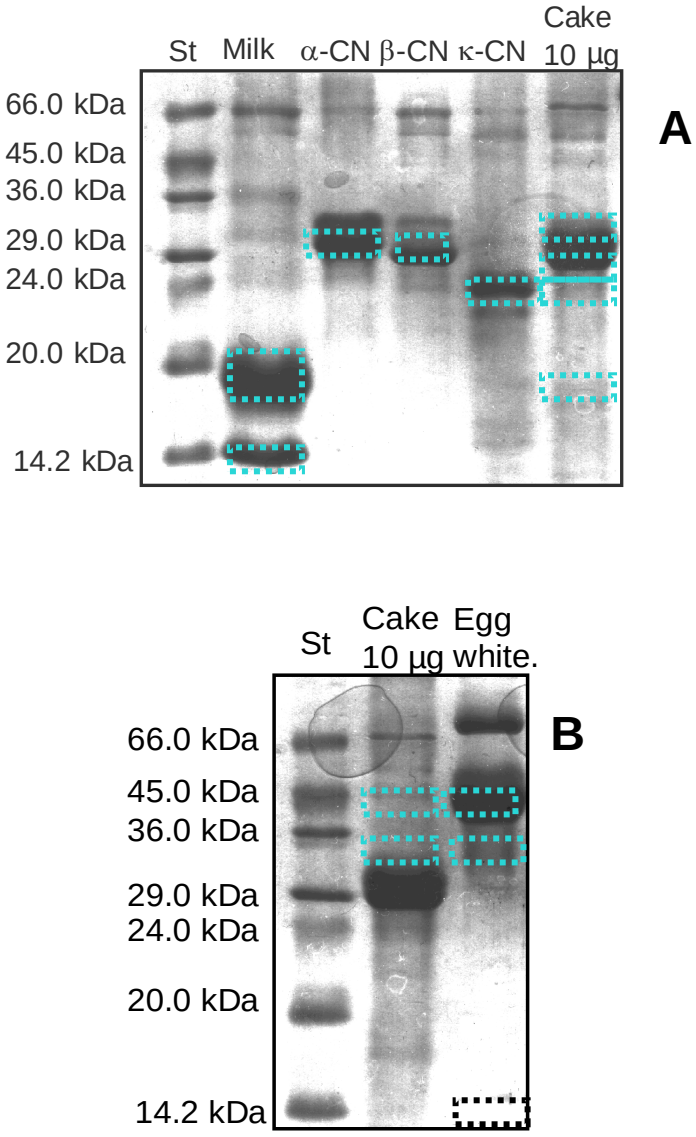


Fig. 32. SDS-PAGE analysis of milk proteins **(A)**, and egg white protein extract **(B)**. In Cake, 10 µg of the cooked cake were loaded. In St standard proteins of known MW were loaded. The spots analyzed by LC-MS/MS are highlighted with a broken line.

Table 10. Results of the tryptic peptides analysis, from milk, egg white and complex cooked food bands, highlighted in Fig. 32. The bands were analyzed by ESI LC-MS/MS mass spectrometry.

Sample	Protein	MW theoretical (kDa)[a]	MW SDS-PAGE (kDa)	Seq. Cov. (%) LC-MS/MS	Peptides LC-MS/MS
Milk	α-LA	14.2	14.2	16	101-112 118-127
	β-LG	18.6	18	46	1-8, 41-60 78-83, 92-100 125-135, 142-162
	α_{s1}-CN	24.5	30	17	23-37, 38-49, 52-57 52-57, 106-115 121-133 Phosph: Ser 56
	α_{s2}-CN	26.1	30	8	96-106, 189-196
	β-CN	25.1	29	7	123-128, 129-136
	κ-CN	21.3	24	5	46-55
Cooked cake	α_{s1}-CN	24.5	32	23	23-37, 38-49, 106-115 121-134 Phosph Ser-130
		24.5	34	32	23-37, 38-49, 106-115, 121-134, 148-166
	α_{s2}-CN	26.2	24	15	96-106, 130-140 153-164 Phosph Ser-158
		26.2	34	15	96-106, 130-140, 153-164 189-196 Phosph Ser-158
	β-CN	25.1	30	8	123-128, 185-191 192-198
	κ-CN	17.8	24	17	15-24 59-76

Sample	Protein	MW theoretical (kDa)[a]	MW SDS-PAGE (kDa)	Seq. Cov. (%) LC-MS/MS	Peptides LC-MS/MS
Egg white	OM	20.1	34	24	107-113 114-136 189-209
	Lys	14.7	14.2	66	6-13, 15-21 22-33, 34-45 46-61, 62-68 98-112, 117-125 112-123, 128-143
	OA	43.2	45	31	144-159, 160-182 188-200, 191-200 265-277, 324-340, 361-370
Cooked cake	Lys	14.7	34	21	52-63 116-130
	OA	43.2	45	13	112-123, 128-143 191-200, 265-277
	serpin (*Triticum aestivum*)	43.2	45	9	138-151, 261-270, 289-300

[a] Calculated ion mass values, obtained by applying cleavage rules to the entries in a protein sequence database

4. Analysis of allergenic proteins in a baked food

The aim of this part of the project was to study the chemical and physical modifications of some proteins after processes like cooking in the presence of sugars and lipids, to study the protein content with chromatographic and electrophoretic techniques and mass spectrometry analysis.

A baked food was made as described in Material and Methods in order to analyze the protein content at different times of cooking. Samples from the cake before cooking, at halfway through the cooking process and after complete cooking, were compared by different techniques, extracting the proteins at neutral and at acidic pH. To have an overall picture of the protein extract from the baked food, an SDS-PAGE electrophoresis was performed (Fig. 33). From the extraction at pH 8.2, is clearly visible the loss of proteins from the beginning of the cooking process. In the raw sample the milk proteins α-LA (14 kDa), β-LG (17 kDa), BSA (66 kDa) and caseins (25-30 kDa) were present, as well as the egg white proteins Lys (14.3 kDa), OM (29 kDa), OA (46 kDa). After cooking, only the bands of the caseins could be identified. The extraction of the proteins at pH 4.5, after caseins precipitation, underline the serum milk and egg white proteins in the row extract, while the cooked samples were composed by not well identifiable bands (Fig. 34).

The protein extracts from raw and cooked samples, extracted at neutral and acidic pH, were analyzed also by RP-HPLC chromatography (Fig. 35- 37), and the peaks content was loaded in a SDS-PAGE electrophoresis gel to check the protein identity (Fig 36-38). The chromatograms of the extraction at pH 8.2 showed many peaks for the raw material, which number was sensibly reduced after cooking. The same thing happened for the protein extract at pH 4.5, and the cooked sample had very low protein content. The two row extracts, compared by electrophoresis of the chromatographic peaks, revealed the presence of OM (30 kDa in gel) in the peak eluted around 15.0 min, Lys (14.3 kDa) at 24.8 min, OT (95 kDa) and α-LA (14 kDa) eluted at 30.0 min, β-LG (17.0 kDa) at 35.0 min and OA (46 kDa) around 41.0 min (Fig. 36). The extract at pH 8.2 presented other peaks eluted at 29.1 min and 31.8 min, identified as α and β-CN respectively. Other peaks were identified as yolk and flour proteins (data not shown). The cooked cake extracts were purified by RP-HPLC chromatography and the peaks analyzed by SDS-PAGE electrophoresis (Fig. 38). Four main peaks composed this chromatogram, and three of them disappeared with the acidic extraction. The bands of the cooked sample extract at neutral conditions were analyzed by LC-MS/MS, the results are shown in Table 13. The sample was mainly composed by α and β-CN, with traces of other proteins like Lys, as we found in the previous analysis of the complex mixture.

This analysis showed that the main part of food proteins was instable to thermal processing. As the matter of fact, egg yolk and flour proteins were not found in the cooked protein extract, only Lys was resistant to high temperature, even if aggregated. Above the milk proteins, all the serum proteins disappeared after cooking, and only caseins were resistant to high temperature. α_{s1} and β-CN were found in a monomeric form, while α_{s2}-CN was aggregated at high MW (Table 13). In SDS-PAGE electrophoresis the low sequence coverage of these proteins could be caused by post translational modifications that alter some amino acids, preventing the sequencing of the peptide. Caseins result to be the only proteins able to resist to thermal stress, probably for their lack of secondary structure or their micellae composition. The other food proteins, not found in the cooked sample, could be aggregated, precipitated, or modified by sugar or other molecules. Part of the proteins could be digested in smaller peptides by proteases present in food, or crosslink with intra or inter-molecular bonds.

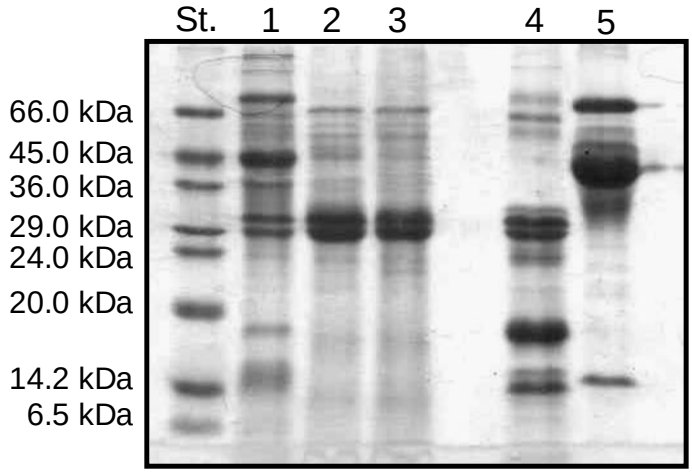


Fig. 33. SDS-PAGE analysis of the protein extracts at pH 8.2: 10 μg of raw cake (1), 10 μg of halfway through the cooking process cake (2), 10 μg of cooked cake (3), milk protein extract (4), and egg white protein extract (5). In St standard proteins of known MW were loaded.

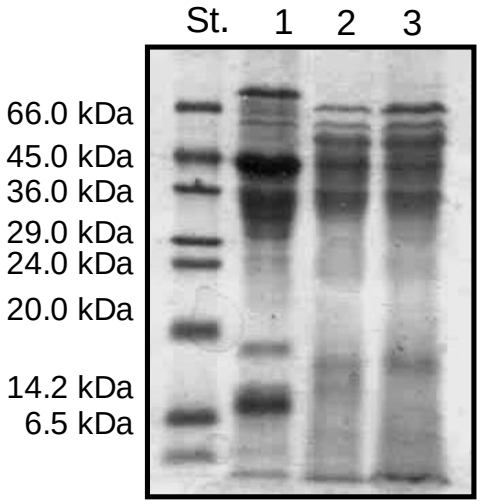


Fig. 34. SDS-PAGE analysis of the protein extracts at pH 4.5: 10 μg of raw cake (1), 10 μg of halfway through the cooking process cake (2), 10 μg of cooked cake (3). In St standard proteins of known MW were loaded.

