UNIVERSITÀ DEGLI STUDI DI FOGGIA

DIPARTIMENTO DI SCIENZE AGRARIE
DEGLI ALIMENTI E DELL'AMBIENTE

DOTTORATO DI RICERCA IN
"BIOTECNOLOGIE DEI PRODOTTI ALIMENTARI" (XXVIII CICLO)

COORDINATORE: PROF. MATTEO ALESSANDRO DEL NOBILE

PROTEOMICS-BASED CHARACTERIZATION OF FOOD ALLERGENS AND EFFECTS FROM FOOD PROCESSING

Dottorando: Francesco Nicola Troiano

Tutor: Dott.ssa Carmen Palermo

ANNO ACCADEMICO 2015

ABSTRACT

Food contains a lot of proteins, but only a small fraction of them are allergens either in their native forms or in products resulting from food processing. There is a need for sensitive and rapid methods for detecting the presence of allergens in foods, as well as analyse the modifications induced by food processing. This research is important if we realize the potential of new analitycal strategies and novel processing techniques that may reduce the allergenicity of foods. Nowadays, mass Spectrometry and Liquid Chromatography (LC) are extensively used for the characterization of allergenic proteins and peptides: the application of proteomics for the analysis of allergenic proteins has been recently termed allergenomics. Unfortunately, the variability of food allergens makes it difficult to develop a generic and universal method for their characterization. A major challenge for MS techniques is sample preparation and its related issues, as in the case of *Apiaceae* allergens, for many of which there is a lack of reference materials and methodologies. This PhD thesis research project aimed to make a contribution for the determination of food allergens by LCMS and to evaluate the effect of some food processing on allergenicity of food through nanoliquid chromatography and electrospray ionization-ion trap mass spectrometry (nanoHPLC-ESI-IT-MS^n). A specific bioinformatic approach was developed, in order to characterize the most important food allergens from soft cheese and plant samples from the *Apiaceae* family and to describe the effects to the allergenic potential of soft cheese samples during several conditioning systems. In this work, the peptides FVAPFPEVF from cow's milk (CM) allergen Bos d 9 and the peptide QEPVLGPVRGPFPIIV from CM allergen Bos d 11 were identified and proposed as valid marker peptides for CM allergens detection and for future analysis in this field, e.g. for quantification purposes. In addition, the results demonstrated the effect of some packaging conditions on reducing the total allergenic power of soft cheese, and showed the condition treatment S1 (based on potassium sorbate) as an optimal solution for reducing the total potential of soft cheese, both in terms of allergenic intensity and in terms of allergen variability. Using a similar strategy, several experiments were also performed for detecting the presence of *Apiaceae* allergens from plant samples and to characterize one or more potential markers for their detection in food using LCMS. Among those, the hydrophilic peptide FYETKDTDILAAFR from the allergen Spi o Rubisco is proposed as a potential marker for routine detection of *Apiaceae* allergens in food through ESI-MS^n. Other aspects regarding this allergen should be investigated, such as its behaviour under food processing and its role in the allergenic reactions (epitope mapping). In this last step, a particular attention was given to the sample preparation and to the selection of eco-compatible extraction buffers, and to the overall optimization of LCMS experimental conditions.

CONTENTS

ABBREVIATION OF AMINO ACIDS

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	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

CHAPTER 1

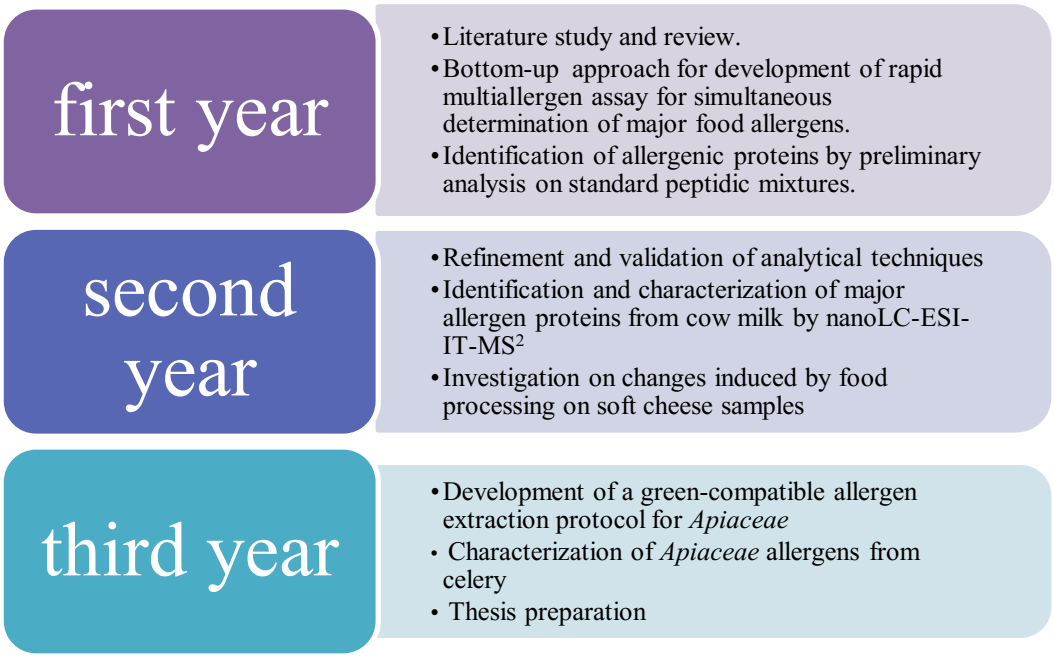
INTRODUCTION

1.1 SIGNIFICANCE AND MANAGEMENT OF DOCTORAL RESEARCH

The last decade has seen a rapid increase in our knowledge of the molecules in foods that cause and trigger allergic reactions. The post-genomic era, with its explosion of information about protein and genome sequences, is allowing us to study molecular relationships in new ways, and notably within the context of evolution. Gel-free proteomics of food allergens by mass spectrometry (MS) analysis is widely used for the identification and characterization of allergenic proteins because of the high specificity and resolution power of this technique. Liquid chromatography (LC) is also used extensively for the separation and isolation of allergens. A major challenge for MS techniques is sample preparation and its related issues, especially for low abundant proteins such as plant allergens. Quantitative MS assays generally rely on measuring marker peptides (biomarkers) using both an internal standard (IS) and a calibration curve for the determination of the analyte's relative concentration or one or more "label-free" methods, that do not require IS (*Stevenson et al*., 2009). Quantitative MS of food allergens is a field under development, and more methods will be developed in the near future, including multiallergen assays for the simultaneous determination of major food allergens from complex processed food matrices. Allergenic proteins have not suddenly appeared on the protein landscape but are the result of a long chain of formative processes that resulted in the creation of protein architectures that are treated as allergenic by an atopic immune system. Allergens are restricted to a very small number of protein families which share characteristic three-dimensional structures or scaffolds. Consequently, allergenicity seems to be linked to certain structural features of molecules that are members of a limited number of protein families but we still do not understand why certain protein (or protein scaffolds) dominate the landscape of allergen structures. Indications are that the relationships between protein structure and allergenicity are very subtle, and for food proteins are further

complicated by relatively poorly understood processing-induced changes (*Verhoeckx et al.,* 2015). Such complex interacting factors underlie the reasons why we still do not comprehend why some food proteins, and not others, cause allergic reactions in man. Only an improved understanding of these factors and the mechanisms underlying the generation of aberrant IgE responses will enable us to understand what makes a protein become an allergen. Investigating the factors that modulate the allergenicity of proteins is a research challenge. This is important if we are to realize the potential of new analitycal strategies and novel processing techniques that may reduce the allergenicity of foods. Such knowledge is important not only for food safety purposes, but also for enabling health professionals to provide patients with the knowledge they need to avoid problems foods effectively and to support the development of specific therapies. The chart below (Fig. 1) shows the steps followed throughout the three years of doctoral research.

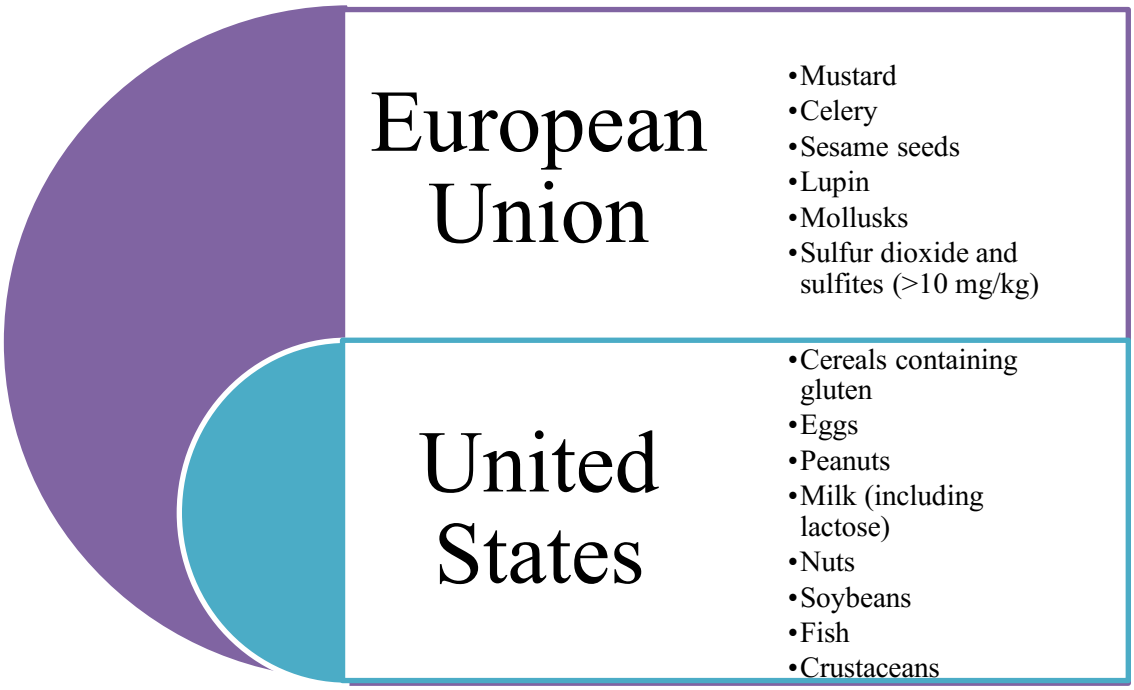
Figure 1: Management of doctoral research



1.2 BACKGROUND

The World Health Organization considers food allergy to be the fourth most important public health problem (*Kirsch et al.,* 2009). Food allergies related problems have consequences for the public health status, economy, and legislation when a big number of individuals must deal with dietary restrictions and food-triggered reactions ranging from mild to life threatening. For this reasons, stringent international food labelling directives have been developed in the last years, requiring food companies to declare ingredients with known allergenic potential (*Reg. EU* 1169/2011). Allergenicity (for allergic individuals) is caused by the food proteins and peptides either in their native forms or in products resulting from food processing interactions and modifications *(Kruse Faeste et al*., 2011). Differences in dietary habits and regulatory management by the food authorities among countries have produced slight variations in food allergen law. In the United States, eight foods (the so called "big-8") have been identified as causing the most of food allergenic reactions: crustaceans, egg, fish, milk, nuts, peanut, soybean, and wheat. Regulations require that the presence of these foods be declared on prepacked food products (*Food Allergen Labeling and Consumer Protection Act ,*2004). In addition to these eight major allergens, sesame must be labelled in Australia and New Zealand (*Australia New Zealand Standard*, 2013) and sesame and mustard must be labelled in Canada (*Canada Gazette*, 2011). The European Union has a longer list (14) of ingredients to be declared (see Fig. 2), including sulphites, wheat and cereals containing gluten and by adding celery, lupin, and molluscs to the most common allergens. In Japan and Korea, allergy to buckwheat is of considerable importance in addition to the eight major allergens (*Japan standards,* 2013). The public management of food allergy requires to develop sensitive and versatile analytical methods able to identify not only the presence of potential allergens in processed foods, but also the modifications induced by food processing and conditioning.

Figure 2: List of food allergens required to be declared in appropriate labels in compliance with legislation in the European Union and the United States

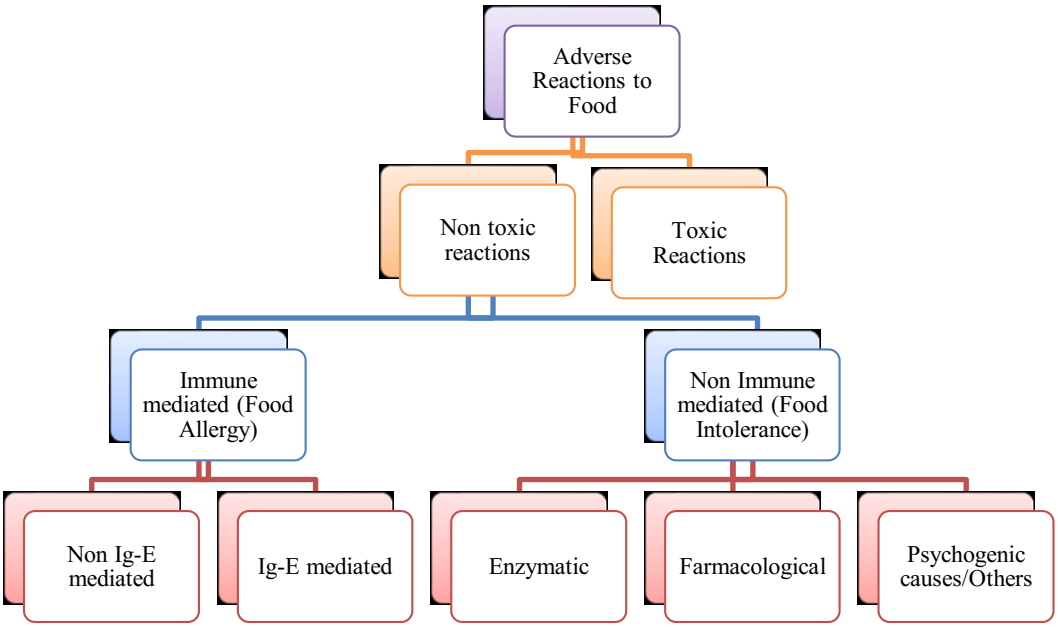


1.3 ADVERSE REACTIONS TO FOOD

The immune system possesses remarkable flexibility in the number of ways in which it works to protect the body from hazards, including infective microorganisms, viruses and parasites, employing both cellular agents to remove and inactivate hazards, as well as molecules, notably immunoglobulins (Igs), which form part of the humoral defense system. Igs are synthesized in a number of different forms, or isotypes, and have been classified on a structural, physicochemical and functional basis including IgA, IgG, IgM and IgE. All are characterized by an antibody-binding site generated to bind specifically to 'non-self' molecules, which are generally known as antigens. These include molecules found in microbial pathogens, parasites, environmental agents such as pollen and dietary proteins. The allergenic determinants or epitopes represent the structures recognized by IgE.

An allergen molecule can be formed by linear epitopes, which constitute a specific sequence of amino acids along its primary structure, and conformational epitopes generated by folding of proteins. High temperature, low pH or enzymatic digestion in food processing can destroy conformational epitopes but may not act over sequential epitopes. *(Mills et al.,* 2012)*)*. An adverse food reaction is a general term that can be applied to a clinically abnormal response to an ingested food or food additive. Adverse food reactions are common and often assumed by patients to be allergic in nature. Adverse reactions to foods are classified as either food hypersensitivity (allergy) or food intolerance (see Fig. 3).

Figure 3: Classification of adverse reactions to food



1.3.1 IG-E MEDIATED REACTIONS

IgE-mediated food-allergic reactions are rapid in onset (usually within minutes to 2 hours) and are the most widely known reactions associated with foods. Specific manifestations of IgE-mediated food hypersensitivity reactions can involve any system within the human body. These reactions frequently involve the skin, respiratory tract, gastrointestinal tract, and cardiovascular system. More severe symptoms and those involving multiple

systems are defined by the term “generalized anaphylaxis” and are often life threatening. In the allergic condition classified as a “type I hypersensitivity reaction”, the antibody repertoire to selected environmental antigens is altered, the body synthesizing larger quantities of the antibody isotype normally produced in response to parasitic infections, IgE. Allergens have been defined by the International Union of Immunological Societies as being molecules that must induce IgE-mediated (atopic) allergy in humans with a prevalence of IgE reactivity above 5%. *(Mills et al., 2012).*

1.3.2 NON-IGE-MEDIATED REACTIONS

Non-IgE-mediated food allergies typically present with more subacute or chronic symptoms isolated to the gastrointestinal tract that present within hours or days of food ingestion. Affected patients commonly present with a classic constellation of features that are consistent with well-described clinical disorders. These disorders include food protein-induced enterocolitis, food protein-induced proctocolitis, food protein induced gastroenteropathy, food-induced contact dermatitis, celiac disease with or without dermatitis herpetiformis (DH), and food-induced pulmonary hemosiderosis.

1.3.3 FOOD INTOLERANCE

Food intolerance is a general term describing an abnormal physiologic response to an ingested food or food additive. This reaction has not been proven to be immunologic in nature, which distinguishes these reactions from those occurring as a result of food allergy. Food intolerance may be caused by many factors including pharmacologic properties of the food (e.g. caffeine in coffee, tyramine in aged cheeses, sulfites in red wine), characteristics of the host such as metabolic disorders (e.g. lactase deficiency), and idiosyncratic responses.

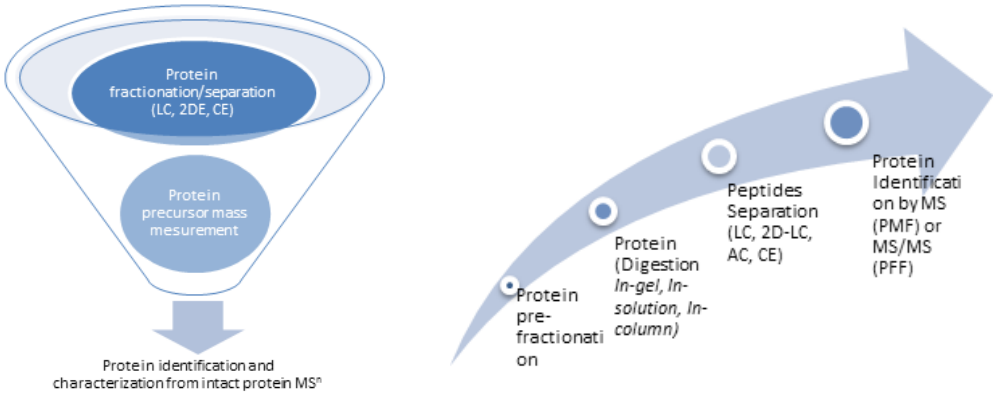
1.4 DIAGNOSTIC TECHNIQUES

Various techniques have been applied to determine the presence of allergens in foods. The methods are qualitative or quantitative and detect either the allergenic protein itself or a marker that typically signifies its presence. Although Immunoenzymatic assay techniques (such as the ELISA tests) are a major tool for screening, they are not sufficiently accurate and precise for quantitative determination of each allergen and become invalid when, for production or preservation processes the food undergoes treatments that alter the chemical structure of the epitope and cancel the antigen-antibody reaction. During the last decade the huge progress in mass spectrometry (MS) technology has increased the application of proteomics for the identification, characterization and quantitation of food allergens, including the study of variations in protein expression from extra- and intracellular conditions, co- and post-translational modifications, splicing variants, covalent and noncovalent associations. The application of proteomic methods for the analysis of allergenic proteins has been termed "allergenomics" (*Akagawa M. et al.,* 2007). Allergenomics can be used for the identification and quantitation of proteins and peptides, the determination of primary sequences, and the detection of protein interactions and modifications. The allergenome comprises many different protein and peptide allergens, and allergenomic information is becoming more and more complete and available in on-line resources like such as the Allergome database (*Mari, A. - Allergy Data Laboratories s.c*., 2003 - 2015). Separation of the proteins before MS analysis is essential to reduce sample complexity. Usually this separation is effectuated by electrophoresis on one-dimensional (1D) or 2D gels or by liquid chromatography (LC). Electrophoresis techniques, despite the many improvements achieved in the last years, are still valid and simple, by the way LC remains the superior separation technology in terms of sensitivity, reproducibility, recovery and versatility, thanks to the extensive availability of stationary phases and the easy interfacing with MS instruments.

1.4.1 PROTEIN IDENTIFICATION BY TOP-DOWN AND BOTTOM-UP PROTEOMICS

There are two principle approaches to proteomic analysis: top down and bottom up (*Aebersold & Mann*, 2003). In the top-down proteomics intact proteins are separated from complex biological matrix using traditional techniques such as liquid chromatography or gel electrophoresis, followed by their analysis. The bottom-up method , also called ‘‘shot-gun’’ proteomics, involves involves direct digestion of a biological sample using one or proteolytic enzyme (such as trypsin) or other digestion methods, that cleaves at defined sites to generate a complex peptide mixture. The digested samples are then analyzed by various techniques that include liquid chromatography and mass spectrometry (see Fig. 4). For very large proteins, a ‘‘middle-down’’ strategy, using limited digestion to produce larger peptides, has also been explored (*Han et al.*, 2008).

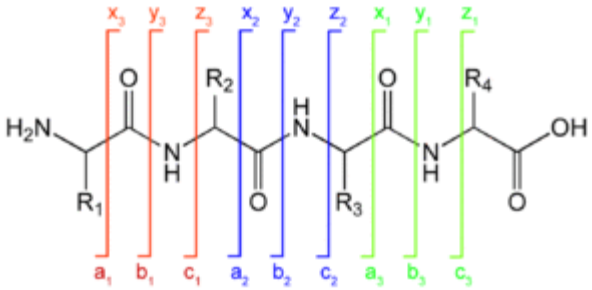
Figure 4. The top down (on the left) and bottom up (on the right) approaches in the proteomic analysis



Proteomics analysis generate a lot of data: *ad hoc* bioinformatic tools compare the spectra from the proteomics analysis with spectra in databases. One of most important resources in proteomic analysis is the fragmentation of analytic precursor ions into its product ions. Peptides generally split at the

peptide bonds in the backbone, leading to six possible fragment ions: the a, b, and c ions containing charged amino terminals and the x, y, and z ions containing charged carboxyl terminals (see Fig. 5).

Figure 5. Peptide fragmentation notation using the scheme of Roepstorff and Fohlman. Source: Wikipedia.



Spectra from single MS experiments are used for peptide mass fingerprinting (PMF) to compare the acquired peptide masses to those in protein databases, whereas multiple MS spectra can be used for either peptide fragmentation mass fingerprinting or to obtain peptide sequence information. The fragmentation is performed in MS^n instruments by a gas, with collision-induced dissociation (CID), by electron transfer dissociation (ETD), electron capture dissociation (ECD), or photo dissociation of the analyte to break up the molecular bonds. The most common series of fragments are the b and y ions, which occur after cleavage of the CO-NH bond. The peptide sequence interpretation is aided by the charge separation in multiply charged ions, which produce complimentary b-y ion pairs in the MS-MS spectrum of a peptide.

1.4.2 LIQUID CHROMATOGRAPHY FOR ALLERGENOMIC ANALYSIS

High-performance Liquid Chromatography (HPLC) is the favourite technology for the separation of allergenic protein or peptides. HPLC has an extensive choice of stationary phases and can also be used to separate small peptides either at physiological or at denaturing pHs; Other aspects contribute to the versatility of HPLC such the automation, the high sensitivity and the easy interfacing with various MS systems. HPLC separation techniques for allergenomic analysis can be combined in two or more dimensions that offer the possibility to separate the most complex mixtures of either digested or entire proteins. The most common combinations of HPLC phases are size exclusion/reversed phase chromatography (SEC-RPLC), ion exchange/reversed phase chromatography (IEX-RPLC) and ion exchange with affinity chromatography (IEX-AC). The choice of the proper buffer is a critical step in LCMS methods developing for allergenomic purposes: when interfacing with MS, LC mobile phases must be chosen in terms of volatility, ion strength and compatibility, in order to avoid undesirable effects such as ion suppression, sensitivity or signal loss and poor quality of the spray. A peptide's retention time in RPLC is directly related to its amino acid composition and depends from peptide's physicochemical characteristics, such as the isoelectrical point (pI). In some RPLC/MS methods several ion-pairing agents (e.g. formic acid or trifluoroacetic acid) are added to the LC eluents in order to positively charge the peptides or proteins and resolve them in a water/organic mobile phase basing on their hydrophobicity or hydrophilicity interaction with the stationary phase. Another powerful aspect of HPLC methods is the very low detection limit: in a common scenario, the separation of femtomole amounts of allergenic proteins or peptides is easily achievable.

1.4.3 MS IN PROTEOMIC ANALYSIS

Mass spectrometry is a powerful tool for allergen detection and characterization. A typical mass spectrometer is composed by:

- An ion source, a device that generates atomic and molecular ions by using several technologies or principles. The electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) are the most frequently used in proteomic analysis. With the ESI the compounds are ionized in multiple charged states and dispersed into an aerosol by applying a high electric voltage; a LC can be easily coupled to an ESI. The MALDI requires the crystallization of analytes in a matrix which absorbs energy from a laser; in the MALDI the compounds are generally ionized in single charged state.

- A mass analyser, that separates ions according to their mass to charge ratio. There are five general types of analysers that can be used for allergen characterization and differ in terms of mass resolution, range, sensitivity and accuracy: the quadrupole (Q) filters the ions by oscillating electrical fields, it has a good reproducibility but limited resolution; time-of-flight (TOF) separates ions by time, basing on their kinetic energy, it has higher accuracy and resolution but limited dynamic range; the quadrupole ion trap (QIT), in which the ions are trapped in an dynamic electric field and sequentially ejected, can retain limited number of ions so it is less suitable for quantitative purposes; the Fourier transform ion cyclotron resonance (FTICR), where the ions are excited at their resonant cyclotron frequencies and the Orbitrap analyser, similar to the FTICR, where the ions are electrostatically trapped in an orbit around an electrode that confines them so that they both orbit and oscillate along the electrode. Both the Orbitrap and the FTICR provides the highest mass accuracy and resolution power. (*Kruse Faeste et al.*, 2011).

- A detector, that records the charge and the current generated on (or across) a surface and produces a mass spectrum, a record of ions as a function of their m/z. Several types of electron multipliers are used for this purpose, and others including Faraday cups and ion to photon detectors.