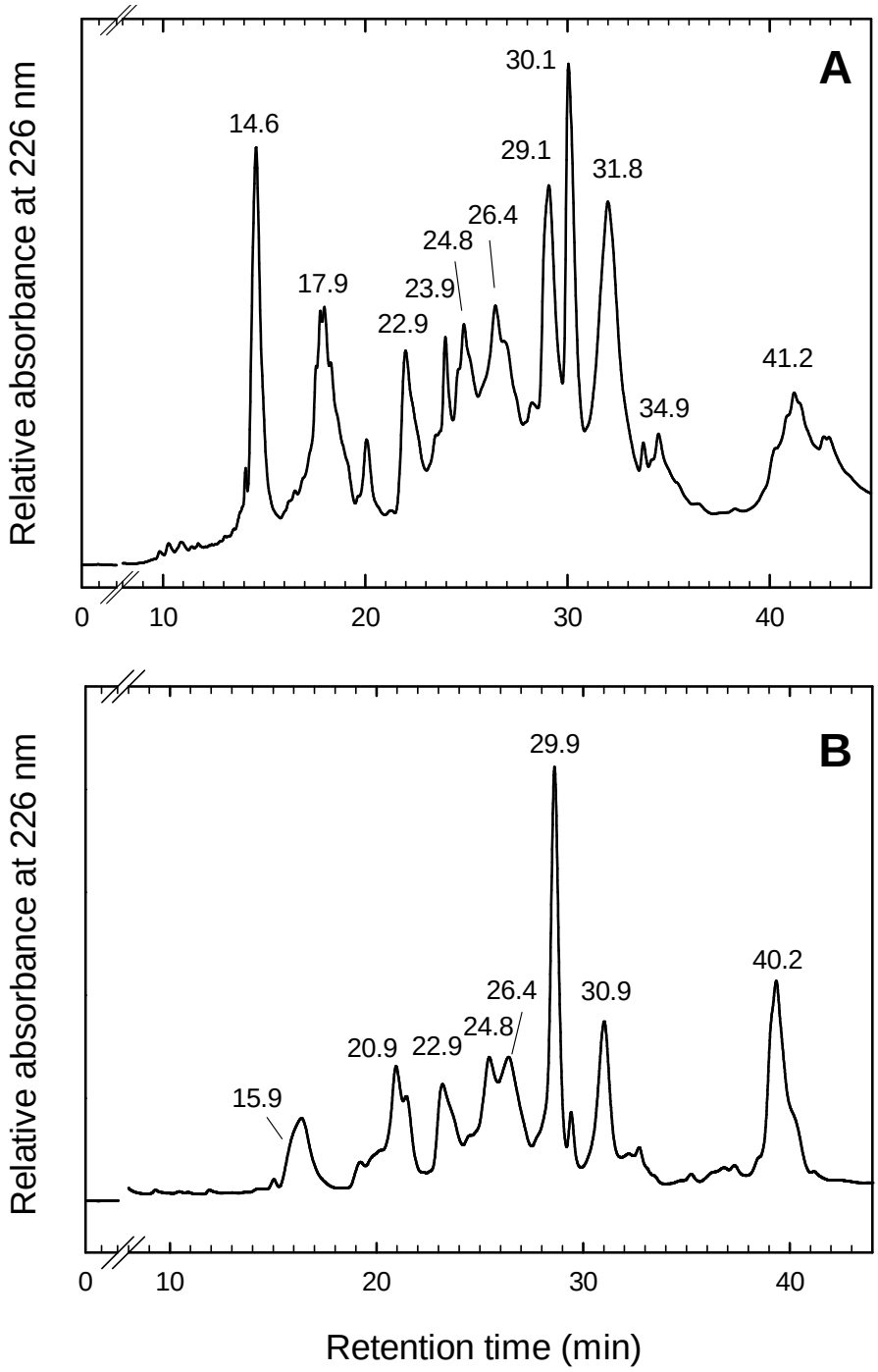


Fig. 35. RP-HPLC chromatography of the protein extract at neutral conditions from raw complex food **(A)** and cooked complex food **(B)**. The column Jupiter C18 has been eluted with water and acetonitrile, with 0.1% (TFA), with an acetonitrile gradient from 5 to 34% in 8 min, from 34 to 60% in 25 min, at a flow rate of 0.8 mL/min.

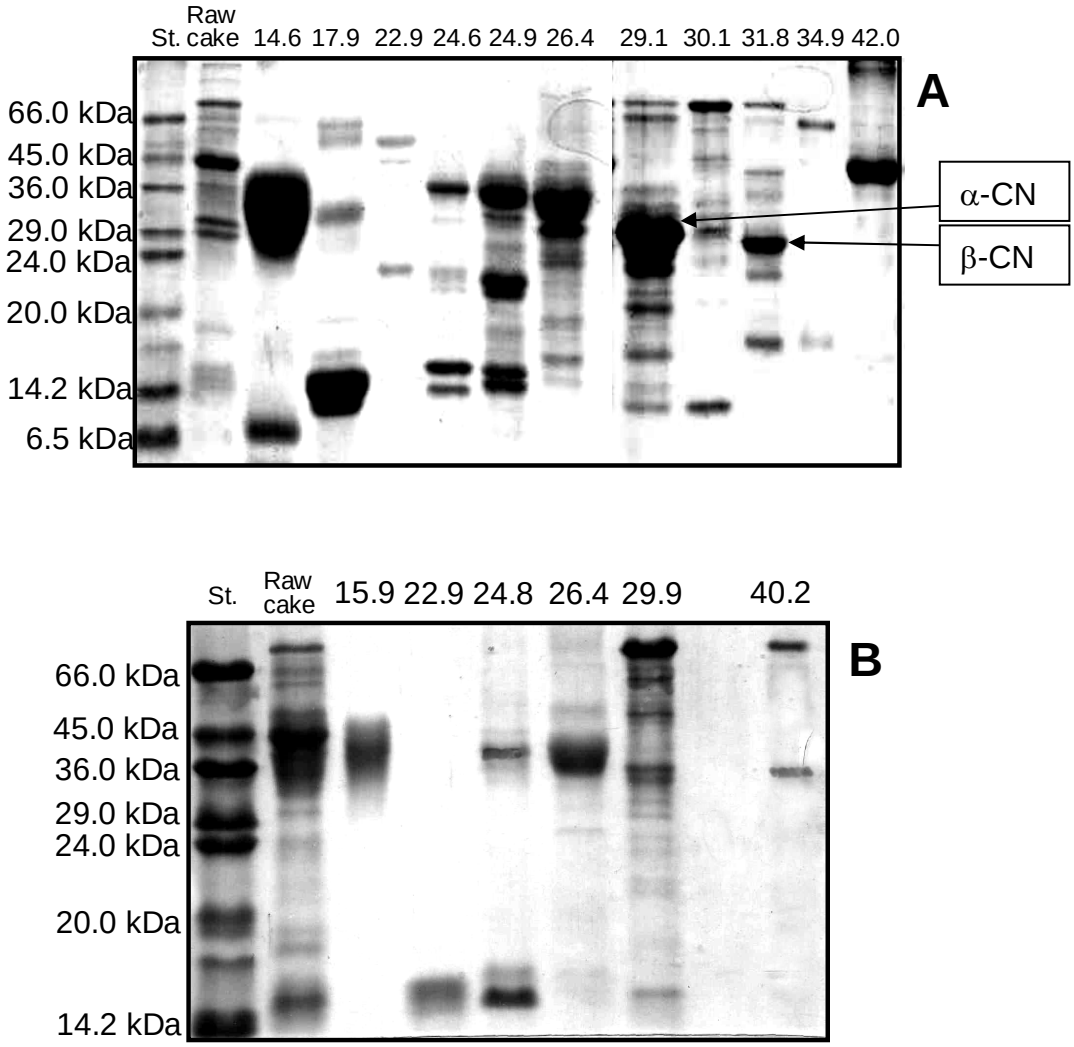


Fig. 36. SDS PAGE electrophoresis analysis of the sample corresponding to the peaks eluted from RP-HPLC in Fig. 35A (**A**) and 35B (**B**). In Row cake a sample from the total protein extract, at pH 8.2 (**A**) and pH 4.5 (**B**) were loaded. In St standard proteins of known MW were loaded.

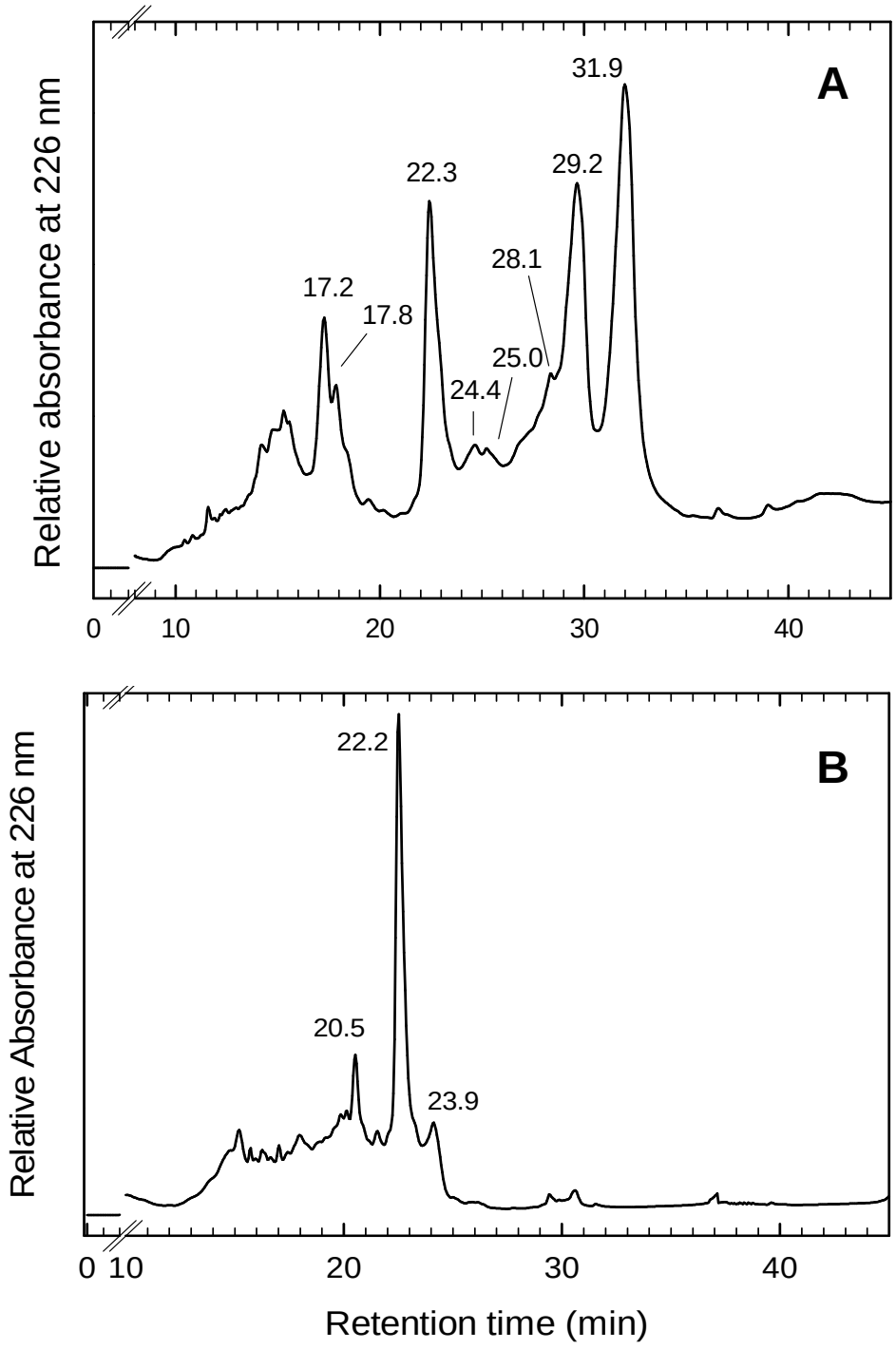


Fig. 37. RP-HPLC chromatography of the protein extract at acidic conditions from raw complex food **(A)** and cooked complex food **(B)**. The column Jupiter C18 has been eluted with water and acetonitrile, with 0.1% (TFA), with an acetonitrile gradient from 5 to 34% in 8 min, from 34 to 60% in 25 min, at a flow rate of 0.8 mL/min.

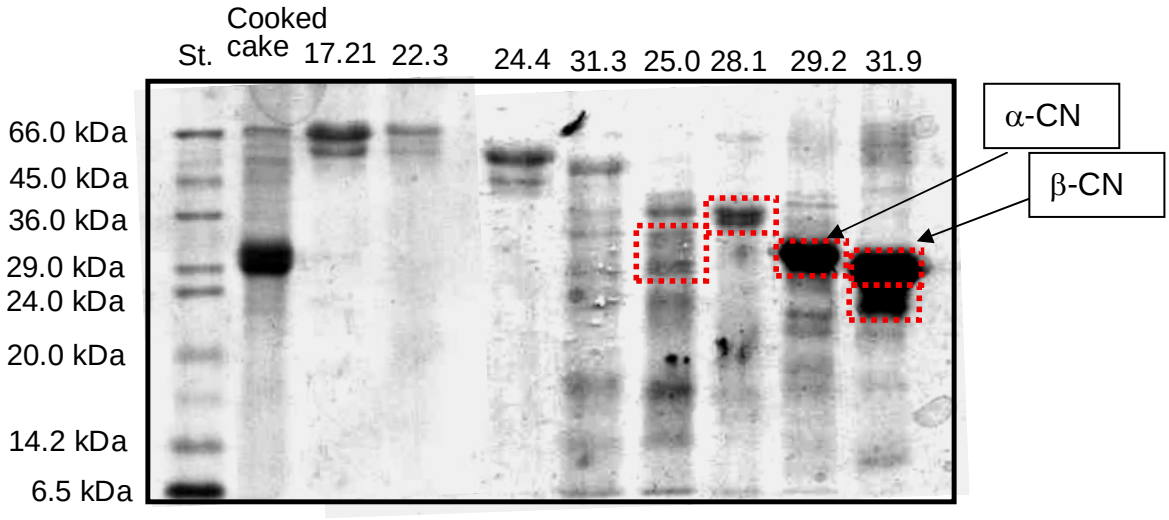


Fig. 38. SDS PAGE electrophoresis analysis of the sample corresponding to the peaks eluted from RP-HPLC in Fig. 37A **(A)**. In Cooked cake a sample from the total protein extract, at pH 8.2 were loaded. In St standard proteins of known MW were loaded. The spots analyzed by LC-MS/MS are highlighted with a broken line.