CHAPTER 2

FOOD ANTIGENS

2.1 INTRODUCING FOOD ALLERGENS

Food allergens are generally glycoproteins with molecular weights ranging from 10 to 70 kDa (*Breiteneder et al*., 2008). Allergenicity in foods can be caused by a variety of proteins with comparable allergenic capacity or by a single dominating allergen. Many food allergens are characterized by their high stability in the presence of heat, low pH, and enzyme digestion and their ability to survive food processing. Allergens are given a designation based on the Latin name of the species from which they originate and composed of the first three letters of the genus, followed by the first letter of the species and finishing with an Arabic number (*King et al.*, 1995). Thus, an allergen from *Mallus domesticus* (apple) is prefaced Mal d followed by a number, which is largely determined by the order in which allergens are identified. For those species where the first three letters of a genus and the first letter of a species are identical, the second letter of the species is also used. Many proteins are post-translationally modified with glycans and such structures can bind IgE, glycan-reactive IgE being found in between 16% and 55% of food-allergic patients (*Mills et al*., 2012). The last 10 years have seen an explosion in the number of allergenic proteins described from a vast array of foods, which has allowed the application of various bioinformatic tools to classify them according to their structure and function into protein families. The majority of allergens fell into around three to 12 families, the remaining allergens belonging to around 14–23 families comprising one to three allergens in each. Thus, around 65% of plant food allergens belonged to just four protein families, known as the prolamin, cupin, Bet v 1 and profilin superfamilies, whereas animal-derived food allergens fall into just three main families, namely the tropomyosins, EF-hand and caseins.

2.2 ALLERGENIC PROTEINS IN THE MAIN FOOD ALLERGENS

CRUSTACEANS, FISH AND MOLLUSCS

The muscle protein parvalbumin from the meat (muscle) is the dominant allergen (*Mills et al*., 2012) responsible for approximately 95% of the allergic incidents associated with fish. In aquatic arthropods (shrimp, crab, lobster and crayfish) the major allergen is the muscle protein tropomyosin which has been identified in 21 different crustaceans to date and is responsible for about 80% of all shrimp-related allergic incidents. A potential cross-reactive allergen, the arginine kinase, has been identified in Tiger shrimp (*Penaeus monodon*, Pen m 2) and Pacific white shrimp (*Litopenaeus vannamei*, Lit v 2). Myosin (Lit v 3) and sarcoplasmatic Ca_2-binding protein (Lit v 4) are considered minor allergens. The muscle protein tropomyosin from molluscs is the most important molluscan allergen and has been identified in several species (shells, clams, mussels, oysters, scallops, squids, octopus, abalones, and cuttlefish).

EGGS AND MILK

The allergens α-lactalbumin (Bos d 4), β-lactoglobulin (Bos d 5), serum albumin (Bos d 6), and immunoglobulin (Bos d 7) have been found in bovine (*Bos domesticus*) whey (see Chapter 3). Allergenic caseins (αS1, αS2, β, and κ) have been identified in milk from cows (Bos d 8), goats (*Capra aegagrus hircus*, Cap h casein), sheep (*Ovis aries*, Ovi a casein), and the domestic water buffalo (*Bubalus bubalis*, Bub b casein). The most important allergens in hen (*Gallus domesticus*) egg white are ovomucoid (Gal d 1), ovalbumin (Gal d 2), ovotransferrin (Gal d 3), lysozyme (Gal d 4), and serum albumin (Gal d 5). In hen egg yolk, the allergens are phosvitin (Gal d 6) and apovitellin VI (Gal d Apo VI).

NUTS AND PEANUTS

The main nut allergens belong mostly to the non–pollen-related seed storage proteins such as the 2S-albumins (Jug r 1, Ber e 1, Ana o 3, Cor a 14, and Pru du 2S albumin). The 11S globulins (Cor a 9 and Ana o 2), 7S globulins (Cor a 11, Jug r 2, Ana o 1, and Pru du 6), and lipid transfer proteins (LTPs: Cor a 8, Jug r 3, Pru du 3, and Pru du 4) also are important allergen families in nuts. These allergens are found in several nut-producing plants, including hazelnut (*Corylus avellana*), walnut (*Juglans regia*), almond (*Prunus dulcis*), Brazil nut (*Bertholletia excelsa*), Cashew nut (*Anacardium occidentale*), Macadamia nut (*Macadamia integrifolia*), and Pecan nut (*Carya illinoinensis*). The majority of the identified allergens in peanut (*Arachis hypogeae*) belong to the same protein families as those found in nuts. Major peanut allergens are the 11S globulin (Ara h 3/4), the 7S globulin (Ara h 1), and the 2S albumins (Ara h 2, Ara h 6, Ara h 7, and Ara h 9).

LUPIN AND SOYBEAN

The most important allergens in soybean (*Glycine max*) belong to the same allergen families as do their homologues in the other allergenic legumes, peanut and lupin. Soybean 11S globulin (Gly m 6), 7S globulin (Gly m 5), 2S albumin (Gly m 2S albumin) and Gly m LTP are stable seed storage allergens, whereas soybean profilin (Gly m 3) is degraded more easily during processing. In lupin, seed storage proteins highly homologous to other leguminous allergens are responsible for the majority of allergenic reactions from this plant. In addition, the lupin 11S and 7S globulins, 2S show considerable allergenicity from sweet lupines.

CELERY, MUSTARD AND SESAME

A few allergens have been characterized in celery (*Apium graveolens*), each belonging to a different protein family (see Chapter 4). The heat-labile pollinosis-associated protein (Api g 1) and profilin (Api g 4) are important for pollen related cross-reactivity, whereas celery LTP (Api g 2) and glycoprotein (Api g 5) are more important in processed foods. Seed storage proteins are the most important allergens in mustard. Sesame allergens (*Sesamum indicum*) come from 7S globulin (Ses i 3), LTP (Ses i LTP), oleosins (Ses i 4 and Ses i 5), and profilin (Ses i 8).

BUCKWHEAT, CEREALS AND WHEAT

Buckwheat (*Fagopyrum esculentum*) is a crop plant commonly used in Asia. The seed storage 2S albumins (Fag e 2 and BWp16) that account for 30% of the total protein content and the 13S globulin (Fag e 1) have considerable allergenic potential. In wheat (*Triticum aestivum*), the storage protein gluten with its ethanol-soluble gliadin is a causative agent of non–IgE-mediated celiac disease, which is an autoimmune disorder. Other cereals such as rye, barley, and durum wheat contain gliadin homologues. The control of gluten-free products is important because of the relatively high prevalence of this disease in the total population. In contrast, cereal-related cases of food allergy, most frequently wheat allergy, occur predominantly in children. Wheat grain proteins can be classified into water-soluble albumins, salt-soluble globulins, ethanol-soluble prolamins, and acid- or alkali-soluble glutenins. A number of different allergens have been identified, e.g., the prolamins gliadin (Tri a 19), glutenin (Tri a 26), and LTP (Tri a 14) and the pollen-related profilin (Tri a 12).

2.3 COMMON PROPERTIES OF ANIMAL FOOD ALLERGENS

TROPOMYOSINS

Tropomyosins are a family of closely related proteins present in muscle and non-muscle cells. Together with actin and myosin, tropomyosins play a key regulatory role in muscle contraction. Tropomyosins form head-to-tail polymers along the length of an actin filament and are the major allergens of two invertebrate groups, *Crustacea* and *Mollusca*, that are generally referred to as shellfish. Shrimp, crab, squid, and abalone are assumed to be largely responsible for seafood allergies. Tropomyosins were originally identified as major shrimp allergens by several laboratories and today they are recognized as invertebrate pan-allergens. Allergenic tropomyosins are heat stable and cross-reactive between the various crustacean and mollusc species.

PARVALBUMINS

The second largest animal food allergen family are the parvalbumins. Abundant in the white muscle of many fish species, paravalbumins are characterized by the possession of a widely found calcium-binding domain which is known as the “EF-hand”. Parvalbumins with bound calcium ions possess a remarkable stability to denaturation by heat. The ability to act as major fish allergens is obviously linked to the stability of parvalbumins to heat, denaturing chemicals, and proteolytic enzymes. Parvalbumins can be subdivided into two distinct evolutionary lineages, the α- and the β-parvalbumins, although their overall architectures are very similar. The β-parvalbumins are generally allergenic.

CASEINS

Structurally mobile proteins, they are found in mammalian milk at a concentration of around 15 mg/ml and are responsible for binding calcium through clusters of phosphoserine and/or phosphothreonine residues in αS1, αS2, and β-caseins although the αS2-casein gene is not expressed in man Caseins are a major food allergen involved in cow's milk allergy, which affects predominantly young children.

2.3.1 MINOR FAMILIES

LIPOCALINS

The lipocalins are a group of diverse proteins sharing about 20% sequence identity with conserved three-dimensional structures characterized by a central tunnel which can often accommodate a diversity of lipophilic ligands. They are thought to function as carriers of odorants, steroids, lipids, and pheromones amongst others although the only lipocalin which acts as a food allergen is the cow's milk allergen, β-lactoglobulin.

LYSOZYME FAMILY

Lysozyme type C and α-lactalbumins belong to the glycoside hydrolase family 22 clan of the O-glycosyl hydrolase superfamily and have probably evolved from a common ancestral protein. Two food allergens belong to this clan, the minor hen's egg allergen, lysozyme (Gal d 4) and the minor cow's milk allergen, α-lactalbumin, these proteins share little sequence homology, but are superimposable three-dimensional structures.

TRANSFERRIN FAMILY

Transferrins are eukaryotic sulfur-rich iron-binding glycoproteins which function in vivo to control the level of free iron. Members of the family that have been identified as minor food allergens include milk lactotransferrin (lactoferrin) and hen egg white ovotransferrin.

SERPINS

Serpins are a class of serine protease inhibitors and are found in all types of organisms with the exception of fungi and are involved in a variety of physiological processes. Only one food allergen has been identified as belonging to this family, the hen's egg allergen ovalbumin.

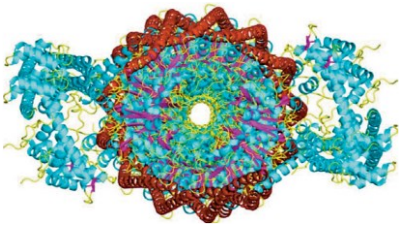
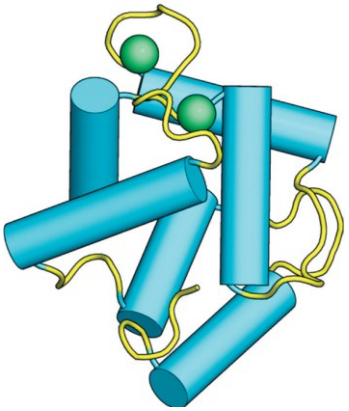
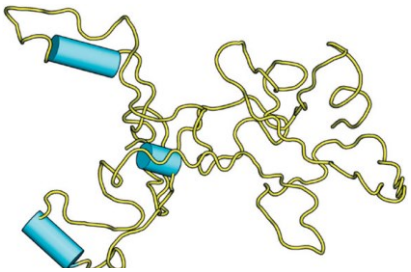
ARGININE KINASES

Arginine kinases have been identified as allergens in invertebrates including food allergens such as in shrimp and as cross-reactive allergens in the Indian meal moth and king prawn, lobster, and mussel.

OVOMUCOIDS

Kazal inhibitors which inhibit a number of serine proteases belong to a family of proteins that includes pancreatic secretory trypsin inhibitor, avian ovomucoid, and elastase inhibitor. These proteins contain between 1 and 7 Kazaltype inhibitor repeats. Avian ovomucoids contain three Kazal-like inhibitory domains. Chicken ovomucoid has been shown to be the dominant hen's egg white allergen Gal d 1.