Table 11. LC-MS/MS results of the trypsin peptides analysis from cooked food bands highlighted in Fig. 38.

Protein	MW (kDa) Theoretical[a]	MW (kDa) Found SDS-PAGE	Tryptic peptide LC-MS/MS	Sequence coverage (%)
α_{s2}-CN	26.2.	31.0	4	13
α_{s1}-CN	24.5	36.0	2	15
		30.0	2	11
β-CN	25.1	28.0	3	9.5
		24.0	1	7.6

[a] Calculated ion mass values, obtained by applying cleavage rules to the entries in a primary sequence database

Maillard reactions of caseins, fluorescence and furosine analysis. Proteins exposed to sugars and lipids undergo to non-enzymatic reactions, called Maillard reactions, and these compounds, really instable, are subjected to further reactions till the formation of Amadori products (Krause *et al.*, 2004).

A qualitative analysis of the AGEs formed during the cooking process can be made by spectroscopy absorbance and fluorescence analysis, choosing specific compounds that absorb or emit light at a specific wavelength (Schmitt *et al.*, 2005; Schmitt *et al.*, 2005). These experiments were conducted on α_{s1} and β-CN, the only proteins from the cooked sample detectable in a monomeric form and that we could purify, by RP-HPLC chromatography.

Furosine (FUR) and furfural compounds can be considered indicators of the extent of Maillard reactions in relation with the type and intensity of the food processing conditions, as well as with the storage conditions (Ferrer *et al.*, 2000). FUR content gives an evaluation of the blocked and therefore not available K and is therefore considered the most specific and the earliest indicator of Maillard reactions. Furfurals are intermediary compounds in the formation of pigments (melanoidins) in the most advanced stages of Maillard reactions. We analyzed hydroxymethyl furfural (HMF), the first furfural formed, as indicator of the extent of Maillard reactions (Morales et al., 1997). FUR and HMF were measured in order to compare the values obtained from α_{s1} and β-CN extracted from milk, raw and cooked cake, and to detect possible differences in quality between the three types of products. The levels were measured in the samples as described in Materials and Methods, using regression curves with standard FUR and HMF as a reference (Table 14). It is important to notice that FUR and HMF are just intermediaries of the reaction, which can either degrade or condensate into a brown polymer. The analysis of α_{s1} and β-CN, extracted from the three samples, had different FUR levels. The modified K was not present in both the caseins extracted from milk, suggesting that after pasteurization these proteins were not modified. The FUR concentrations increased in the proteins extracted from the raw cake, more in α_{s1} than in β-CN. After cooking, the molecule level in α_{s1} still increased, while it was stable in the β-CN sample. HMF was not present in milk extracts, as expected, but it was found at a quite high concentration in the raw α_{s1} sample; this concentration decreased after cooking. β-CN, as well, present a significant level of HMF in the raw extract, which decreased after the high temperature process. According to these results we can assert that α_{s1}-CN, not modified in milk, reacts in a complex food in presence of other molecules like sugar and lipids, forming not only early stage glycations but also more advanced products. After cooking, some K still react to form FUR, while HMF, which has a shorter life, goes on with further reactions. β-CN

data showed a low but constant FUR level in raw and cooked sample, while the HMF concentration, lower than α_{s1}-CN, increased after cooking. The number of K in the sequence, 14 in α_{s1}-CN and 11 in β-CN can be the reason why the level of modifications are higher in α_{s1}-CN than in β-CN. Some K of β-CN could also be busy in other bindings with other molecules, as well as with other proteins or lipids and therefore they are not free to bind dicarbonyl compounds to form FUR or HMF.

Some products of Maillard reactions are fluorescent molecules, like 5-methylimidazolinone, derived from the modification of K with MG. Consequently, it is possible to measure the degree of a modification in a protein by observing the level of fluorescence, which increases at specific wavelength. The excitation of the sample at 330 nm, registering the emission maxima at 420-430 nm, is specific to measure the level of argpyrimidine and pentosidine, the AGEs compounds derived from the amino acid R, as described in literature. The protein intrinsic fluorescence, measured by exciting at 280 nm an registering the emission maxima at 320-350 nm, is subject to a strong decrease with the increase of AGEs, due to the conformational changes of the protein structure and the quenching, a process which decreases the intensity of the fluorescence emission, between the amino acids chromophores and the new cyclic molecules produced by glycation. The sample subject to Maillard reactions, excited at 280 nm, showed a progressive decreasing of fluorescent intensity at 350 nm and an increasing of fluorescent intensity at 420-430 nm.

First of all, we analyzed the standard proteins α and β-CN, incubated with MG in a concentration of 1 mM at 37°C.

Caseins are proteins without a proper tertiary structure, which in physiological conditions form micellae, so the spectra of standard proteins, taken as t=0, were slightly different from the usual protein spectra. Analyzing the spectra, exciting at 330 nm, it is clear that the proteins, incubated only 6 hours with 1 mM MG, had a quick increment in the fluorescent emission at 430 nm, due to AGEs formation (Fig. 39). The spectra relative to the intrinsic fluorescence of α and β-CN showed a rapid decrease of the fluorescence emission at 340-350 nm, as a complete denaturation process or a strong quenching phenomenon in few hours; this process was quicker for β-CN. In parallel with the peak decrease at 350 nm, the spectra of the two proteins showed an increase of fluorescence emission at 430 nm, the specific emission wavelength of AGEs molecules. These data proved the quick and specific reactions between MG and α-CN o̩r β-CN standard. These information could be only interpreted as qualitative data only, due to the heterogeneity of the molecules that contribute

to the fluorescent emission; nevertheless, these data are important because they highlight the reactions occurred in the proteins.

α-CN and β-CN were purified from milk, raw and cooked cake protein extracts, and analyzed by fluorescence spectroscopy. The spectra collected were qualitative indicators of the state of these proteins, after many processes like milk pasteurization, mixing and stirring with others foods, and after cooking at high temperature. The spectra of excitation at 330 nm of α-CN revealed a constant increase of the fluorescence emission at 420-430 nm, signaling that the content of fluorescent AGEs grew with the mixing with other molecules and the high temperature treatment (Fig. 40A). On the contrary, the β-CN spectra showed a different behavior of this protein, which had a prominent emission of fluorescence at 430 nm already after extraction from milk (Fig. 40B). β-CN from raw and cooked food had approximately the same fluorescent intensity, a little lower for the cooked protein. These data can be explained either by an early glycation process of the protein during the milk pasteurization, with the formation of important fluorescent molecules, or by the presence of a fluorescent compound linked to β-CN, as retinoic acid (Peterson & Rask, 1971). Further experiments will be done to prove one of these hypotheses. The excitation of α-CN at 280 nm revealed the loss of the intrinsic protein fluorescence emission at 350 nm, both in the raw and in the cooked extracts (Fig. 40C). α_{s1}-CN extracted from milk had the same structure of a standard protein, while in the other two samples the loss of fluorescence at 350 nm was balanced by an increase of fluorescent intensity at 430 nm, the proper wavelength of the fluorescent AGEs. The spectra of β-CN, with excitation wavelength at 280 nm, showed the loss of protein structure already in the sample extracted from milk, with a concurrent fluorescence peak at 430 nm (Fig. 40D). β-CN from raw and cooked cake lost completely the fluorescence peak at 350 nm, showing a large peak at 420-440 nm, higher for the raw sample than the cooked one. Again, the possible explanation of these results could be the presence of another molecule, like retinoic acid, linked to the protein, which was released at high temperature, living the fluorescence intensity depending only from the AGEs compounds (Andersen *et al.*, 2005).

Table 12. FUR and HMF levels in α_{s1}-CN and β-CN extracted from milk, raw and cooked cake. The FUR and HMF contents were measured by absorbance measure at specific wavelength.

Sample (100 μg)	FUROSINE (μg) Detection at 280 nm	HMF(μg) Detection at 284 nm
α_{s1}-CN milk	-	-
α_{s1}-CN raw cake	0.015	0.126
α_{s1}-CN cooked cake	0.025	0.072
β-CN milk	-	-
β-CN raw cake	0.004	0.05
β-CN cooked cake	0.004	0.007

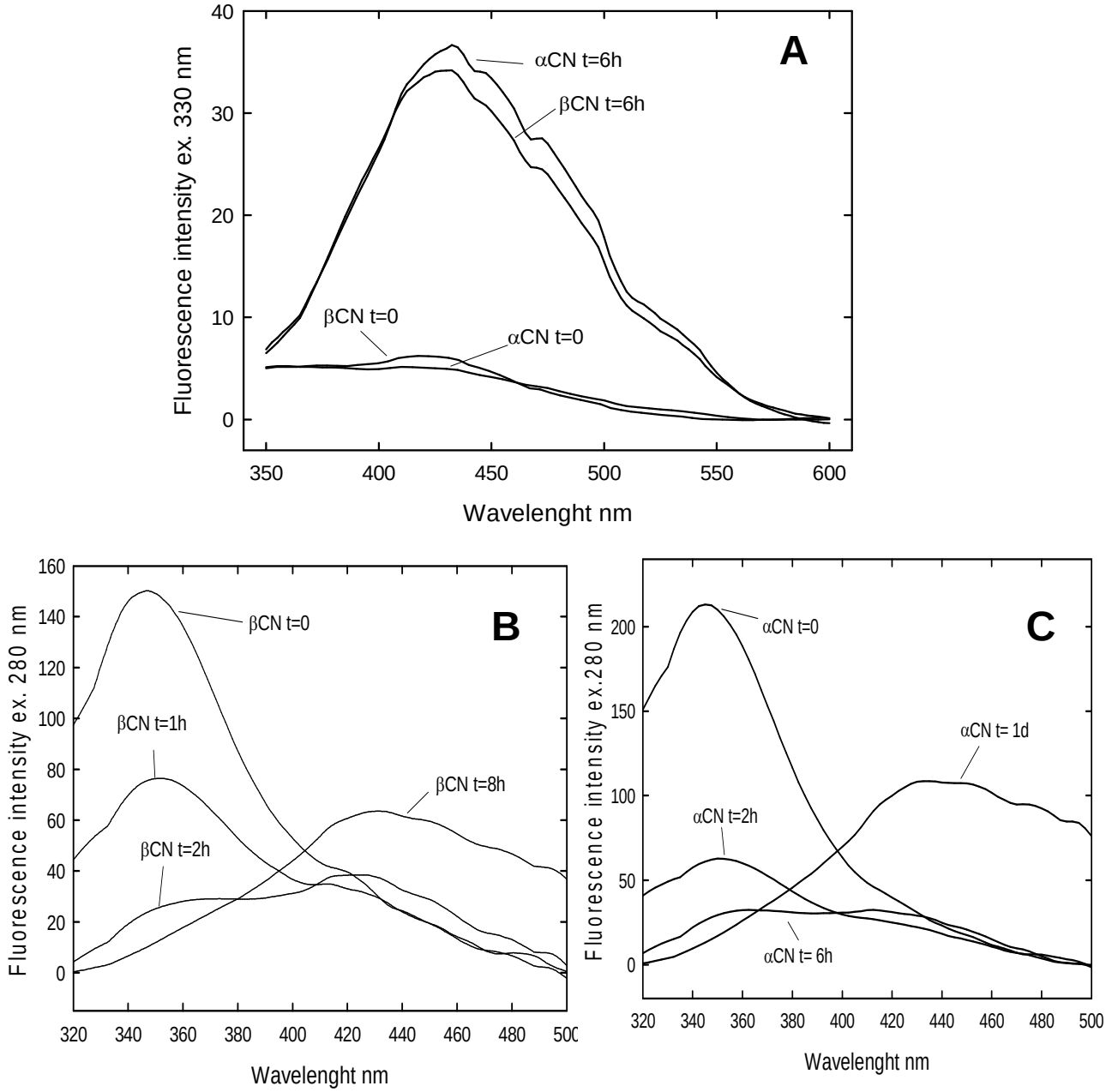


Fig. 39. Fluorescence spectra of α_{s1}-CN and β-CN standard incubated with MG 1 mM. α_{s1}-CN and β-CN were excited at 330 nm, the emission spectra was registered between 350 and 600 nm **(A)**. α_{s1}-CN was excited at 280 nm, the emission spectra was registered between 320 and 500 nm **(B)**. β-CN was excited at 280 nm, the emission spectra was registered between 320 and 500 nm **(C)**. The samples were in 20 mM Na_2HPO_4, pH 7.5.

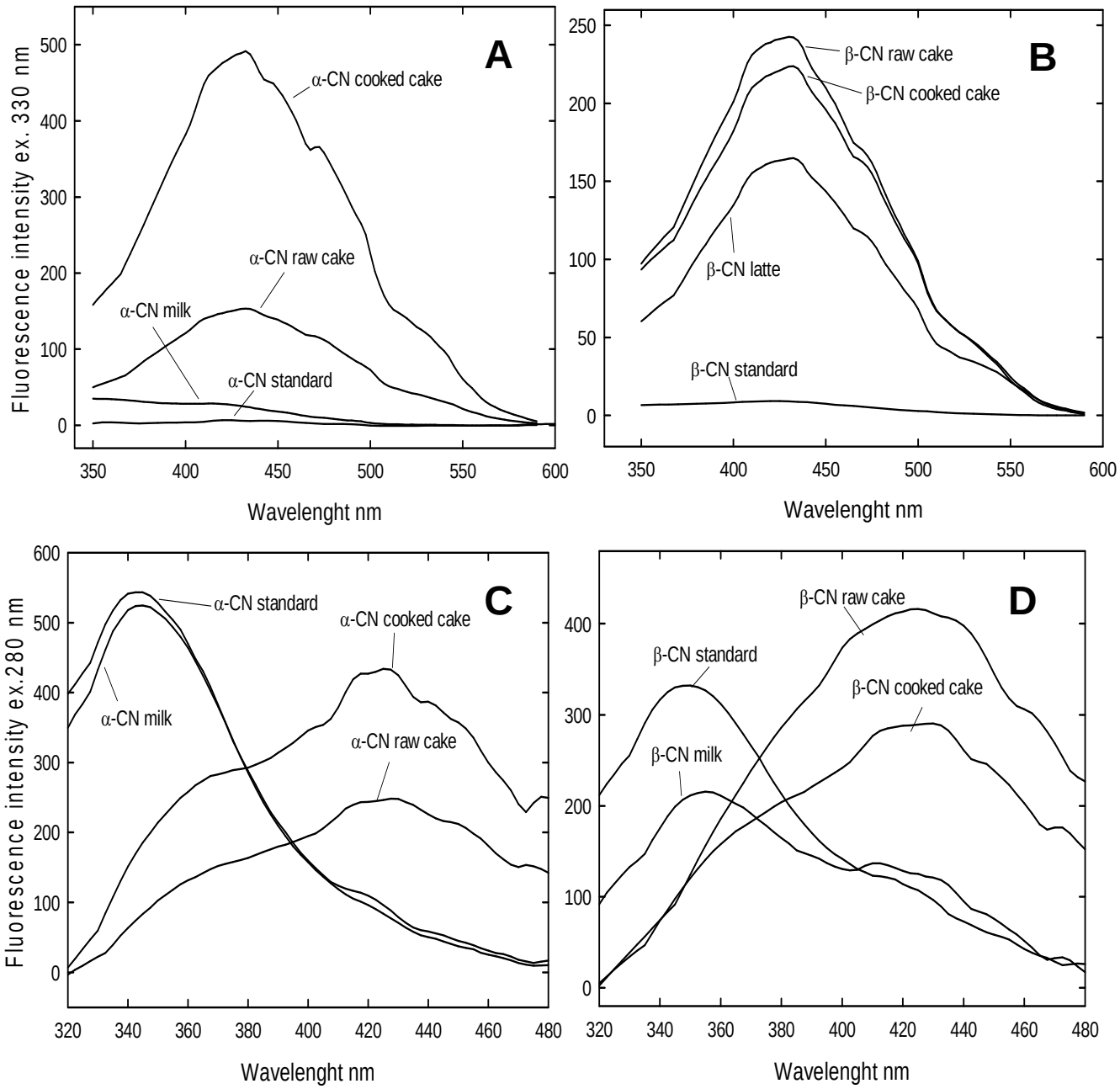


Fig. 40. Fluorescence spectra of α_{s1}-CN and β-CN standard and extracted from milk, raw coke and cooked cake. α_{s1}-CN was excited at 330 nm, the emission spectra was registered between 350 and 600 nm **(A)**. β-CN was excited at 330 nm, the emission spectra was registered between 350 and 600 nm **(B)**. α_{s1}-CN was excited at 280 nm, the emission spectra was registered between 320 and 500 nm **(C)**. β-CN was excited at 280 nm, the emission spectra was registered between 320 and 500 nm **(D)**. The samples were in 20 mM Na_2HPO_4, pH 7.5.

Fingerprinting and LC-MS/MS analysis. α_{s1}-CN and β-CN, extracted from milk, raw cake and cooked cake were analyzed by LC-MS/MS mass spectrometry in order to detect important AGEs products due to the cooking process. The proteins were purified from the samples by RP-HPLC chromatography, afterward, they were isolated from all the possible coeluting proteins by SDS-PAGE. The trypsin digestion was performed on the bands correspondent to the caseins, then peptides were analyzed by a ESI Q-Tof Premiere mass spectrometer, similar to the normal ESI Q-Tof, but with a greater sensibility and accuracy. The proteins were identified with high sequence coverage, in all the samples. Some post translational modifications, as the α-CN phosphorylation, were identified, both in the protein extracted from milk and from the complex food (Fig. 41). Also some oxidations were identified, which are typical reactions between the amino acid M and the oxygen in the air, frequent when the protein is exposed to the atmosphere (Fig. 42). The goal of this analysis was the detection and the sequencing of an α-CN peptide with a non enzymatic modification (Fig. 43). This peptide, from α-CN from cooked food, was composed by two peptides usually cleaved by trypsin. The K amino group in the middle of the sequence reacted with a dicarbonyl compound, to produce the AGEs N^{ε}(carboxymethyl) lysine, a well known Maillard modification formed upon the reaction between K and glyoxal, generated during the glycation reaction or the lipid peroxidation. This α-CN modification could not be seen with fluorescence spectroscopy, because it is not fluorescent, only by mass spectrometry we could establish its presence. The instable nature of AGEs does not permit a simple identification, usually the modified proteins aggregate in high MW polymers, but the analysis of caseins in the monomeric form was important to check the real damage and amino acid loss in food proteins and especially in a complex food, subject to various chemical and physical stresses.

Table 13. LC-MS/MS results of the tryptic peptides derived from α_{s1}-CN and β-CN extracted from milk, raw cake and cooked cake. The proteins were purified by RP-HPLC chromatography and SDS-PAGE electrophoresis.

Protein	MW (kDa) Theoretical[a]	MW (kDa) Found	Tryptic peptide LC-MS/MS	Sequence coverage (%) Post-translational modifications
α_{s1}-CN from milk	24.5	24.513	9	52 Deamidation Q108, oxidation M135, phosphorylation Ser-115
α_{s1}-CN from raw food	24.5	24.513	12	55 Deamidation Q108, oxidation M54,135, phosphorylation Ser-130
α_{s1}-CN from cooked food	24.5	24.513	8	49 Deamidation Q108
β-CN from milk	24.5	25.200	4	24.4 Deamidation Q194, oxidation M185
β-CN from raw food	25.1	23.608	4	32 Deamidation Q194, oxidation M185
β-CN from cooked food	25.1	25.200	3	20 Deamidation Q194, oxidation M185

[a] Calculated ion mass values obtained by applying cleavage rules to the entries in a protein sequence database.

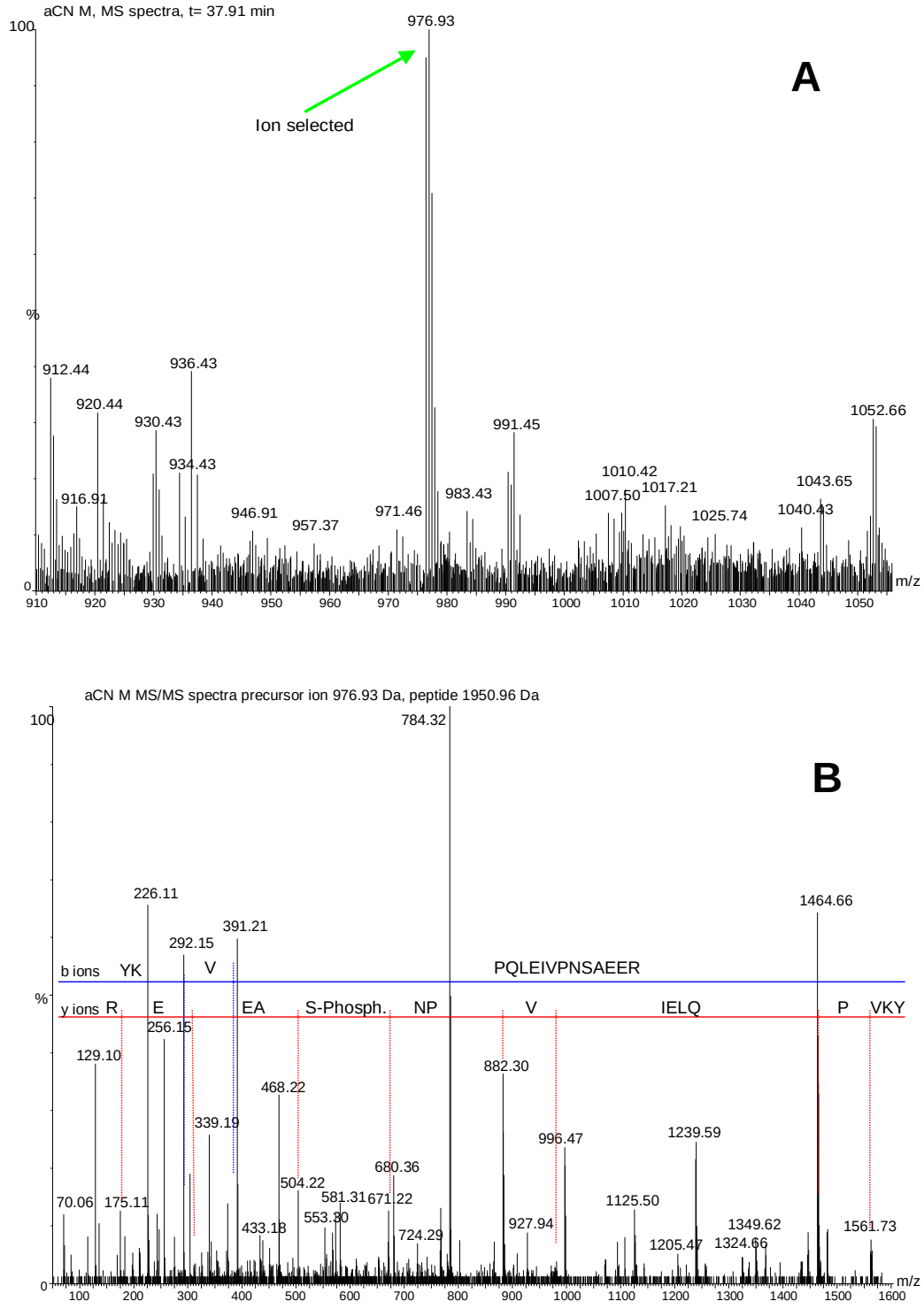


Fig. 41. LC MS/MS analysis of the peptide 1950.96 Da, from α_{s1}-CN milk sample. Mass spectra of the peptide eluted in the nano LC column during the experiment **(A)**. MS/MS spectrum of the double charged ion at m/z 976.93 and the interpreted amino acid sequence, identified as peptide 104-119 of α_{s1}-CN, with S115 phosphorylation **(B)**.