Table 1: Major allergens of animal origin. 3D images are from Allergome®

Allergen family	Source	Allergen name	Sequence accession	Protein 3D pattern example
Tropomyosin superfamily	Brown shrimp (Farfantepenaeus aztecus)	Pen a 1	AAZ76743	
	Greasy backed shrimp (*Metapenaeus ensis*)	Met e 1	Q25456	
	Black tiger shrimp (*Penaeus monodon*)	Pen m 1	ADM34184.1	
	Indian prawn (*Fenneropenaeus indicus*)	Pen i 1	Peptides only	
	Snail (*Helix aspersa*)	Hel as 1	O97192	
	Squid (*Todarodes pacificus*)	Tod p 1	Peptides only	Cross-sectional view of tropomyosin (shown in red). Other proteins are troponin and actin. α-Helices and loops are shown in cyan and yellow, respectively.
	Oyster (*Crassostrea gigas*)	Cra g 1 Cra g 2 Cra g 1.03	Q95WY0	
	Crab (*Charybdis feriatus*)	Cha f 1	Q9N2R3	
	Abalone (*Haliotis diversicolor*)	Hal d 1	Q9GZ71	
Parvalbumin superfamily	Cod (*Gadus morhua*)	Gad c 1	Q90YK9 Q90YL0	
	Carp (*Cyprinus carpio*)	Cyp c 1.01 Cyp c 1.02	Q8UUS3 Q8UUS2	
	Salmon (*Salmo salar*)	Sal s 1.01 Sal s 1.02	Q91482 Q91483	
	Tuna (*Thunnus tonggol*)	Thu o 1.01 Thu o 1.02	None	
	Edible frog (*Rana esculenta*)	Ran e 1	Q8JIU2 P02627	
		Ran e 2	Q8JIU1 P02617	
				Calcium-liganded parvalbumin with two calcium-binding sites (α-helix) around the calcium cation.
Caseins	Domestic cow (*Bos taurus*)	Bos d 8		
		αs1*n*	P02662	
		αs2	P02663	
		β-casein	P02666	
		κ-casein	P02668	
	Goat (*Capra hircus*)	αs1*n*	P18626	
		αs2	P33049	
		β-casein	P33048	
		κ-casein	P02670	
	Sheep (*Ovis aries*)	αs1*n*	P04653	Bovine β-casein: α-Helices and loops are shown in cyan and yellow, respectively.
		αs2	P04654	
		β-casein	P11839	
		κ-casein	P02669	

2.4 COMMON PROPERTIES OF PLANT FOOD ALLERGENS

THE PROLAMIN SUPERFAMILY

These included two types of cereal seed proteins, namely the sulfur-rich prolamins and the α-amylase/ trypsin inhibitors of monocotyledonous cereal seeds, together with the 2S storage albumins found in a variety of dicotyledonous seeds including castor bean and oilseed rape. Subsequently other low molecular weight (LMW) allergenic proteins have been identified as belonging to this superfamily including soybean hydrophobic protein, non-specific lipid-transfer proteins (nsLTPs) and α-globulins. Compact proteins, the disulfide bonds in the prolamin superfamily members, are responsible for maintaining the three-dimensional structure even after heating, which is associated with retaining their allergenic properties after cooking. Their structure and IgE-binding properties are only being altered if severe heating results in hydrolysis of these bonds. These same structural attributes underlie their resistance to proteolysis, with several members, including the 2S albumins and nsLTP allergens being highly resistant to gastric and duodenal digestion.

CEREAL PROLAMINS

They comprise around half of the protein found in grain from the related cereals, wheat, barley, and rye, those from wheat being able to form large disulfide linked polymers which comprise the viscoleastic protein fraction known as gluten. These proteins are characteristically insoluble in dilute salt solutions, either in the native state or after reduction of inter-chain disulfide bonds, being soluble instead in aqueous alcohols. In addition to their role in triggering celiac disease, several types of cereal storage prolamins have been identified as triggering IgE-mediated allergies including γ-, α-, and ω-5 gliadins in addition to the polymeric HMW and LMW subunits of glutenin. Cooking appears to affect allergenicity and one study suggested that baking may be essential for allergenicity of cereal prolamins.

BIFUNCTIONAL INHIBITORS

The other group of prolamin superfamily allergens unique to cereals are the α-amylase/trypsin inhibitors which have been found to sensitize individuals via the lungs resulting in occupational allergies to wheat flour such as Baker's asthma or via the gastrointestinal tract for cereal-containing foods including wheat, barley, and rice. The individual subunits are either inactive or inhibitory to trypsin (and sometimes other proteinases), α-amylases from insects (including pests) or both enzymes (i.e. the inhibitors are bifunctional). The best characterized allergens are the α-amylase inhibitors of rice grain. Allergens with Mr of 16,000 have also been characterized in corn and beer (originating from barley) which appear to belong to the α-amylase inhibitor family

2S ALBUMINS

The 2S albumins are a major family of storage proteins and appear to be restricted to seeds of dicotyledonous plants where they may accompany the cupin globulin seed storage proteins 2S albumins can act as both occupational (sensitizing through inhalation of dusts) and food allergens, having been identified as the major allergenic components of many foods including the peanut allergens Ara h 2, 6, and 7, oriental and yellow mustard allergens Bra j 1 and Sin a 1, the walnut allergen Jug r 1, Ses i 1 and 2 from sesame, Ber e 1 from Brazil nut, and 2S albumins from almond and sunflower seeds. There is also some evidence that the 2S albumins of soy and chickpea are also allergenic.

NON-SPECIFIC LIPID-TRANSFER PROTEINS

One of the most important groups of allergens to have been identified in the last decade are the nsLTPs which appear to be involved in severe allergies to fresh fruits such as peach in the south of Europe around the Mediterranean. They have been termed as "pan-allergens" and are the most widely distributed type of prolamin being found in a variety of plant organs including seeds, fruits, and vegetative tissues. Thus, in addition to being identified in many different fruits and seeds, they have also been characterized in pollen of plant species such as olive and *Parietaria judaica*. nsLTPs as major allergens have been identified in fruits such as Pru p 3 in peach, Mal d 3 in apple, and Vit v 1 in grape.

Allergenic nsLTPs have also been characterized in vegetables such as asparagus, cereals such as maize, and in a number of nuts including hazelnut.

THE CUPIN SUPERFAMILY

The cupins are a functionally diverse protein superfamily which has probably evolved from a prokaryotic ancestor but has not found its way into the animal kingdom. They possess a characteristic β-barrel structure, the name "cupin" being derived from Latin for barrel. The 11S–12S globulins are found in the seeds of many monocotyledonous and dicotyledonous plants with homologs having been identified in gymnosperms (including conifers). They are sometimes termed legumins because they are particularly found in legume seeds and are oligomers of Mr 300,000–450,000. The subunits are again the products of a multigene family and also undergo proteolytic processing and glycosylation the extent of which varies depending on the plant species. Major allergens include the 7S and 11S globulins of soybean, Ara h 1 and Ara h 3 of peanut, Ana c 1 and Ana c 2 of cashew nut, the 7S globulins Jug r 2 of walnut and Len c 1 of lentil, and the 7S globulins of sesame and hazelnut. The 11S globulins have also been shown to be allergens in almond, also known as almond major protein (AMP), and in hazelnut. In general, these vicilin- and legumin-like seed globulins exhibit a high degree of thermostability, requiring temperatures in excess of 70 °C for denaturation, and have a propensity to form large aggregates on heating which still retain, to a large degree, their native secondary structure. Since the globulins are partially or fully insoluble between pH 3.5 and 6.5 it is likely that only limited solubilisation of globulins would occur when they enter the stomach. However, the 7S globulins seem to be highly susceptible to pepsinolysis, although several LMW polypeptides seem to persist following digestion of the peanut 7S globulin allergen Ara h 1, and there is evidence that they still possess IgE-binding sites following proteolysis (*Breiteneder et al.,* 2008). Similarly in vitro simulated gastrointestinal digestion results in rapid and almost complete degradation of the protein to relatively small polypeptides although these retain their allergenic properties.

THE BET V 1 FAMILY

The association of plant food allergies with birch pollen allergy is the most frequently observed of the cross-reactivity syndromes. The clinical symptoms of the birch pollen allergy-related OAS are caused by cross-reactive IgE between the major birch pollen allergen Bet v 1 and its homologs in a wide range of fruits and vegetables, including apple, celery, peanut, mung bean, sharon fruit, and even jackfruit.

2.4.1 MINOR FAMILIES

CLASS I CHITINASES

Chitinases are enzymes that catalyze the hydrolysis of chitin polymers. Chitinases are members of the glycoside hydrolase families 18 or 19. Endochitinases from plants belong to 19 (also known as classes IA or I and IB or II) and are able to degrade chitin, a major structural component of the exoskeleton of insects and of the cell walls of many pathogenic fungi. Class I chitinases from fruits such as avocado, banana, and chestnut have been identified as major allergens that cross-react with Hev b 6.02. Pers a 1, an allergenic class I chitinase from avocado, was extensively degraded when subjected to simulated gastric fluid digestion.

CYSTEINE PROTEASE SUPERFAMILY

Cysteine proteases of the C1, or papain-like, family were originally characterized by having a cysteine residue as part of their catalytic site, which has now been extended to include conserved glutamine, cysteine, histidine, and asparagine residues. Two major food allergens belong to this family, actinidin (Act c 1) from kiwi fruit and an allergen involved in soybean-induced atopic dermatitis known as Gly m Bd 30K, Gly m 1, or P34.

PROFILINS

Profilins from higher plants constitute a family of highly conserved proteins with sequence identities of at least 75% even between members of distantly related organisms. Profilins are cytosolic proteins of 12–15 kDa in size that are found in all eukaryotic cells. Profilins bind to monomeric actin and a number of other proteins, thus regulating the actin polymerization and depolymerization during processes such as cell movement, cytokinesis, and signaling. Structures of three plant profilins have been elucidated so far, those from *Arabidopsis thaliana* pollen, birch pollen, and *Hevea brasiliensis* latex. Since profilin-specific IgE cross-reacts with homologs from virtually every plant source, sensitization to these allergens has been considered a risk factor for allergic reactions to multiple pollen sources and for pollen-associated food allergy.

PROTEASE INHIBITORS AND LECTINS

The Kunitz/bovine pancreatic trypsin inhibitor family is active against serine, thiol, aspartic, and subtilisin proteases. They are generally small (50 residue) with three disulfide bonds constraining the proteins three-dimensional structure and belong to a superfamily of structurally related proteins, which share no sequence similarity and that includes such diverse proteins as interleukin-1 proteins, heparin-binding growth factors (HBGF), and histactophilin. In plants they probably play a role in defense against pests and pathogens. Minor allergens have been identified belonging to the Kunitz inhibitor family in soybean and potato. It is thought that their stability to processing and digestion is important for their allergenic activity. In addition to agglutinin, a lectin found in peanut has been identified as a minor allergen.

THAUMATIN-LIKE PROTEINS

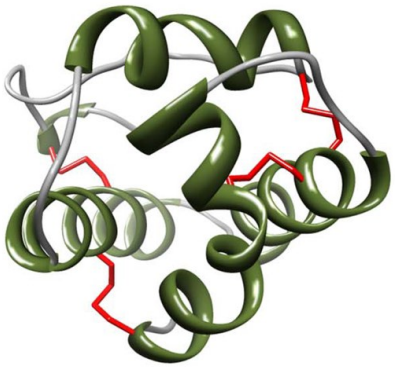
Thaumatin-like proteins (TLPs) derive their name from their sequence similarities to thaumatin, an intensely sweet tasting protein isolated from the fruits of the West African rain forest shrub *Thaumatococcus danielii*. Several allergenic TLPs from fruits have been described.

These include Mal d 2 from apple, Cap a 1 from bell pepper, Pru av 2 from sweet cherry, Act c 2 from kiwi, and an allergenic TLP from grape. The conformation of TLPs is stabilized by eight disulfide bonds.

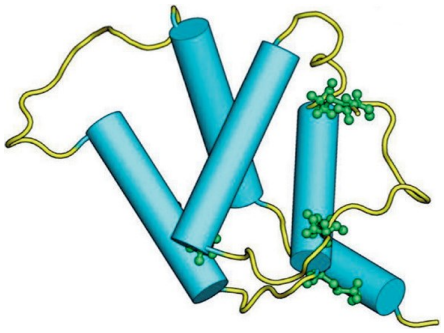
This extensive cross-linking confers high stability to proteolysis to the TLP scaffold and this is also the reason why the allergenic TLPs produced by grape berries persist during the entire vinification process and are among the major proteins present in wine.

Table 2: Major allergens of plant origin. 3D images are from Allergome®

Allergen family	Source	Allergen name	Sequence accession	Protein 3D pattern example
Prolamin superfamily				
Prolamins		α- and γ-gliadin	BAA12318 BAA11251 BAA23162	
		LMW glutenin		
Non-specific lipid-transfer proteins (nsLTPs)	Apple (*Malus domestica*)	Mal d 3	Q9M5X7	
	Apricot (*Prunus armeniaca*)	Pru ar 3	G7404406	
	Cherry (*Prunus avium*)	Pru av 3	AAF26449	
	Peach (*Prunus persica*)	Pru p 3 (originally Pru p 1)	P81402	
	Garden plum (*Prunus domestica*)	Pru d 3	P82534	
	Strawberry (*Fragaria ananassa*)	Fra a 3	Q4PLT5-9 Q4PLU0	
	Orange (*Citrus sinensis*)	Cit s 3	P84161	
	Grape (*Vitis vinifera*)	Vit v 1	P80274	
	Chestnut (*Castanea sativa*)	Cas s 8	*N*-terminus only	
	Walnut (*Juglans regia*)	Jug r 3	ACI47547.1	
	Asparagus (*Asparagus officinalis*)	Aspa o 1		
	Lettuce (*Lactuca sativa*)	Lac s 1	*N*-terminus only	
	Maize (*Zea mays*)	Zea m 14	P19656	
α-amylase/trypsin inhibitors	Barley (*Hordeum vulgaris*)	Hor v 1	CAA45085	
	Rice (*Oryza sativa*)	Rag 1, 2 5, 5.b, 14, 14b, 16, 17	Q01881 S59922 S59924 S59925 BAA01997 BAA01998 BAA01999 BAA02000	
2S albumins	Walnut (*Juglans regia*)	Jug r 1	JRU66866	
	Almond (*Prunus dulcis*)		P82944	
	Brazil nut (*Bertholletia excelsa*)	Ber e 1	CAA38362	
	Cashew nut (*Anacardium occidentale*)	Ana o 3	Q8H2B8	

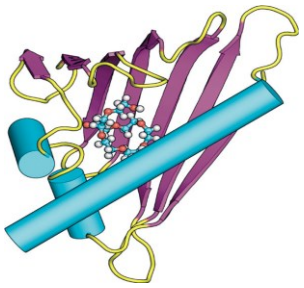


NsLTP allergen Pru p 3 from plum NsLTPs share a common fold that is composed of 4 a-helices (in green) and stabilized by 4 disulfide bonds (shown in red) to form a central tunnel for ligand interaction



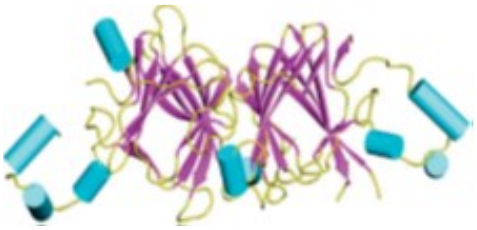
2S albumin from peanut: α-Helices and loops are shown in cyan and yellow, respectively. Disulfide bridges are shown in green ball-and-stick form.

	White mustard (*Sinapis alba*)	Sin a 1	P15322
	Black mustard (*Brassica juncea*)	Bra j 1	P80207
	Chickpea (*Cicer* arietinum)	None	None
	Peanut (*Arachis hypogaea*)	Ara h 2 Ara h 6 Ara h 7	L77197 AF091737 AF092846
	Sesame (*Sesamum indicum*)	Ses i 1 Ses i 2	AAD42943 Q9XHP1
	Sunflower (*Helianthus annuus*)	SFA-8	
Bet v 1 superfamily	Apple (*Malus domestica*)	Mal d 1	P43211
	Pear (*Pyrus communis*)	Pyr c 1	O65200
	Cherry (*Prunus avium*)	Pru av 1 (originally Pru a 1)	O24248 Q6QHU3 Q6QHU2
	Strawberry (*Fragaria ananassa*)		AY679601
	Carrot (*Daucus carota*)	Dau c 1	O04298
	Celery root (*Apium graveolens*)	Api g 1	P49372 P92918
Cupin superfamily			
7S (vicilin like) Globulins	Peanut (*Arachis hypogaea*)	Conarachin, Ara h 1	L34402
	Soy (*Glycine max*)	β-conglycinin	αP13916 α ‘: P11827 β:P25974
	Buckwheat (*Fagopyrum esculentum*)	BWI-1c BWI-2 BWI-2b BWI-2c	Q6QJL1
	Almond (*Prunus dulcis*)	Conglutin Gamma	P82952
	Walnut (*Juglans regia*)	Jug r 2	AAB41308



Carrot allergen Dau c 1 from the bet v 1 family of allergens. the structure is complexed with polyethylene glycol oligomer. α-helices are shown as cyan cylinders. single β-pleated sheets and loops are shown in magenta and yellow, respectively.

	Hazelnut (*Corylus avellana*)	Cor a 11	Q8S4P9
	Cashew nut (*Anacardium occidentale*)	Ana o 1.0101 1.0102	Q8L5L5 Q8L5L6
	Sesame (*Sesamum indicum*)	Ses i 3	Q9AUD0
11S (legumin like) Globulins	Peanut (*Arachis hypogaea*)	Arachin Ara h 3 Ara h 4	 AF093541 AF086821
	Soy (*Glycine max*)	Glycinin	Gy1 (A1aBx): P04776 Gy2 (A2B1a): P04405 Gy3 (AB): P11828 Gy4 (A4/5B3): P02858 Gy (A3B4): P04347
	Buckwheat (*Fagopyrum esculentum*)	BW24KD Fag e 1 FAGAG1 FA02 FA18	Q9XFM4 O23880 O23878
	Almond (*Prunus dulcis*)	Major almond protein, amandin, prunin	Q43607
	Cashew nut (*Anacardium* Brazil nut (*Bertholletia excelsa*)	Ana o 2	Q8GZP6
Cysteine protease C 1 family	Kiwi (*Actinidia deliciosa*)	Actinidin	Q43367
	Soy (*Glycine max*)	Gly m Bd 30K; P34; Gly m 1	P22895



"Chain B" of native soybean β-conglycinin trimer The structure consists of three chains, A, B and D. α-Helices are shown as cyan cylinders. β-Pleated sheets and loops are shown in magenta and yellow, respectively.

CHAPTER 3

CHARACTERIZATION OF MILK ALLERGENS AND THEIR CHANGES ARISING FROM SEVERAL CONDITIONING SYSTEMS

3.1 INTRODUCTION

Food allergy is an IgE-mediated abnormal response to normally tolerated food proteins. Cow's milk (CM) is the third most common food (after peanuts and tree nuts) causing anaphylactic reactions and one of the first and most common causes of food allergy in early childhood. The prevalence of cow's milk allergy (CMA) is increasing, and this may be explained by a decrease in breast-feeding and an increase in feeding babies with cow's milk-based formulas. In CMA the immediate and IgE-associated symptoms appear immediately or within 1–2 h after ingestion and affect the skin, the respiratory system and the gastrointestinal tract, often with severe effects. In addition, non- IgE-mediated mechanisms of cows' milk hypersensitivity (more common in adults) can cause other effects, from 2 hours to several days after CM consumption. The clinical symptoms of non Ig-E reactions mainly affect the gastrointestinal and respiratory systems. Cow's milk contains more than 25 different proteins, with around 30–35 g of proteins per litre, but only some of these are known to be allergenic. The casein fraction consists of four proteins: αS1-casein (Bos d 9), αS2-casein (Bos d 10), β-casein (Bos d 11) and k-casein (Bos d 12); αS1-casein is the most important allergen of the casein fraction. Allergens found in the whey fraction are α-lactalbumin (Bos d 4), β-lactoglobulin (Bos d 5), immunoglobulins (Bos d 7), bovine serum albumin (BSA, Bos d 6) and traces of lactoferrin (Bos d lactoferrin). The human IgE response to CM proteins is very variable and no single allergen or particular structure has been identified as being responsible for most allergenicity in milk, even if there is a higher prevalence of some of them *(Hochwallner et al., 2014)*. Several analytical approaches available for the detection of food allergens detect either the allergenic proteins or markers (e.g. specific proteins, peptides or DNA fragments) that indicate the presence of the

allergen. The enzyme-linked immunosorbent assay (ELISA) is a popular analytical method becasue it is easy to use, rapid and very sensitive, but it does not function when the antigen-antibody reaction is lost during heating and any other technological process that lead to changes in the protein structure. Mass spectrometry (MS) and liquid chromatography (LC) are increasingly used as reference techniques for the detection of allergens, with promising outcomes both for the characterization of allergens and for their quantification *(Kruse Faeste et al.,* 2011*)*. Other authors have described the urgent need to establish certain aspects of the analytical protocols in this field, e.g. the target analyte, the source of allergenic food (processed or unprocessed, before or after processing), the reporting units to be used (mass ratio of allergenic food, total proteins, allergenic protein target) and digestion conditions *(Pedreschi et al.,* 2012*)*. Furthermore, the variability of food allergens makes it difficult to develop a generic and universal method for multi-allergen detection and the multi-allergen quantification. There are many unresolved issues regarding this question.

During the second year research activities, several experiments were carried out on soft cream cheese under different conditioning systems, all at t days of refrigerated storage (5 °C). The goals of this period were: 1) to create a suitable platform for the detection and characterization of allergenic peptides using milk allergens and dairy food and 2) to analyse the effects of different conditioning systems on the allergenomic profile of cheese samples. The peptide pattern modification was monitored using nanoliquid chromatography and an electrospray ionization source coupled to an ion trap mass spectrometer (nanoHPLC-ESI-IT-MS). A specific bioinformatic approach was used, in order to characterize the most important food CM allergens and to describe the effects of different conditioning systems. This strategy was useful for selecting the best system to reduce the overall allergenic potential.

3.2 MATERIALS AND METHODS

3.2.1 CHEMICALS

Water, methanol, acetonitrile (LC–MS CHROMASOLV®, ≥ 99.9%), and formic acid (≥ 98%), were from Fluka (Steinheim, Germany); sodium hydroxide was from Merck (Darmstadt, Germany); trisodium citrate dehydrate (≥99%), standards of α-casein (≥ 70%), β-casein (≥ 98%) and k-casein (≥ 70%), bovine serum albumin (BSA, ≥ 98%) and 1,4-dithiothreitol (DTT) were purchased from Sigma Aldrich (St. Louis, USA); urea was purchased from Carlo Erba Reagenti (Rodano, IT). Soft cream cheese samples were provided by our Department's Food Technology group. The "allergenomics approach" was used to analyse a control group with no treatment (CNT) and seven different conditioning treatments at t days, all refrigerated at 5 °C,

The treatments were the following: CNT MAP (control sample with modified atmosphere), CNT MAP LIS MICRO (control with modified atmosphere and antimicrobial agents LIS and MICRO), L (antimicrobial agent LIS), M (antimicrobial agent MICRO), S1, S2, and S3 (antimicrobial agent SORB at three different concentrations). The hydrophilic peptides were extracted in duplicate from the samples following the Kunda's protocol *(Kunda, Benavente, & Català- Clariana, 2012)*. In particular, 2.5 g of the sample was suspended in 12.5 mL of a reduction buffer prepared with 73 mg of trisodiumcitrate dihydrate and 38 mg of DTT in 37.5 mL of 8 M urea. The pH value was adjusted to 8 with 1 M sodium hydroxide in a 50 mL Falcon tube (water was added to fill the total volume) and the samples were incubated for 1 h at room temperature. A 4,600 rpm centrifugation was performed for 30 min and the solution below the fat film was aspirated with a 10 mL syringe and filtered twice: through 0.45 μm nylon filters and then through 0.22 μm nylon filters (OlimPeak, Teknokroma). Solid phase extraction (SPE) cartridge TELOS C18 (6 mL/500 mg of sorbent, Kinesis) was preconditioned with 6 mL of methanol and 6 mL of water; 12 mL of the sample extract was loaded into the cartridge and the compounds were eluted with 1.2 mL of 80:20 (v/v) methanol:0.1% formic acid (FA) in water. The eluate was evaporated at room temperature under nitrogen flow.

The residue was reconstituted with 1.2 mL of water, then 1 mL was loaded through the SPE cartridge TELOS neo PRP (1 mL/30 mg of sorbent, Kinesis), previously conditioned with 1 mL of methanol and 1 mL of water, respectively. The cartridge was washed with 1 mL of 5% methanol, and finally the peptide fraction was eluted with 1 mL of 80:20 (v/v) methanol:0.1% formic acid in water and evaporated to dryness under nitrogen at room temperature. The final residue was dissolved in 1 mL of 0.1% FA and analysed or stored at – 30 °C. The same procedure was also performed on a mixture of casein standards for assuring that the sample preparation was specific for the peptides extraction.

3.2.2 LIQUID CHROMATOGRAPHY AND MASS SPECTROMETRY

Peptides were separated on a nano-LC reversed phase chromatography system. NanoLC-ESI-IT-MS/MS analysis were performed using a NanoLC Ultimate 3000 (Thermo Fisher Scientific Inc., US), equipped with an autosampler, a low pressure gradient micropump with flow managers, a column thermostat, a UV detector set at 214 nm and an ESI-Ion Trap HCTultra ETD II Basic System (Bruker Daltonics Srl, Bremen, Germany). Chromatographic separation used a PepMap C18 nanotrap precolumn (300 µm i.d. × 5mm, 5-µm particle size, LC Packings) and a PepMap™ C18 analytical column (15 cm length × 75 µm i.d., 3-µm particle size, 100-Å pore diameter; LC Packings). The Chromeleon™ (Dionex) and Hystar™ 2.3 (Bruker Daltonics) softwear packages were used for system management.

The mobile phase consisted of 0.1% FA in water (eluent A) and ACN 0.1% FA in water (80:20, v/v) (eluent B). The following gradient was used (flow rate 300 uL/min): 6 min isocratic step at 96% A and 4% B; 29 min linear gradient at 50% A and 50% B; 11 min isocratic step at 10% A and 90% B; 24 min with 96% A and 4% B, for reconditioning. A full-scan MS acquisition in the 300–1500 m/z range was performed with the acquisition of Base Peak Chromatogram (BPC) and Total Ion Current (TIC) profile by using a high capacity ion trap, coupled to a nano-ESI source.

The following parameters were set: positive ion mode, spray voltage 4.5 kV, sheath gas: nitrogen at 10 L/min, capillary voltage 1.5 V, heated capillary

temperature 160 °C. MS/MS analysis by collision-induced dissociation (CID) was performed using unattended data-dependant acquisition with the following modes: auto-MS/MS (scan range 100–3000 m/z), number of precursor ions: 3, absolute threshold of minimal signal required for precursor ion: 10,000. Extraction of mass spectra peak-lists, mass annotation and deconvolution were performed using Data Analysis™ 4.0. The acquired MS and MS/MS data were submitted to NCBI-nr and SwissProt database searches using the Biotools™ 3.2 software and the MASCOT® platform as a search engine (*Matrix Science, London, UK*) with the following parameters: category: "other mammalia"; maximum number missed cleavages: 3; MS/MS ion tolerance and peptide tolerance: 0.3 Da. Subsequent database searches were performed by setting pepsin, chimotrypsin or semi-trypsin as proteases. The oxidation of methionine and the phosphorylations of serine, threonine and tyrosine were chosen as variable modifications. A specific Bioinformatic approach was used for peptide identification, based on MASCOT scores higher than an arbitrary cut-off level of 48 in order to focus on the most significant peptides and minimize the risk of false-positives. The alignment of allergenic sequences was performed on the Allergome® database using the algorithm NCBI blastp (2.2.18) with the following parameters: identity (BLAST) and similarity (FASTA) from 90 % to 100 %; matrix BLOSUM62; gap costs Open:11 Ext:1; expect threshold 100 and word size: 3.