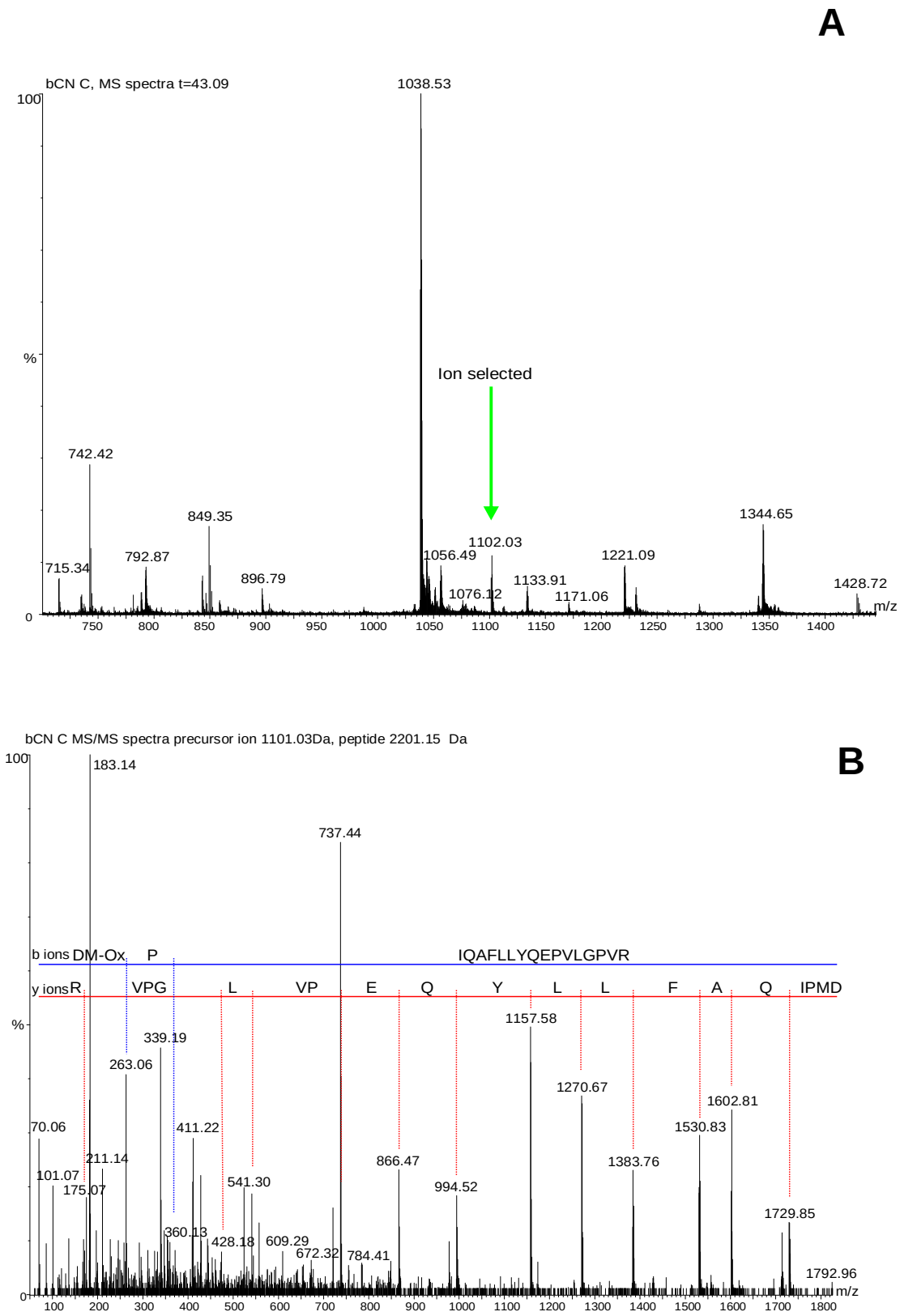


Fig. 42. LC MS/MS analysis of the peptide 2201.15 Da, from β-CN from cooked cake sample. Mass spectra of the peptide eluted in the nano LC column during the experiment **(A)**. MS/MS spectrum of the double charged ion at m/z 1102.03 and the interpreted amino acid sequence, identified as peptide 184-202 of β-CN, with M 185 oxidation **(B)**.

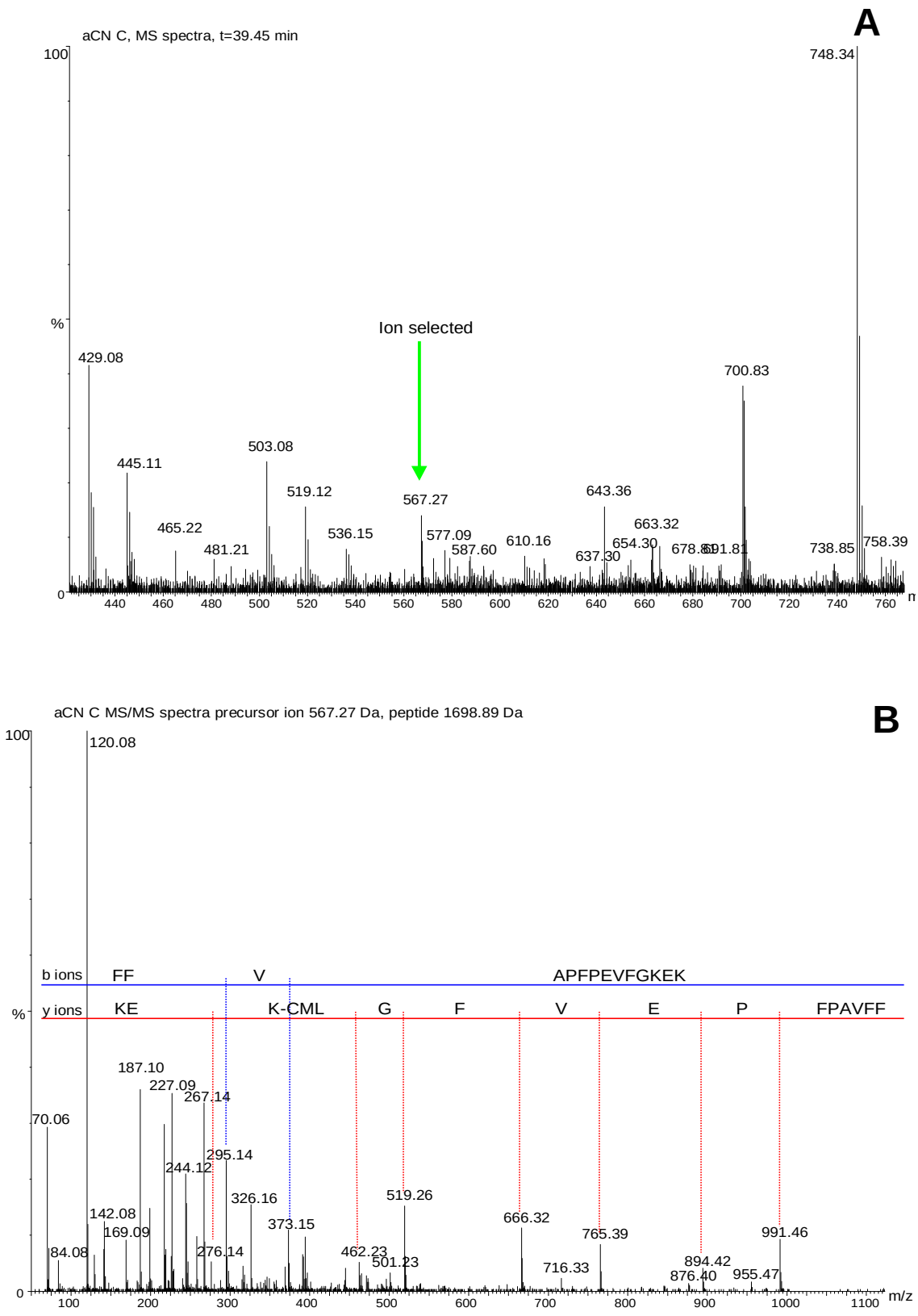


Fig. 43. LC MS/MS analysis of the peptide 1698.89 Da, from α_{s1}-CN from cooked cake sample. Mass spectra of the peptide eluted in the nano LC column during the experiment **(A)**. MS/MS spectrum of the double charged ion at m/z 567.27 and the interpreted amino acid sequence, identified as peptide 23-36 of α_{s1}-CN, with K34 CML **(B)**

IV. CONCLUSIONS

In the present work, we have developed methods to detect and analyze allergenic proteins in foods by using modern techniques of mass spectrometry (MS). Firstly, we prepared protein samples from row food for a direct analysis by MS, and we optimized the MS methods with known standard allergenic proteins. We have clearly demonstrated the possibility to detect protein allergens in a whole protein extract, like peanuts and soybean, using the fingerprint technique, sequencing by tandem MS and database searches. The allergens Ara h1 and Gly m bd 28K were identified by MS, from peanut and soybean extracts respectively; then we have detected and characterized allergen proteins from a complex matrix, made from different ingredients, to test our MS method in a more complex sample.

As a matter of facts, food proteins and especially protein allergens are often modified, making the detection of hidden allergens difficult in the traditional immunochemical analysis. Moreover, thermal processes can have a dramatic effect on proteins, changing the structure and the allergic characteristics of allergens. The MS method that we developed proved to be really useful to analyze complex foods after thermal processing, as food is used to being prepared. In particular, we analyzed the proteins contained in a home-made baked food and thus being composed of a variety of potential allergenic proteins of different origin. First of all, we identified proteins which are resistant to the heating process, characterizing the level of protein modifications, and the allergens still present in the food; then, by MS, we were able to identify the main chemical and physicochemical protein modifications caused by cooking. The analysis by MS could not only detect a protein, but also establish the chemical modifications that could alter the allergenicity of the protein.

Finally, we demonstrated that MS techniques can be applied for identifying allergenic proteins contained in complex matrices of foods. The advantages of these techniques, compared to the immunochemical methods, are the possibility to identify proteins even after their damaging in their chemical or physical structure, and identify the specific modification of proteins that could increase or decrease their allergenic power.

REFERENCES

Ahmed N., Mirshekar-Syahkal B., Kennish L., Karachalias N., Babaei-Jadidi R. and Thornalley P.J. (2005). Assay of advanced glycation endproducts in selected beverages and food by liquid chromatography with tandem mass spectrometric detection. *Mol. Nutr. Food Res*. **49**, 691–699.

Ahmed N. and Thornalley P.J. (2002). Chromatographic assay of glycation adducts in human serum albumin glycated *in vitro* by derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl-carbamate and intrinsic fluorescence. *Biochem. J.* **364**, 15–24.

Andersen C.M., Vishart M. and Holm V.K. (2005). Application of fluorescence spectroscopy in the evaluation of light-induced oxidation in cheese. *J. Agric. Food Chem*. **53**, 9985–9992.

Ankrah N.A. and Appiah-Opong R. (1999). Toxicity of low levels of methylglyoxal: depletion of blood glutathione and adverse effect on glucose tolerance in mice. *Toxicol. Lett*. **109**, 61–67.

Baisier, W.M. and Labuza, T.P. (1992). Maillard browning kinetics in a liquid model system. *J. Agric. Food Chem.*, **40**, 707–713.

Bando N., Tsuji H., Hiemori M., Yoshizumi K., Yamanishi R., Kimoto M. and Ogawa T. (1998). Quantitative analysis of Gly m Bd 28K in soybean products by a sandwich enzyme-linked immunosorbent assay. *J. Nutr. Sci. Vitaminol.* **44**, 655-664.

Besler M. and Steinhart H., Paschke A. (2001). Stability of food allergens and allergenicity of processed foods. *J. Chromatogr. B Biomed. Sci. Appl*. **756**, 207–228.

Biemann K. (1992). Mass spectrometry of peptides and proteins. *Annu. Rev. Biochem.* **61**, 977–1010.

Birlouez-Aragon I., Leclere J., Quedraogo C.L., Birlouez E. and Grongnet J.F. (2001). The FAST method, a rapid approach of the nutritional quality of heat-treated foods. *Nahrung*. **45**, 201-205.

Bischoff S.C., Mayer J.H. and Manns M.P. (2000). Allergy and the gut. *Int. Arch. Allergy Immunol*. **121**,270–283.

Boccagni P., Favari F., Zanoni G., Pezzini A. and Tridente G. (1994). Comparison of four in vitro assays for specific IgE detection. *Int. J. Clin. Lab. Res*. **24**, 102–105.

Bradford M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of protein-dye binding. *Annal. Biochem.* **72**, 248–254.

Breiteneder H. and Radauer C. (2004). A classification of plant food allergens. *J. Allergy. Clin. Immunol*. **113**, 821-830.

Breiteneder H. and Mills E.N. (2005). Molecular properties of food allergens. *J. Allergy Clin. Immunol*. **115**, 14–23.

Bucala R., Mitchell R., Arnold K., Innerarity T., Vlassara H. and Cerami A. (1995). Identification of the major site of apolipoprotein B modification by advanced glycosylation end products blocking uptake by the low density lipoprotein receptor. *J. Biol. Chem.* **270**, 10828–10832.

Burks, A.W., Cockvell, G., Connaughton, C., and Helm and R.M. (1994). Epitope specificity and immunoaffinity purification of the major peanut allergen Ara h1. *J. Allergy Clin. Immunol*. **93**, 743–750.