3.3 RESULTS AND DISCUSSIONS

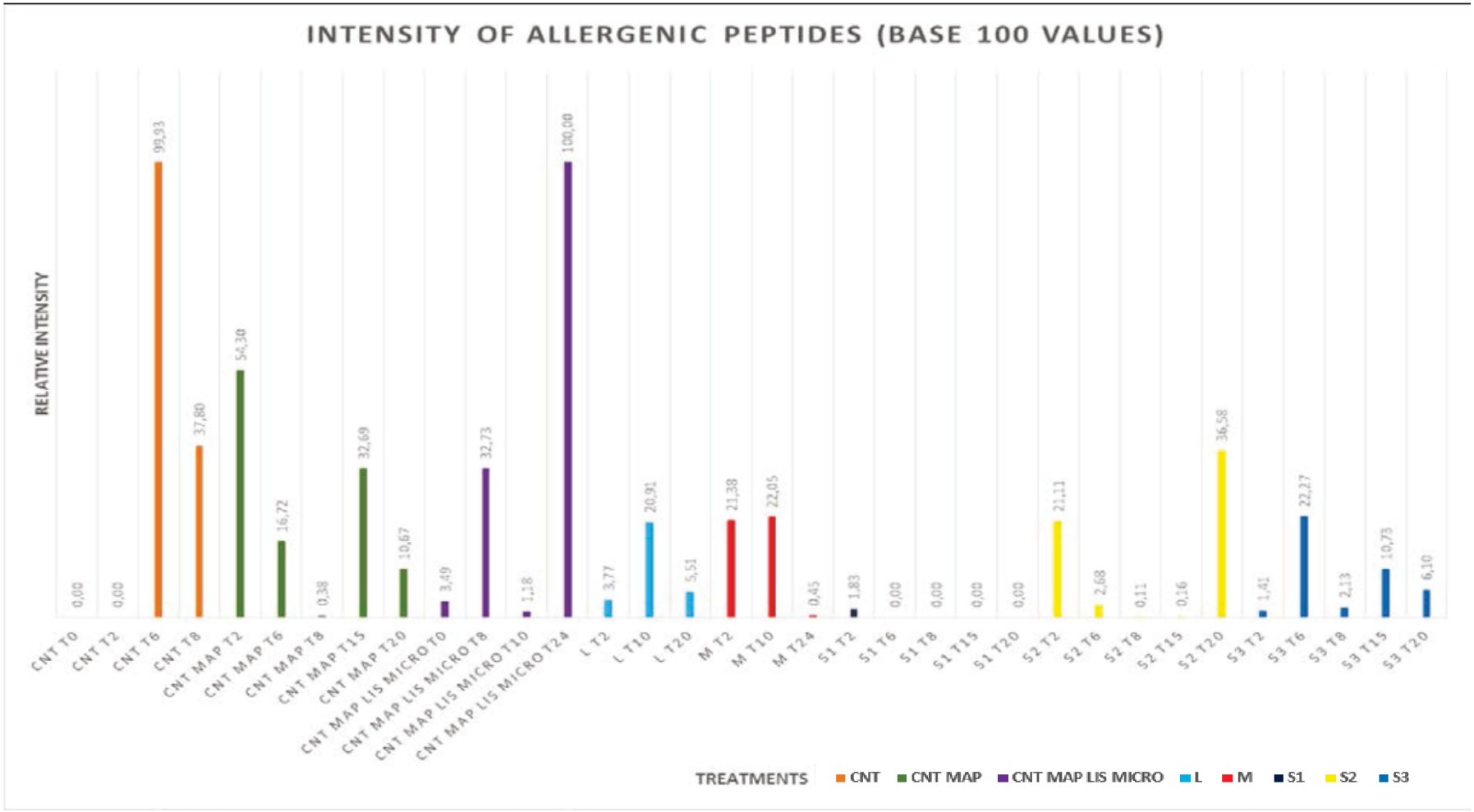
Monitoring involved six different packaging conditions and a control sample without treatment at t days of 5 °C refrigerated storage in order to find the optimal conditions for reducing the overall allergenic potential of soft cheese samples. The study has been carried on independently from other important aspects of conditioning, with the only purpose being to find out the effects of packaging on the allergenic pattern of this product. Protein degradation during the refrigerated storage led to 30 allergenic hydrophilic peptides, as described in Table 3. Most of the sequences (29) were generated from the casein fraction, and were related to the allergens Bos d 9, Bos d 10, Bos d 11 and Bos d 12. All allergens were differently distributed among the treatments. Only one sequence was related to the allergen Bos d 5 (beta-lactoglobulin).

Bos d 9 was the most prevalent allergen in the samples and it is also the allergen most involved in CMA, with a prevalence of 60-100% of allergenic patients *(Hochwallner et al.,* 2014*).* The total intensities of free allergenic peptides ranged from a minimum of 0 to a maximum of 553 x 10^6. Figure 6 shows the normalized intensities (base 100) of all systems during the days of monitoring.

Table 3: Detected peptides sequences and their distribution across the different days of conditioning

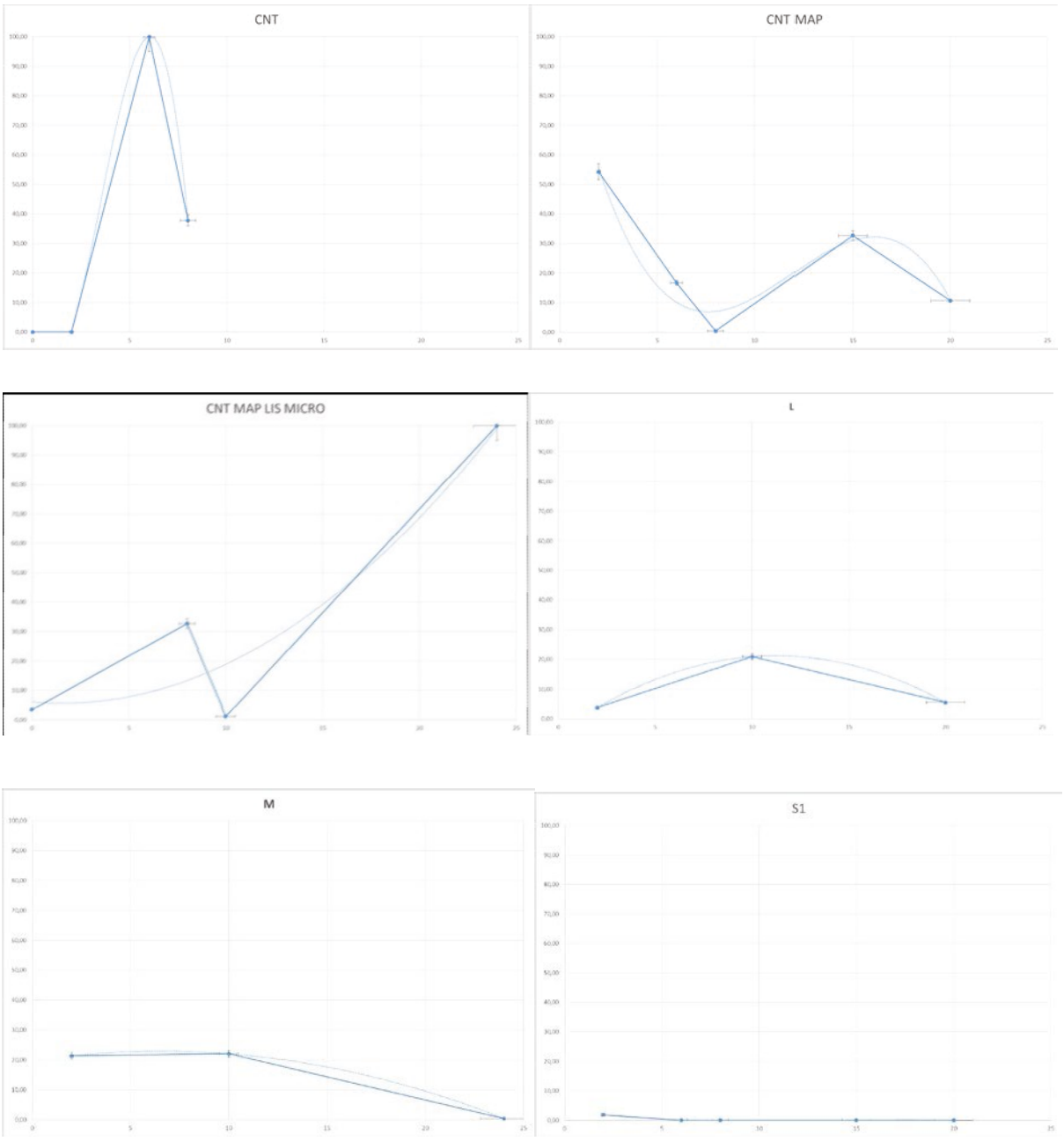
Allergen	Allergenic peptides	Meas. m/z	z	Treatments							
				Cnt	Cnt Map	Cnt Map Lis Micro	L	M	S1	S2	S3
Bos d 5	IIVTQTMKGLDIQKVAGTW	701,35	3+			t8		t2			
Bos d 9	FVAPFPEVF	526,710-526,770	2+	t6, t8	t2, t6, t8, t15	t8, t10, t24	t2, t10,	t2, t24		t2, t20	t6, t8, t15, t20
	NENLLRF	453,160-453,250	2+	t6	t2, t6, t15, t20	t0, t8, t10, t24	t2, t10, t20	t2, t10	t2	t2, t6, t20	t6, t8, t15, t20,
	GYLEQLLRL	552,740-552,850	2+	t8	t2, t15	t24	t10, t20	t2		t20,	t15, t20
	VAPFPEVF	453,23	2+	t8	t20		t10				
	KTTMPLW	438,660-438,730	2+		t2, t15, t20	t24	t20		t2	t8, t15, t20	
	RFFVAPFPEVF	678,35	2+			t24					
	LRFFVAPFPEVF	734,91	2+					t2			
Bos d 10	NSEKTTMPLW	603,650-603,780	2+		t20		t20				
	FVAPFPEVFGKEKVNEL	975,48	2+			t24					
Bos d 11	YQEPVLGPVRGPFPIIV	940,920-941,010	2+	t6	t2, t6	t10, t24	t2	t10		t2	t2, t15
	QEPVLGPVRGPFPIIV	859,370-859,490	2+	t6, t8	t2, t15, t20	t0, t8, t24	t10, t20	t2, t10		t20	t6, t8, t15, t20
	HKEMPFPKYPVEPF	873,360-873,370	2+	t6		t24					
	EMPFPKYPVEPF	740,800-740,830	2+			t8, t24	t10	t2, t10			
	KIEKFQSEE	609,200-609,290	2+		t15		t10	t2			t20
	TESQSLTL	439,690-439,720	2+	t8			t10, t20				
	GPVRGPFPIIV	576,27	2+		t20						
	LYQEPVLGPVRGPFPIIV	997,540	2+				t10				
	YQEPVLGPVRGPFPIL	891,450-891,470	2+			t0, t8, t10		t2			
	VLGPVRGPFPIIV	682,410-682,440	2+			t10		t10			
	FLLYQEPVLGPVRGPFPIIV	1127,63	2+					t2			
	QEPVLGPVRGPFPIL	809,94	2+					t2			
	HLPLPLLQSW	602,780	2+					t2			
	EMPFPKYPVEPFTE	855,88	2+					t2			
Bos d 12	TVQVTSTAV	453,160-453,250	2+		t20		t20	t2		t2, t6	t2, t6
	SRYPSYGLNYY	691,790-691,820	2+				t10	t2			
	ARHPHPHLSF	599,8	2+				t20	t2			
	INNQFLPYPY	634,8	2+				t10				
	SRYPSYGINYY	691,78	2+				t20				
	SRYPSYGLNY	610,3	2+				t10				

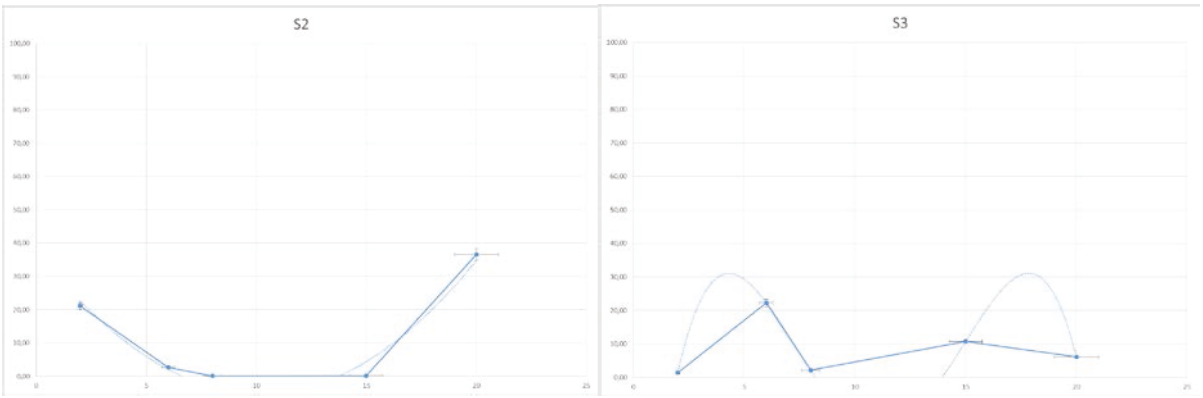
Figure 6: Relative mass intensities of Bos taurus allergens for each treatment at t day of 5 °C refrigerated storage



While allergenic potential is a function of the intensity of free allergenic peptides, it is evident that conditioning system S1, involving the use potassium sorbate, showed the lowest allergenicity. Peptides from dairy foods are generated by the action of microorganisms and microbial species, and sorbic acid is recognized as a safe food preservative which can inhibit the growth of some bacteria like *Listeria monocytogenes, Salmonella typhimurium, Escherichia coli 0157:H7* and some fungi in many low acid foods, like cold-packed cheeses (*Cruz-Romero et al, 2013).* On the contrary, the worst effect was observed in the treatment CNT MAP LIC MICRO, with an increase in proteolytic degradation, with values higher than the control series; the use of this option, as well as the other treatments CNT MAP, L and M should be discouraged for allergen-free food technology purposes. On the other hand, a less important effect was present in the samples from S2 and S3 treatments, with an unstable trend (Figure 7) and the generation of three different free allergens (Figure 8) compared to S1 (one sequence from Bos d 9 only). These values suggest that it is important to consider the effect of concentration of this additive in future food technology research. The samples past day 8 of control sample CNT were previously excluded because unacceptable for sensorial and microbiological reasons. A slight reduction was observed in treatment L that showed characteristic trend during shelf life, as shown in Figure 7.