Businco L., Bruno G. and Giampietro P.G. (1999). Prevention and management of food allergy. *Acta Paediatr. Suppl.* **88,** 104–109.

Cordle C.T. (2004). Soy protein allergy: incidence and relative severity. *J. Nutr*. **134**, 1213–1219.

Coussons P.J., Jacoby J., McKay A., Kelly S.M., Price N.C. and Hunt J.V. (1997). Glucose modification of human serum albumin: a structural study. *Free Radic. Biol. Med*. **22**, 1217–1227.

Czerwenka C., Maier I., Pittner F. and Lindner W. (2006). Investigation of the lactosylation of whey proteins by liquid chromatography-mass spectrometry. *J. Agric. Food Chem*. **54**, 8874–8882.

David T.J. (2000). Adverse reactions and intolerance to foods. *Br. Med. Bull*. **56**, 34–50.

Davis P.J., Smales C.M. and James D.C. (2001). How can thermal processing modify the antigenicity of proteins? *Allergy*. **56**, 56–60.

Docena G., Rozenfeld P., Fernández R. and Fossati C.A. (2002). Evaluation of the residual antigenicity and allergenicity of cow's milk substitutes by in vitro tests. *Allergy*. **57**, 83–91.

Docena G.H., Fernandez R., Chirdo F.G. and Fossati C.A. (1996). Identification of casein as the major allergenic and antigenic protein of cow's milk. *Allergy*. **51**, 412–416.

Ehn B.M., Allmere T., Telemo E., Bengtsson U. and Ekstrand B.(2005). Modification of IgE binding to beta-lactoglobulin by fermentation and proteolysis of cow's milk. *J. Agric. Food Chem*. **53**, 3743–3748.

Elsayed S., Hill D.J. and Do T.V. (2004). Evaluation of the allergenicity and antigenicity of bovine-milk alphas1-casein using extensively purified synthetic peptides. *Scand. J. Immunol*. **60**, 486–493.

Erbersdobler H.F. and Somoza V. (2007). Forty years of furosine - forty years of using Maillard reaction products as indicators of the nutritional quality of foods. *Mol. Nutr. Food Res*. **51**, 423–430.

Farrell H.M. Jr., Qi P.X., Wickham E.D. and Unruh J.J. (2002). Secondary structural studies of bovine caseins: structure and temperature dependence of beta-casein phosphopeptide (1-

25) as analyzed by circular dichroism, FTIR spectroscopy, and analytical ultracentrifugation. *J. Protein Chem*. **21**, 307–321.

Fay L.B. and Brevard H. (2005). Contribution of mass spectrometry to the study of the Maillard reaction in food. *Mass Spectrom. Rev*. **24**, 487–507.

Fenn J.B., Mann M., Meng C.K., Wong S.F. and Whitehouse C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71.

Ferreira F., Hawranek T., Gruber P., Wopfner N. and Mari A. (2004). Allergic cross-reactivity: from gene to the clinic. *Allergy*. **59**, 243–267.

Ferrer E., Alegría A., Farré R., Abellán P. and Romero F. (2000). Effects of thermal processing and storage on available lysine and furfural compounds contents of infant formulas. *J. Agric. Food Chem*. **48**, 1817–1822.

Fogliano V., Monti S.M., Visconti A., Randazzo G., Facchiano A.M., Colonna G. and Ritieni A. (1998). Identification of a beta-lactoglobulin lactosylation site. *Biochim. Biophys. Acta* **1388**, 295–304.

Fu M.X., Requena J.R., Jenkins A.J., Lyons T.J., Baynes J.W. and Thorpe S.R. (1996). The advanced glycation end product, Nepsilon-(carboxymethyl)lysine, is a product of both lipid peroxidation and glycoxidation reactions. *J. Biol. Chem*. **271**, 9982–9986.

Fujita M., Baba R., Shimamoto M., Sakuma Y. and Fujimoto S. (2007). Molecular morphology of the digestive tract; macromolecules and food allergens are transferred intact across the intestinal absorptive cells during the neonatal-suckling period. *Med. Mol. Morphol*. **40**, 1–7.

Galvani M., Hamdan M., Herbert B. and Righetti P.G. (2001). Alkylation kinetics of proteins in preparation for two-dimensional maps: a matrix assisted laser desorption/ionization-mass spectrometry investigation. *Electrophoresis*. **22**, 2058–2065.

Gerrard J.A., Meade S.J., Miller A.G., Brown P.K., Yasir S.B., Sutton K.H. and Newberry M.P. (2005). Protein cross-linking in food. *Ann. N. Y. Acad. Sci.* **1043**, 97–103.

Gómez-Ollés S., Cruz M.J., Renström A., Doekes G., Morell F. and Rodrigo M.J. (2006). An amplified sandwich EIA for the measurement of soy aeroallergens. *Clin. Exp. Allergy*. **36**, 1176–1183.

Hanash S. (2003). Disease proteomics. *Nature* **422**, 226–232.

Hasenkopf K., Rönner B., Hiller H. and Pischetsrieder M. (2002). Analysis of glycated and ascorbylated proteins by gas chromatography-mass spectrometry. J. *Agric. Food Chem*. **50**, 5697–5703.

Heddleson R.A., Park O. and Allen J.C. (1997). Immunogenicity of casein phosphopeptides derived from tryptic hydrolysis of beta-casein. *J. Dairy Sci*. **80**, 1971–1976.

Hiemori M., Bando N., Ogawa T., Shimada H., Tsuji H., Yamanishi R. and Terao J. (2000). Occurrence of IgE antibody-recognizing N-linked glycan moiety of a soybean allergen, Gly m Bd 28K. *Int. Arch. Allergy Immunol*. **122**, 238–245.

Humeny A., Kislinger T., Becker C.M. and Pischetsrieder M. (2002). Qualitative determination of specific protein glycation products by matrix-assisted laser desorption/ionization mass spectrometry Peptide mapping. *J. Agric. Food Chem*. **50**, 2153–2160.

Ikeda K., Nagai R., Sakamoto T., Sano H., Araki T., Sakata N., Nakayama H., Yoshida M., Ueda S. and Horiuchi S. (1998). Immunochemical approaches to AGE-structures: characterization of anti-AGE antibodies. *J. Immunol. Methods*. **215**, 95–104.

Kasai N. and Ikehara H. (2005). Stepwise extraction of proteins and carbohydrates from soybean seed. *J. Agric. Food Chem.* **53**, 4245–4252.

Kato Y., Watanabe H. and Matsuda T. (2000). Ovomucoid rendered insoluble by heating with wheat gluten but not with milk casein. *Biosci. Biotechnol. Biochem*. **64**, 198-201.

Koppelman, S. J., Bruijnzeel-Koomen, C.A., Hessing, M. and de Jongh, H.H. (1999). Heat-induced conformational changes of Ara h1, a major peanut allergen, do not affect its allergenic properties. *J. Biol. Chem*. **274,** 4770–4777.

Kopper R.A., Odum N.J., Sen M., Helm R.M., Steve Stanley J. and Wesley Burks A. (2004). Peanut protein allergens: gastric digestion is carried out exclusively by pepsin. *J. Allergy Clin. Immunol*. **114**, 614–618.

Krause R., Kühn J., Penndorf I., Knoll K. and Henle T. (2004). N-Terminal pyrazinones: a new class of peptide-bound advanced glycation end-products. *Amino Acids*. **27**, 9–18.

Kumosinski T.F., Brown E.M. and Farrell H.M. Jr. (1991). Three-dimensional molecular modeling of bovine caseins: kappa-casein. *J. Dairy Sci*. **74**, 2879–8287.

Lane C.S. (2005). Mass spectrometry-based proteomics in the life sciences. *Cell Mol. Life Sci*. **62**, 848–869.

L'Hocine L., Boye J.I. and Munyana C. (2007). Detection and quantification of soy allergens in food: study of two commercial enzyme-linked immunosorbent assays. *J. Food Sci.* **72**, 145–153.

Lo T.W., Westwood M.E., McLellan A.C., Selwood T. and Thornalley P.J. (1994). Binding and modification of proteins by methylglyoxal under physiological conditions. A kinetic and mechanistic study with Nalpha-acetylarginine, Nalpha-acetylcysteine and Nalpha-acetyllysine and bovine serum albumin. *J. Biol. Chem*. **269**, 32299–32305.

Luz Sanz M., Corzo-Martínez M., Rastall R. A., Olano A. and Moreno F.J. (2007). Characterization and in vitro digestibility of bovine β-lactoglobulin glycated with galacto-oligosaccharides. *J. Agric. Food Chem.* **55**, 7916–7925.

Maleki, S. J., Chung, S. Y, Champagne, E. T, and Raufman, J. P. (2000). The effects of roasting on the allergenic properties of peanut proteins. *J. Allergy Clin. Immunol*. **106**, 763–768.

Malin E.L., Brown E.M., Wickham E.D. and Farrell H.M. Jr. (2005). Contributions of terminal peptides to the associative behaviour of α1-casein. *J. Dairy Sci*. **88**, 2318–2328.

Marvin L.F., Parisod V., Fay L.B. and Guy P.A. (2002). Characterization of lactosylated proteins of infant formula powders using two-dimensional gel electrophoresis and nanoelectrospray mass spectrometry. *Electrophoresis*. **23**, 2505–2512.

Miller A.G. and Gerrard J.A. (2005). Assessment of protein function following cross-linking by alpha-dicarbonyls. *Ann. N. Y. Acad. Sci*. **1043**, 195–200.

Mills E.N. and Breiteneder H. (2005). Food allergy and its relevance to industrial food proteins. *Biotechnol. Adv*. **23**, 409-414.

Minikiewicz P., Slangen C.J., Lagerwerf F.M., Haverkamp J., Rollema H.S. and Visser S. (1996). Reversed-phase high-performance liquid chromatographic separation of bovine kappa-casein macropeptide and characterization of isolated fractions. *J. Chromatogr. A*. **743**, 123–135.

Monaci L. and van Hengel A.J. (2007). Effect of heat treatment on the detection of intact bovine beta-lactoglobulins by LC mass spectrometry. *J. Agric. Food Chem*. **55**, 2985–2992.

Morales F., J., Romero C. and Jimenez-Perez S., (1996). Evaluation of heat-induced changes in Spanish commercial milk: hydroxymethylfurfural and available lysine content. *Int. J. of Food Sci. Tech.* **31,** 411-418.

Morales F.J., Romero C. and Jiménez-Pérez S. (1997). Chromatographic determination of bound hydroxymethylfurfural as an index of milk protein glycosylation. *J. Agric. Food Chem.* **45**, 1570–1573.

Natale M., Bisson C., Monti G., Peltran A., Garoffo L.P., Valentini S., Fabris C., Bertino E., Coscia A. and Conti A. (2004). Cow's milk allergens identification by two-dimensional immunoblotting and mass spectrometry. *Mol. Nutr. Food Res*. **48**, 363–369.

Newkirk M.M., Goldbach-Mansky R., Lee J., Hoxworth J., McCoy A., Yarboro C., Klippel J. and El-Gabalawy H.S. (2003). Advanced glycation end-product (AGE)-damaged IgG and IgM autoantibodies to IgG-AGE in patients with early synovitis. *Arthritis Res. Ther*. **5**, 82–90.

Nielsen H.K., De Weck D., Finot P.A., Liardon R. and Hurrell R.F. (1985). Stability of tryptophan during food processing and storage. 1. Comparative losses of tryptophan, lysine and methionine in different model systems. *Br. J. Nutr*. **53**, 281–292.

Odani H., Shinzato T., Usami J., Matsumoto Y., Brinkmann Frye E., Baynes J.W. and Maeda K. (1998). Imidazolium crosslinks derived from reaction of lysine with glyoxal and methylglyoxal are increased in serum proteins of uremic patients: evidence for increased oxidative stress in uremia. *FEBS Lett*. **427**, 381–385.

Oya T., Hattori N., Mizuno Y., Miyata S., Maeda S., Osawa T. and Uchida K. (1999). Methylglyoxal modification of protein. Chemical and immunochemical characterization of methylglyoxal-arginine adducts. *J. Biol. Chem.* **274**, 18492–18502.

Parker E.A., Donato L. and Dalgleish D.G. (2005). Effects of added sodium caseinate on the formation of particles in heated milk. *J. Agric. Food Chem*. **53**, 8265–8272.

Perkins D.N., Pappin D.J., Creasy D.M. and Cottrell J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **20**, 3551–3567.

Peterson C.G., Hansson T., Skott A., Bengtsson U., Ahlstedt S. and Magnussons J. (2007). Detection of local mast-cell activity in patients with food hypersensitivity. *J. Investig. Allergol. Clin. Immunol.* **17,** 314–320.