Figure 7: Trends of allergenic relative intensities (y-values) in different packaging systems at t days (x-values) of refrigerated storage

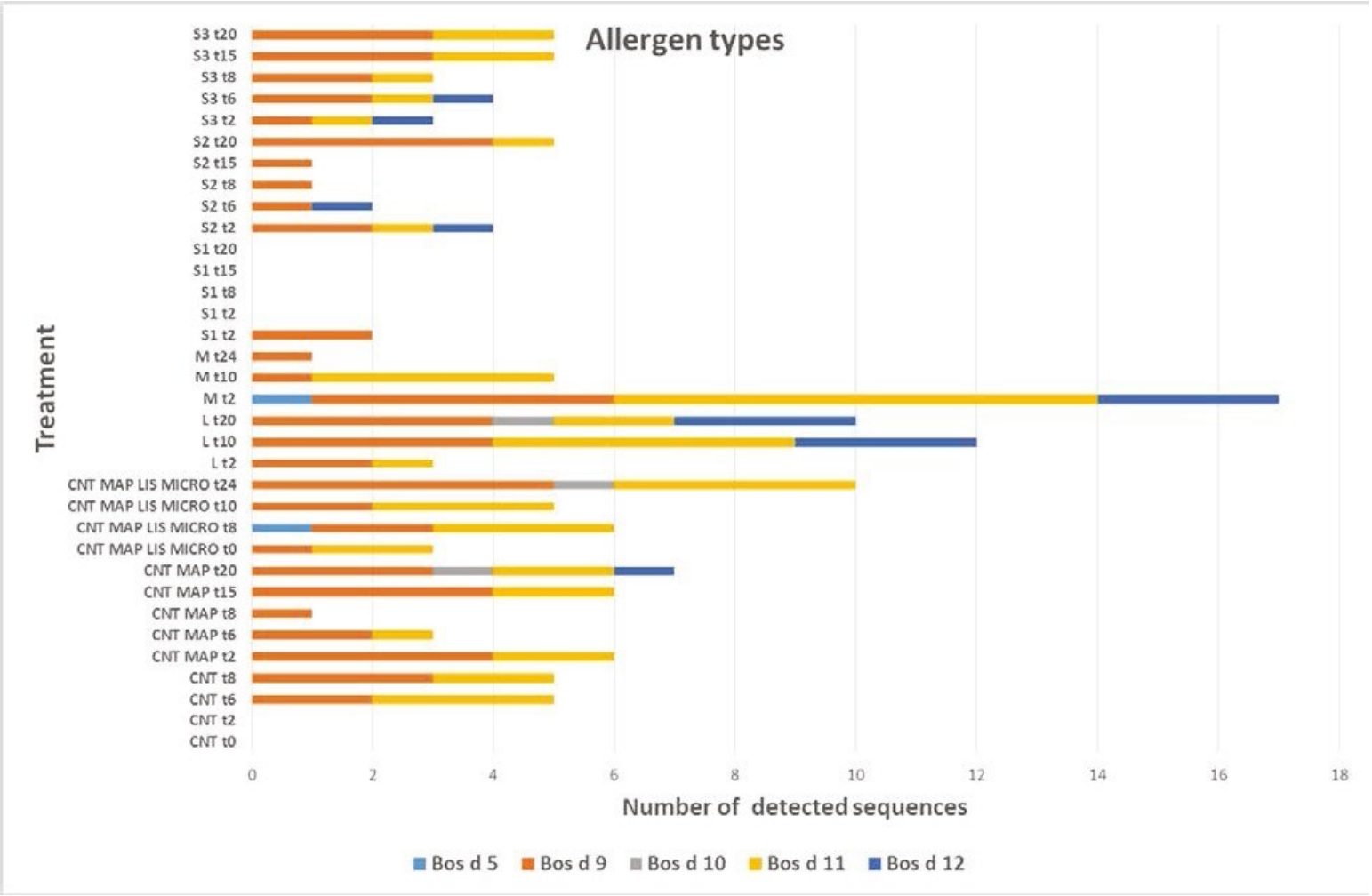




Besides the intensity values, another aspect that should be considered is the distribution of the different kinds of allergens. As described in the first section, CMA and food hypersensitivity in general, is a response often not given by only one protein or peptide, so the allergenic composition of a food plays a fundamental role in eliciting the reaction *(Hochwallner et al., 2014)*. Figure 8 shows the counts of all sequences found on the basis of their allergenic origin. As shown, more than three different free allergens were generated from samples o L, M and CNT MAP series; in particular sample M at t2 showed a total of 17 different allergenic peptides, including allergen Bos d 5 which was only generated in this samples CNT MAP LIS MICRO at day 8 and M at t2.

Peptides from allergen Bos d 10 were present only in samples CNT MAP at day 20, in CNT MAP LIS MICRO at day 24 and L at day 20. The most prevalent peptides were those from allergens Bos d 9 (68 sequences found), and those from allergens Bos d 11 (53 sequences). The best solution, not only for allergenic intensity but also for allergenic composition, was given by conditioning system S1.

Figure 8: Cow milk allergens found in each treatment at t days of refrigerated storage (numbers of sequences)



Allergenicity studies have shown the existence of numerous allergenic IgE and IgG binding epitopes along each of the four milk allergens. Furthermore, an effective LC-MS/MS strategy in the field of food allergy should be rapid and simple and focus only on the most relevant peptides. For these reasons, a last important aspect to consider is the presence of specific and very abundant peptides from the allergens Bos d 9 and Bos d 11. In particular, the peptides FVAPFPEVF from Bos d 9 and QEPVLGPVRGPFPIIV from Bos d 11 could be used as marker peptides in future analysis in this field, e.g. in selective ion monitoring experiments and for quantification purposes. The peptide FVAPFPEVF is comprised in the sequence FFVAPFPEVFGK, while the peptide QEPVLGPVRGPFPIIV incorporates the sequence GPFPIIV in hidden form. Both these peptides have been classified as "critical" peptides in the proteomic identification of CMA *(Ansari et al.,* 2011*)*. Overall, these data demonstrated the effect of different packaging conditions on reducing the total allergenic power of refrigerated soft cheese, and showed the condition treatment S1 as an optimal solution for reducing the total potential of soft cheese, both in terms of allergenic intensity and in terms of allergen variability. On the contrary, packaging solutions CNT MAP LIS MICRO, L and M should be excluded. The information about other treatments S2 and S3 can be useful for future food processing studies

3.4 CONCLUSIONS

In this part of research, a nano-HPLC MS/MS, coupled with a specific Bioinformatics approach, has been applied for the sensitive and selective detection of allergenic peptides in soft cheese samples. Furthermore, different packaging techniques were investigated in order to find the best solution for reducing the overall allergenic power of a complex processed food, demonstrating the applicability of targeted analyses using the Allergenomics approach as an effective strategy for food analysis. This investigation identified the hydrophilic peptides FVAPFPEVF and QEPVLGPVRGPFPIIV and proposed them as valid markers for the absolute detection of the two most important CMA Bos d 9 and Bos d 11. The best packaging conditions for allergenic purposes was found to be treatment S. In addition, a good platform has been created for the following activities of this PhD thesis.

CHAPTER 4:

CHARACTERIZATION OF *APIACEAE* ALLERGENS

4.1 INTRODUCTION

Apiaceae, (*Umbelliferae*, *nomen conservandum*) includes about 400 genera of plants distributed throughout a wide variety of habitats. Many species of *Apiaceae* are poisonous, including poison hemlock (*Conium maculatum*), water hemlock (*Cicuta maculata*), and fool's parsley (*Aethusa cynapium*). Other species, however, are widely used vegetables, including parsley (*Petroselinum crispum*), carrot (*Daucus carota*), celery (*Apium graveolens*), parsnip (*Pastinaca sativa*), and fennel (*Foeniculum vulgare*). Other *Apiaceae* species are used as herbs and spices, including anise (*Pimpinella anisum*), dill (*Anethum graveolens*), coriander (*Coriandrum sativum*), caraway (*Carum carvi*), and cumin (*Cuminum cyminum*). Allergies to *Apiaceae* are particularly important in Southern Italy (*Caiaffa et al*, 2011; *Asero et al*, 2009) although there is a lack of reference materials and methodologies for their absolute detection using liquid chromatography and mass spectrometry (LCMS). European Union law (*Reg. No. 1169*, 2011) requires food manufacturers to declare the presence of celery (and products therof) on labels of processed food in order to protect consumers' health.

Based on the data available from the International Union of Immunological Societies (*IUIS*, 2015) the most important allergens from carrot are the following: Dau c 1, a Bet v 1-homologue, and major allergen; Dau c 3, a lipid transfer protein, which is heat-stable; Dau c 4, a 12 kDa protein (a profilin) and Dau c cyclophilin, a 20 kDa protein. Allergens unique to fennel, which have already been characterized, are Foe v 1, a Bet v 1-related protein and Foe v 2, a profilin. The most important allergens from celery are: Api g 1, the major allergen, a 16 kDa protein and a Bet v 1 homologue; Api g 3, a chlorophyll Ab-binding protein; Api g 4, a 14.3 kDa protein, a profilin and a minor allergen, and Api g 5, a 60 kDa protein, isolated from the tuber, with homology to FAD-containing

oxidases. There is often an extensive cross-reactivity among the different individual species of *Apiaceae* and non-*Apiaceae* species, frequently manifested (*Halmepuro et al*, 1985). It is not known why a protein becomes an allergen, but there are several biochemical factors that evoke food allergy (FA): allergens are usually proteins (or peptides) with a molecular weight lower than 70 kD and they are soluble in water, with few exceptions like peanut oleosins (Schwager *et al*, 2015). In most cases they are glycosylated and resistant to denaturation and proteolysis, and they present linear epitopes and/or structures that interact with lipids (*Anpelkovic´ et al,* 2015; *Mills*, 2005). Most allergens are present in the foods themselve, while others are generated by several types of food processing (both thermal and not), with a carousel of reactions that can involve, among others, interaction by adjuvant compounds, like lipids or glycans (*Verhoeckx et al*, 2015).

Sample preparation is an essential step not only for allergen detection but also for reliable allergen quantification: if the allergen recovery is not complete, quantification is underestimated. Allergen extraction from plant tissues is very challenging because of the presence of interfering compounds like phenols, lipids and sugars. Several protocols have been adapted, according to food matrix. Food processing must also be taken into account in this step because, for example, some thermal treatments or sanitization processes have an important impact on allergen stability and, consequently, on extraction efficiency (*Sharma et al*, 2009). The first extraction method for plant allergens was defined by Björkstén in 1980 (*Björkstén et al,* 1980) and was based on potassium phosphate. Other more complex protocols are described by Pastorello (*Pastorello et al*, 2001) and other authors, and involve the use of several saline buffers, and/or precipitation reagents (*Rudolf et al*, 2012). In some cases the precipitation of proteins is performed using organic solvents, but the use of saline buffers or high salt buffers is often the preferred strategy (*L'Hocine & Pitre*, 2015).

During the third year's research, several experiments were carried out in order to achieve the following goals: 1) to find a valid method for the extraction of allergens from carrot, celery and fennel; 2) to apply the platform developed during the second year on plant samples, in order to characterize new or already known allergens, with particular attention to the allergens which are most important in Southern Italy; 3) to characterize one or more valid markers for the

reliable and rapid detection of *Apiaceae* allergens in food using LCMS.

Allergen extraction was performed with several saline buffers, as described in the experimental section.Where possible, particular attention was given to the selection of eco-compatible buffers, for green chemistry purposes. The overall allergenic pattern of the samples was monitored using nanoliquid chromatography and an electrospray ionization source coupled to an ion trap mass spectrometer (nanoHPLC-ESI-IT-MS). A specific bioinformatic strategy, basing on the platform developed in the second year was successfully applied to the results obtained from LCMS analysis.