Peterson P.A. and Rask L. (1971). Studies on the fluorescence of the human vitamin A-transporting plasma protein complex and its individual components. *J. Biol. Chem.* **246**, 7544–7550.

Phizicky E., Bastiaens P.I., Zhu H., Snyder M. and Fields S. (2003). Protein analysis on a proteomic scale. *Nature* **422**, 208–215.

Plaza-Martín A.M., Jiménez-Feijoo R., Andaluz C., Giner-Muñoz M.T., Martin-Mateos M.A., Piquer-Gibert M. and Sierra-Martínez J.I. (2007). Polysensitization to aeroallergens and food in eosinophilic esophagitis in a pediatric population. *Allergol. Immunopathol.* **35**, 35–37.

Ring J., Brockow K. (2002). Adverse drug reactions: mechanisms and assessment. *Eur. Surg. Res*. **34,** 170–175.

Roepstorff P. and Fohlman J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom*. **11**, 601.

Roychaudhuri R., Sarath G., Zeece M. and Markwell J. (2004). Stability of the allergenic soybean Kunitz trypsin inhibitor. *Biochim. Biophys. Acta.* **1,** 207–212.

Rupa P. and Mine Y. (2003). Immunological comparison of native and recombinant egg allergen, ovalbumin, expressed in *Escherichia coli*. *Biotechnol. Lett*. **25**, 1917–1924.
Rytkönen J., Valkonen K.H., Virtanen V., Foxwell R.A., Kyd J.M., Cripps A.W. and Karttunen T.J. (2006). Enterocyte and M-cell transport of native and heat-denatured bovine beta-lactoglobulin: significance of heat denaturation. *J. Agric. Food Chem*. **54**, 1500–1507.

Sali A., Glaeser R., Earnest T. and Baumeister W. (2003). From words to literature in structural proteomics. *Nature* **422**, 216–225.

Scaloni A., Perillo V., Franco P., Fedele E., Froio R., Ferrara L. and Bergamo P. (2002). Characterization of heat-induced lactosylation products in caseins by immunoenzymatic and mass spectrometric methodologies. *Biochim. Biophys. Acta*. **1598**, 30–39.

Schmitt A., Gasic-Milenkovic J. and Schmitt J. (2005). Characterization of advanced glycation end products: mass changes in correlation to side chain modifications. *Anal. Biochem*. **346**, 101–106.

Schmitt A., Schmitt J., Münch G. and Gasic-Milencovic J. (2005). Characterization of advanced glycation end products for biochemical studies: side chain modifications and fluorescence characteristics. *Anal. Biochem*. **338**, 201-215.

Sedmark J.J. and Grossberg SE (1977). A rapid, sensitive, and versatile assay for protein using Coomassie brilliant blue G250. *Anal Biochem*. **79,** 544–552.

Seiquer I., Díaz-Alguacil J., Delgado-Andrade C., López-Frías M., Muñoz Hoyos A., Galdó G. and Navarro M.P. (2006). Diets rich in Maillard reaction products affect protein digestibility in adolescent males aged 11-14 y. *Am. J. Clin. Nutr*. **83**, 1082–1088.

Sharma S., Kumar P., Betzel C. and Singh T.P. (2001). Structure and function of proteins involved in milk allergies. *J. Chromatogr. B Biomed. Sci. Appl*. **756**, 183-187.

Shin, D. S., Compadre, C. M., Maleki, S. J., Kopper, R. A., Sampson, H., Huang, S. K., Burks, A. W. and Bannon, G. A. (1998). Biochemical and structural analysis of the IgE binding sites on Ara h1, an abundant and highly allergenic peanut protein. *J. Biol. Chem*. **22**, 13753–13759.

Sicherer S.H., Morrow E.H. and Sampson H.A. (2000). Dose-response in double-blind, placebo-controlled oral food challenges in children with atopic dermatitis. *J. Allergy Clin. Immunol*. **105**, 582–586.

Sicherer S.H. and Sampson H.A. (2007). Peanut allergy: emerging concepts and approaches for an apparent epidemic. *J. Allergy Clin. Immunol*. **120**, 491-503.

Siciliano R., Rega B., Amoresano A. and Pucci P. (2000). Modern mass spectrometric methodologies in monitoring milk quality. *Anal Chem*. **72**, 408–415.

Skripak J.M., Matsui E.C., Mudd K. and Wood R.A. (2007). The natural history of IgE-mediated cow's milk allergy. *J. Allergy Clin. Immunol*. **120,** 1172–1177.

Smyth E., Clegg R.A. and Holt C. (2004). A biological perspective on the structure and function of caseins and casein micelles *Int. J. Dairy Technol.* **57**, 121–126.
Sun H.S., Hill Park J. and Won K.S. (2007). Determination of proteins in infant formula by high-performance liquid chromatography–electrospray tandem mass spectrometry *J. Chromatography B.* **845, 69–7**3.

Takagi K., Teshima R., Okunuki H., Itoh S., Kawasaki N., Kawanishi T., Hayakawa T., Kohno Y., Urisu A. and Sawada J. (2005). Kinetic analysis of pepsin digestion of chicken egg white ovomucoid and allergenic potential of pepsin fragments. *Int. Arch. Allergy Immunol*. **136**, 23–32.

Tatsumi E. and Hirose M. (1997). Highly ordered molten globule-like state of ovalbumin at acidic pH: native-like fragmentation by protease and selective modification of Cys367 with dithiodipyridine. *J. Biochem.* **122**, 300–308.

Teufel M., Biedermann T., Rapps N., Hausteiner C., Henningsen P., Enck P. and Zipfel S. (2007). Psychological burden of food allergy. *World J. Gastroenterol*. **13,** 3456-3465.

Thornalley P.J. (2005). Measurement of protein glycation, glycated peptides and glycation free adducts. *Perit. Dial. Int*. **25**, 522–533.

Tyers M. and Mann M. (2003). From genomics to proteomics. *Nature* **422**, 193–197.

Uchida K., Khor O.T., Oya T., Osawa T., Yasuda Y. and Miyata T. (1997). Protein modification by a Maillard reaction intermediate methylglyoxal. Immunochemical detection of fluorescent 5-methylimidazolone derivatives in vivo. *FEBS Lett*. **410**, 313-318.

Urisu A., Ando H., Morita Y., Wada E., Yasaki T., Yamada K., Komada K., Torii S., Goto M. and Wakamatsu T. (1997). Allergenic activity of heated and ovomucoid-depleted egg white. *J. Allergy Clin. Immunol*. **100**, 171-176.

Van der Plancken I., Van Remoortere M., Van Loey A. and Hendrickx M.E. (2004). Trypsin inhibition activity of heat-denatured ovomucoid: a kinetic study. *Biotechnol. Prog.* **20**, 82-86.

van Ree R. (2004). Clinical importance of cross-reactivity in food allergy. *Curr. Opin. Allergy Clin. Immunol*. **4**, 235-240.

Vieths S., Scheurer S. and Ballmer-Weber B. (2002). Current understanding of cross-reactivity of food allergens and pollen. *Ann. N. Y. Acad. Sci*. **964,** 47-68.

Viquez, O. M., Konan, K. N. and Dodo, H. W. (2003). Structure and organization of the major peanut allergene Ara h1. *Mol. Immunol.* **40**, 565–571.

Wal J. M. (2004). Bovine milk allergenicity. *Ann. Allergy Asthma Immunol*. **93,** S2–11.

Xiang P., Haas E.J., Zeece M.G., Markwell J. and Sarath G. (2004). C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen. *Planta* **22**, 56–63.

Yoshino K., Sakai K., Mizuha Y., Shimizuike A. and Yamamoto S. (2004). Peptic digestibility of raw and heat-coagulated hen's egg white proteins at acidic pH range. *Int. J. Food Sci. Nutr*. **55**, 635–640.

Zhu H., Bilgin M. and Snyder M. (2003). Proteomics. *Annu. Rev. Biochem.* **72**, 783–812.

Imperial College London

Faculty of Natural Science
Division of Molecular Bioscience

Analysing the proteomic basis of post-mating behaviours in *Anopheles* mosquitoes

Correlatore: Doct. Maria PANICO

Dottoranda: Francesca BATTAGLIA

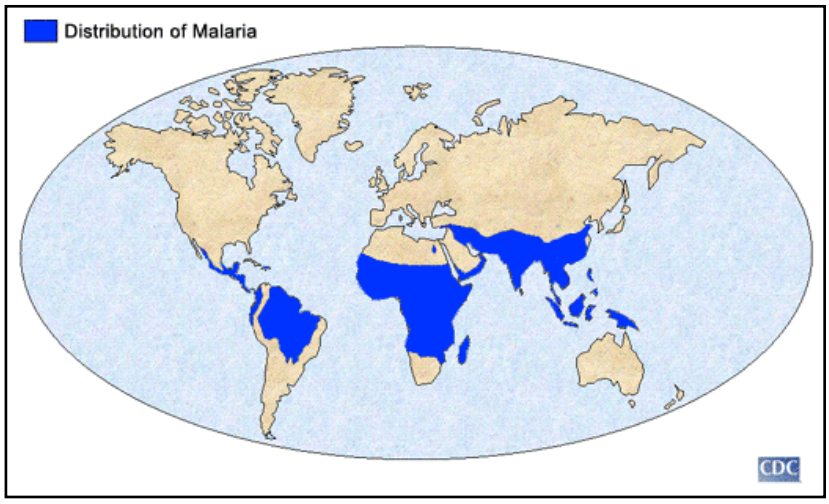
31st January 2008

Analysing the proteomic basis of post-mating behaviours in *Anopheles* mosquitoes

1. INTRODUCTION

MALARIA DISEASE

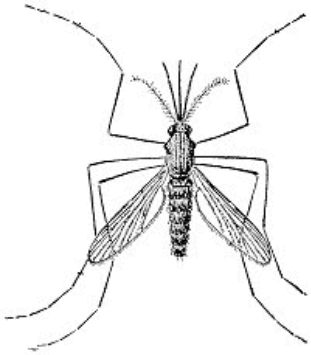
Malaria is caused by an infection with protozoa of the genus *Plasmodium.* The name malaria, derived from the Italian *mala aria* expression, meaning bad air, comes from the linkage suggested by Lancisi (1717) of malaria with the poisonous vapors of swamps. An estimated three billion people, almost half the world's population, live in areas where malaria transmission occurs. Malaria is endemic in 107 countries and territories in tropical and subtropical regions. Between 350 and 500 million cases of clinical malaria occur each year, leading to an estimated 1 million deaths. Over 80 per cent of these deaths, or at least 800,000 a year, occur among African children under age five. Malaria is a disease of warm, humid climates where reservoirs of water constitute perfect breeding grounds for the Anopheles mosquito. With the bite of the mosquito, malaria parasites are transmitted from infected to healthy people. Once in the bloodstream, the mature parasites reach the liver where they multiply. The rapid multiplication of the parasite causes the destruction of red blood cells and the infection of new cells throughout the body. Depending upon the species of infected Anopheles mosquito, the infected person will become ill with malaria after about a week to several months, but mostly within 7-21 days. The most important sign of malaria is fever. The symptoms in children and adults infected with malaria might also include shivering, severe pain in the joints, headaches, vomiting, generalized convulsions and coma, but also coughing and diarrhoea. If children, in particular, are not treated within a day, the disease can lead to death. In many cases, severe anaemia is the attributable cause of death. Prompt and effective treatment of malaria within 24 hours of the onset of symptoms is necessary to prevent life-threatening complications.



The World Health Organization now recommends treating malaria using artemisinin-based combination therapies, which are based on combinations of artemisinin, extracted from the plant *Artemisia annua*, with other effective antimalarial medicines. When combined with other medicines, artemisinin derivatives are highly powerful, fast acting and very well tolerated. An effective vaccine against malaria would be a critical component to aid in the control of this disease, and at the moment there are several vaccines under studies and trials.

ANOPHELES GAMBIAE

Mosquitoes are insects from the family Culicidae. They have a pair of scaled wings, a pair of halteres, a slender body, and long legs. The females of many mosquito species suck blood (hematophagy) from other animals: that has made them one of the most deadly disease vectors known to man, killing millions of people over hundreds of years and continuing to kill hundreds of thousand per year by the spread of diseases. Both male and female mosquitoes are nectar feeders, but the female is also capable of haematophagy, they do not require blood for survival, but they need supplemental proteins for the development and laying of their eggs. In its life cycle the mosquito undergoes complete metamorphosis, going through four distinct stages: egg, larva, pupa, and adult (Timmermann & Briegel, 1999). Female mosquitoes lay their eggs one at a time or together in rafts of hundreds eggs on the surface in any stagnant water. This requires still water: mosquitoes do not breed in fast-moving water. The total time to go through all four stages depends on the temperature and the type of mosquito, but typically it takes 14 days or less in warmer weather. In various species the time varies from 4 to 30 days.



In many female mosquitoes, the mouth parts form a long proboscis for piercing the skin of mammals to suck their blood. Male mosquitoes are distinctly smaller than females, with features such as feathered antennae and having no audible sound during flight, while female mosquitoes in flight emit a distinctive high-pitched buzz.

VECTOR PROGRAM

Mosquito species of the *Anopheles gambiae* complex represents the major vectors of human malaria causing enormous troubles on global health and economies. While many insect pests have long been defeated by population control measures such as insecticides or release of sterile males, for others, including *A. gambiae*, classical control measures have largely failed to deliver long-term solutions. Disease endemic countries often do not have the

economic resources and the logistics to sustain control efforts like the massive and prolonged use of insecticides. This global health problem has stimulated scientific research to study the biology and the genetics of *Anopheline* mosquitoes, with the aim of generating new molecular tools to fight the vector of malaria, starting from the *A. gambiae* genome sequencing which was completed a few years ago. These molecular advances have made available the expression of genes that can block the transmission of *Plasmodium* in model systems or express traits facilitating the implementation of sterile insect techniques for vector control. The translation of these achievements into suitable control measures still represents a major scientific and technical challenge.

The research on post-mating behavior of *Anopheles* females is not yet completely studied, but it could be very important for vector control programs (Catteruccia *et al.*, 2005). Mating induces many behavioural and physiological changes in females of various insect species, such as reduced readiness to mate, induced ovulation and oviposition, and reduced lifespan. In *Anopheles* mosquitoes for instance, females mate once in their lifetime, after that they become refractory to further mating, with polyandry occurring only in a small percentage of individuals (Tripet *et al.*, 2005). Up to now, the biological bases of female post-mating behaviours have been largely studied in *Drosophila*, and the results show that male accessory glands (MAGs) secretions are really important in the transfer of sperm from the male to the female reproductive apparatus, and in a variety of functions to ensure that the male will sire a significant proportion of the female's progeny, if not all (Swanson *et al.*, 2001). The major components of MAGs secretions are proteins, named accessory glands proteins (Acps), other secretions include carbohydrates, protein-bound lipids, small amounts of amino acids and amines. Acps are the most important modulators of female post-mating behaviours and can be divided in 3 groups, based on molecular size. The first group is made up of peptides, generally smaller than 100 amino acids. These comprise the majority (about 75%) of the *Drosophila* Acps (Muller et al., 2004). They are implicated in female refractoriness in many insects, including mosquitoes. The second group contains proteins of 200-400aa. Usually these proteins are glycosylated and need to be activated by cleavage. The third group is composed of large, often glycosylated, proteins.

Knowledge of the role of Acps in *Anopheles* is instead still very limited and no Acp has been identified as yet, but there is evidence that in many organisms, including *Anopheles*, the refractoriness behaviour depends on the female nervous system and that innervations of the reproductive organs is essential in modulating behaviour (Ottiger *et al.*, 2000). A deeper study of mosquito Acps, their molecular mechanisms and functions, can provide new tools for

vector control, for instance through the development of transgenic mosquitoes in which these signals are constitutively switched on in females, therefore bringing about a permanent refractory state, or through the design of chemical formulations that mimic their function and block relevant female receptors (Windbichler *et al.*, 2007).

This project has focused on the identification of the proteinaceous factors that induce behavioural changes in *Anopheles* females upon mating, using mass spectrometry to analyze the Acps transferred from males to females. These analyses have identified a series of genes that may play an important role in modulating female post-mating behaviours, and may constitute novel targets for vector control in the field.

2. METHODS AND RESULTS

Tissues from *Anopheles* male MAGs, atrium and spermatheca from mated and virgin females were dissected and the protein extracts analyzed immediately by SDS-PAGE. The atrium from mated female contents the mating plug, produced by the mosquito male and transferred to the female during mating. The atrium from the virgin female is analyzed as control. For the sample proteins separation, precast NuPAGE 4-12% Bis –Tris gels were used. They were 1.0 mm x 10 well, made by Invitrogen (Carlsbad, CA, USA). To stain the proteins, an Invitrogen NOVEX colloidal blue stain kit was used. The gel spots of interest were excited from the gel and fragmented. The gel pieces were rehydrated with 0.1 M NH_4HCO_3 containing 10 mM DTT, to perform the reduction. After an acetonitrile wash, a solution of 55 mM iodoacetic acid in 0.1 M NH_4HCO_3 was added to the sample, to cap the reduced cysteines. Afterwards, the sample was dehydrated with 50 ng/mL of trypsin solution in 100 mM NH_4HCO_3, and incubated overnight at 37°C. The peptides produced by the proteolysis were collected by washing the sample with acetonitrile and 5% formic acid. The mass spectrometry analysis was conducted with Applied biosystem mass spectrometer API Qstar Pulsar (Foster City, CA, USA), connected with a micro-HPLC Ultimate by LC Packings from Dionex, (Leeds United Kingdom). The MS/MS spectra were analyzed with the MASCOT software (http://www.matrixscience.com) for the database research, selecting trypsin as proteolytic enzyme. There was no restriction of the database for a specific organism, and oxidation of Met was selected among all the possible modifications of the aminoacid side chain.

The aim of this project was to understand which proteins are produced in the accessory gland by the male mosquito and which of them are transferred to the female. Basically, the fresh sample from mosquito was analyzed by SDS-PAGE in the same day of

extraction, in order to avoid degradation and aggregation of the proteins (Fig. 1). All the bands in the gel were cut and submitted to trypsin digestion. To identify the proteins produced by the accessory glands, isolated accessory glands from male mosquitoes were prepared, analyzed by SDS-PAGE. 16 spots, of different stain intensity, were digested with trypsin and analyzed by LC-MS/MS (Table 1). After that, a second sample extracted from a female mosquito was analyzed according the same protocol, and a third sample, from non-mated female, was similarly processed as blank control, to understand if the proteins found both in male and mated female samples were passed in the female by mating or were already present in the female reproductive tract (data not shown). The peptides sequenced by the software Information Dependent Acquisition (IDA), that helps select the best ions to target for MS/MS data acquisition during an LC analysis,, have been manually validated as well.

The analysis of male accessory gland proteins and the mating plug extracted from mated females permitted the isolation and the characterization of peptides directly passed from the male to the female. Several peptides were found and analyzed, while other control studies were carried out to confirm the data. In particular, a protein of 65566 Da was identified in both male and female samples. Its identity remains confidential until the completion of biological studies by our collaborators, but it is already clear that it is a very promising candidate for modulating female post-mating behaviour. In summary, this project has yielded a wealth of proteomic information and has opened up new avenues of malaria research.

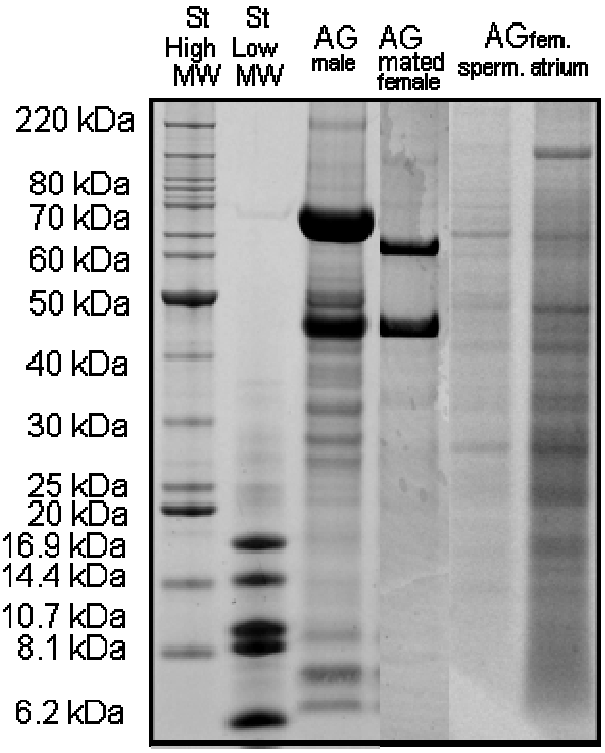


Fig. 1. SDS-PAGE of the biological extracts from male and female *Anopheles Gambiae* (AG) mosquitoes. AG male: protein extract from accessory glands. AG mated female: protein content of the mating plug extracted from mated female. AG female spermatheca and atrium: protein extracts from the non mated mosquitoes organs.

Table 1. *Anopheles Gambiae* proteins identified in the Male Accessory Gland Proteins sample, by LC-MS/MS analysis and database search.

Spot MW	Found SDS-PAGE (kDa)	Protein	MW Calculated (Da)	Number Peptides sequenced	% coverage
1	6.5	AGAP012830-PA	8846	3	50
		AGAP009358-PA	8828	3	43
2	7.0	AGAP009370-PA	11174	3	26
		AGAP003412-PA	7536	3	41
3	10.0	AGAP001502-PA	10891	3	30
		AGAP001502-PA	8330	3	32
4	14.0	8437	239085	3	2
5	17.0	AGAP010933-PA	17757	4	19
6	27.0	AGAP007162-PA	21667	6	32
		AGAP005792-PA	27510	5	24
		AGAP004960-PA	28799	3	12
		AGAP005791-PA	26986	2	9
		AGAP007643-PA	28349	8	30
		AGAP007643-PC	28443	9	26
		Plug	65566	7	13
7	28.0	AGAP001053-PA	24881	7	30
		AGAP012364-PA	29619	5	22
		AGAP001151-PA	29375	4	15
		AGAP001053-PC	23775	5	18
		AGAP002401-PA	25863	3	11
8	30.0	AGAP008071-PA	42420	6	15
		AGAP006418-PA	29899	5	30
		AGAP006782-PA	33103	5	14
9	38.0	AGAP001797-PA	79021	15	16
		AGAP001799-PA	32667	13	42
		AGAP001903-PA	35717	11	32
		AGAP009623-PA	35676	5	15
		AGAP009673-PA	40203	4	11

10	40.0	AGAP007420-PA	48010	3	7
		AGAP003592-PA	34057	2	8
		AGAP005264-PA	24738	3	18
		AGAP005095-PA	42156	8	35
		AGAP011516-PA	42710	8	31
11	45.0	AGAP005095-PA	42156	8	35
		AGAP011516-PA	42710	8	31
12	50.0	Plug	65566	18	40
		AGAP008019-PA	59531	10	19
		AGAP007406-PA	50700	5	9
		AGAP005095-PA	42156	5	16
		AGAP000883-PA	48799	6	15
		AGAP005558-PA	52302	2	6
		AGAP006037-PA	40264	2	6
		AGAP005784-PA	44770	3	7
		AGAP009362-PA	24470	4	23
13	72.0	Plug	65566	22	46
14	80.0	AGAP001497-PA	107098	20	24
		AGAP004877-PA	102923	17	22
		AGAP007532-PA	107554	9	10
15	100.0	Plug	65566	16	35
		AGAP004877-PA	102923	10	13
16	220.0	AGAP010147-PA	225159	85	45
		AGAP004877-PA	102923	4	8

REFERENCES

Catteruccia F., Benton J.P., Crisanti A. (2005). An Anopheles transgenic sexing strain for vector control. *Nat. Biotechnol.,* **23,** 1414–1417.

Mueller J.L., Ripoll D.R., Aquadro C.F., Wolfner M.F. (2004). Comparative structural modeling and inference of conserved protein classes in Drosophila seminal fluid. *Proc Natl Acad Sci U S A,* **101,** 13542–13547.

Ottiger M., Soller M., Stocker R.F., Kubli E. (2000). Binding sites of Drosophila melanogaster sex peptide pheromones. *J. Neurobiol.,* **44,** 57–71.

Swanson W.J., Clark A.G., Waldrip-Dail H.M., Wolfner M.F., Aquadro C.F. (2001). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. *Proc. Natl. Acad. Sci. U S A,* **98,** 7375–7379.

Tripet F., Thiemann T., Lanzaro G.C. (2005). Effect of seminal fluids in mating between M and S forms of Anopheles gambiae. *Med Entomol.*, **42,** 596–603.

Windbichler N., Papathanos P.A., Catteruccia F., Ranson H., Burt A., Crisanti A. (2007). Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. *Nucleic Acids Res*. **35**, 5922–5933.