4.2 MATERIALS AND METHODS

4.2.1 CHEMICAL REAGENTS

Water, methanol, acetonitrile (LC–MS CHROMASOLV®, ≥ 99.9%), and formic acid (reagent grade, ≥ 98%) were from Fluka (Steinheim, Germany); sodium carbonate anhydrous (reagent grade, ≥ 99.5%) was from Mallinckrodt Baker (Deventer, Holland); sodium hydroxide was from Merck (Darmstadt, Germany); albumin from bovine serum (electrophoresis grade ≥ 98.0%), 1,4-dithiothreitol (DTT), iodoacetamide (Bioultra), sodium bicarbonate (reagent grade, ≥ 99.5%), Bradford reagent, sodium phosphate dibasic (reagent grade, ≥ 99.95%), sodium phosphate monobasic (reagent grade, ≥ 99.0%), Trizma® hydrochloride (reagent grade, ≥ 99.9%) and potassium chloride (reagent grade, ≥ 99.0%) were purchased from Sigma Aldrich (St. Louis, USA); urea was purchased from Carlo Erba Reagenti (Rodano, IT); pepsin (sequencing grade), chymotrypsin (sequencing grade) and trypsin were purchased from Princeton Separations (Adelphia, USA). Celery, carrot and fennel samples were purchased from local stores.

4.2.2 ALLERGEN EXTRACTION, PROTEIN QUANTIFICATION

Proteins were extracted from the samples following three extraction protocols, using three different extraction buffers: 50 mM Tris HCl, pH 8.2 (TRIS), 100 mM Carbonate buffer, pH 8 (CARB) and 10 mM phosphate buffer, KCl 1 M, pH 4 (HSPB). For each of the three extraction protocols, four experimental replicates were performed. 150 g of washed whole plant sample was homogenized with around 250 mL of each buffer; the resulting suspension was grossly filtered in a filter funnel and left at 50 °C for 4 hours for protein extraction. The insoluble matter was removed by centrifugation at 6,000 rpm, at 25 °C for 1 hour. The supernatant from each sample was recovered and filtered through 0.45 μm nylon filters (OlimPeak, Teknokroma). The three protein samples were then quantified by Bradford assay using BSA as a standard (Table 1). The filtered solution was evaporated under nitrogen flow at 25 °C, then reconstituted for further digestion in 5 mL 50 mM Tris HCL, pH8.0

(for chymotrypsin and trypsin digestion) or in 5 mL 50 mM ammonium acetate, pH 4.5 (for pepsin digestion).

4.2.3 ENZYMATIC DIGESTION OF PROTEIN EXTRACTS

Three different proteolytic enzymes were used: pepsin, chymotrypsin and trypsin. 1 mL aliquots were reduced and alkylated using 0.1 M DTT and 0.1 M iodoacetamide. The enzyme solution (1 mg/ml in water) was added at a ratio of 1:50 (enzyme to protein, by weight). The sample was incubated overnight at 37 °C. The reaction was stopped by adding 5% formic acid.

4.2.4 SAMPLE PURIFICATION

The digested solution was loaded through the SPE cartridge TELOS neo PRP (1 mL/30 mg of sorbent, Kinesis), previously conditioned with 1 mL of methanol and 1 mL of water, respectively. The cartridge was washed with 1 mL of 5% methanol, and finally the peptide fraction was eluted with 1 mL of 80:20 (v/v) methanol:0.1% formic acid in water and evaporated to dryness under nitrogen at room temperature (around 25 °C). The final residue was dissolved in 1 mL of 0.1% FA and analyzed or stored at – 30 °C.

4.2.5 LIQUID CHROMATOGRAPHY AND MASS SPECTROMETRY

Peptides were separated on a nano-LC reversed phase chromatography system. NanoLC-ESI-IT-MS/MS analysis was performed using a NanoLC Ultimate 3000 (Thermo Fisher Scientific Inc., US), equipped with an autosampler, a low pressure gradient micropump with flow managers, a column thermostat, a UV detector set at 214 nm and an ESI-Ion Trap HCTultra ETD II Basic System (Bruker Daltonics Srl, Bremen, Germany). Chromatographic separation used a PepMap C18 nanotrap precolumn (300 µm i.d. × 5mm, 5-µm particle size, LC Packings) and a PepMap™ C18 nano column (15 cm length × 75 µm i.d., 3-µm particle size, 100-Å pore diameter; LC Packings). The Chromeleon™ (Dionex) and Hystar™ 2.3 (Bruker Daltonics) software packages were used for system management.

The mobile phase consisted of 0.1% FA in water (eluent A) and ACN 0.1% FA in water (80:20, v/v) (eluent B). The following gradient was used (flow rate 300 µL/min): 5 min isocratic step at 96% A and 4% B; 5 min gradient (curve 7) at 96% A and 4% B; 90 min gradient (curve 7) at 85% A and 15% B; 40 min gradient (curve 7) at 10% A and 90% B; 10 min with 96% A and 4% B, for reconditioning. A full-scan MS acquisition in the 300–1500 m/z range was performed with the acquisition of Base Peak Chromatogram (BPC) and Total Ion Current (TIC) profile by using a high capacity ion trap, coupled to a nano-ESI source. The following parameters were set: positive ion mode, spray voltage 4.5 kV, sheath gas: nitrogen at 10 L/min, capillary voltage 1.5 V, heated capillary temperature 160 °C. MS/MS analysis by collision-induced dissociation (CID) was performed using unattended data- dependent acquisition with the following modes: auto-MS/MS (scan range 100–3000 m/z), number of precursor ions: 3, absolute threshold of minimal signal required for precursor ion: 10,000. Extraction of mass spectra peak-lists, mass annotation and deconvolution were performed using Data Analysis™ 4.0. The acquired MS and MS/MS data were submitted to NCBI-nr and SwissProt database searches using the Biotools™ 3.2 software and the MASCOT® platform as a search engine (Matrix Science, London, UK) with the following parameters: category: "*Viridiplantae* (Green Plants)"; maximum number missed cleavages: 3; MS/MS ion tolerance and peptide tolerance: 0.6 Da. Subsequent database searches were performed by setting pepsin, chymotrypsin or trypsin as proteases. The oxidation of methionine was chosen as variable modification. A specific bioinformatic approach was used for peptide identification, based on MASCOT scores higher than an arbitrary cut-off level of 20 in order to focus on the most significant peptides and minimize the risk of false-positives. The alignment of allergenic sequences was performed on the Allergome® database using the algorithm NCBI blastp (2.2.18) with the following parameters: identity (BLAST) and similarity (FASTA) from 90 % to 100 %; matrix BLOSUM62; gap costs Open:11 Ext:1; expect threshold 100 and word size: 3.

4.3 RESULTS AND DISCUSSION

4.3.1 SAMPLE TREATMENT, OPTIMIZATION OF EXTRACTION CONDITIONS AND PROTEIN CONTENT IN ALLERGENIC EXTRACTS.

Three different allergenic foods from the *Apiaceae* family were chosen for this study: fennel, celery and carrot. The incidence of allergies to these foods is increasing in the Mediterranean Diet *(Caiaffa et al* , 2011). Allergy to celery is one of the most common food allergies in Europe, causing digestive disorders, respiratory distress, and skin reactions when ingested. The known allergens in celery are the 16 kD heat-labile protein Api g1, which is homologous to the profilin Api g 4, the glycoprotein Api g 5, and the protein Bet v 1 (*Mari*, 2015). Several methods are reported which detect the presence of celery allergens in food, mostly based on DNA detection (*Zahradnik et al*, 2014; *Fuchs et al*, 2013) but no sufficient characterization data are available at the present. Allergen extraction requires choice of the proper buffer system able to efficiently solubilize food sample allergenic proteins. The isolation of allergens is the most critical step in allergenomic studies, making the discipline more complex than studies on other proteins. Plant allergens are present in relatively low concentrations and are a highly heterogeneous population because of their functionality and variability. One of the biggest obstacles for allergenomic analysis (and for proteomics in general) is the presence of highly abundant proteins, which may include up to 40% of protein in plant tissue. In an optimistic scenario, the presence of interfering compounds in plant cells, such as multiple protease, polyphenols, tannins, pigments, waxes and carbohydrates, already reduces recovery of the proteome by about 25% (*Wienkoop et al*, 2014). There is no extraction protocol capable of capturing the entire proteome and consequently, a carousel of extractive methods is present in the literature that involve the use of single or multiple solvents, buffers, permutations, physical treatments and so on. In general, the protein extraction requires initial destruction of the tissue, cell lysis induced mechanically or by sonication, enzymes or solvents and the dissolution of the proteins in a buffer extractant as similar as possible to the cellular environment

of origin (respect to pH, polarity, ionic strength) (*Westphal et al*, 2004).

For this purpose, three buffer solutions were evaluated among those mostly used for allergen extraction (*Pastorello et al*, 2001), and chosen in terms of the total protein extraction yield, the number of found allergens and their ecological compatibility. Three allergenic foods were extracted with the following buffers of different pH and ionic strengths, as described in the experimental section: TRIS-HCl 50 mM, pH 8,2 (Tris); Carbonate buffer 100 mM, pH 8 (Carb), and phosphate buffer 10 mM, KCl 1 M, pH 4 (HSPB). Extraction was performed (see experimental section) after physical disruption, for three hours in a thermic bath set at 50 °C. The best conditions for the efficient extraction of total proteins from fennel, celery and carrot were chosen by using the Bradford protein assay (*Bradford*, 1976) as a discriminating factor. Thus, in a second step, the best extraction conditions were chosen according to the highest numbers of allergenic peptides found in each extract. Table 4 summarizes the protein content found in the different extracts. The type of extraction buffer used had a significant effect on the amount of protein extracted from the samples. Regarding total protein content, it was surprising to find that the high salt phosphate buffer (HSPB) extracted significantly less protein than Tris and Carb, since the phosphate buffer is known to be widely effective in allergens extractions (*Rudolf*, 2012); significant differences were also found between the protein amounts extracted by these last two solutions.

Table 4: Protein content in three allergenic food extracts using three extraction buffers. Values are expressed as mean ± standard deviation (n=4 for all samples). Mean within each column with different superscripts differs at $p < 0.05$

	Protein concentration, µg/mL		
Extraction buffers	**Fennel**	**Celery**	**Carrot**
TRIS-HCl 50 mM, pH 8,2	38 ± 3^{a}	102 ± 8^{a}	39 ± 2^{a}
Carbonate buffer 100 mM, pH 8	73 ± 4^{b}	56 ± 4^{b}	122 ± 10
Phosphate buffer 10 mM, KCl 1 M, pH 4	10 ± 1^{c}	16 ± 1^{c}	20 ± 4^{c}

Differences among extracts by Tris and Carb were more marked in the cases of celery and carrot than with the fennel extract, which contained significantly fewer proteins. In the case of celery, Tris extract contained double the amount of protein in the Carb extract. On the contrary, in the case of carrot, the Carb extract contained three times the amount of protein in the Tris extract. The results from HSPB extraction suggest that either low pH or high ionic strength are not the best conditions for the solubilization of these kind of plant allergens. Considering the number of different allergens extracted, as described below (Table 5), no allergens were found in all extracts from carrot and fennel using this method. On the contrary, 15 allergenic peptides were found in the celery extract: 11 from Carb extraction and 4 from Tris extraction. The differences either in the quantity or in the quality of allergens may be due to some aspects such as, to cite a few, buffer compatibility and some possible modifications of proteins due to various reactions with polyphenols and lipids.

In order to optimize the method, protein extraction was performed on the same amount (100 g) of homogenized sample kept at 50 °C for each buffer. Different conditions were tested and the following extraction parameters evaluated: (a) extraction volume, (b) extraction time, (c) temperature, and (d) sonication. The results showed that total protein content was optimum with 250 mL of extraction buffer, for 4 hours and at 50 °C, without sonication. When the same extractions were conducted with sonication, the results were significantly worse than the standard protocol. In a previous work, Monaci confirmed this effect just for plant proteins, in comparison with other proteins, such as milk caseins, in which sonication was useful for breaking down the casein micelles (*Monaci et al*, 2014).

Because Tris showed the best sequence coverage for the most of investigated proteins, it was chosen as the default extraction buffer. Although different amounts of proteins were extracted from all samples, no allergenic peptides from carrot and fennel were found. Another relevant piece of information related to celery is that Tris extraction gave a very different protein content compared with the other buffer. As will be shown below, the Tris extract contained almost all allergens from one protein, the RuBisCo, whereas allergens in the Carb extract were from different proteins.

There is an urgent need to clarify some important aspects regarding the extraction conditions to use and some aspects of the protocols, such as the selection of proper buffers for plant allergen extraction, sample pre-treatment and, eventually, the addition of denaturants and/or detergents. Additional information on the qualitative aspects of the extraction is needed in order to better characterize the ability of different buffers to extract proteins from food. The overall results of this investigation confirmed the there is no universal extraction buffer for all food allergens from the *Apiaceae* family. Possible matrix effects should be taken into account since all buffers were unable to extract allergens from celery and fennel. Each extraction buffer need to be evaluated for every food and for every target allergen, therefore the results from this kind of analysis should include the selection of several mass spectrometry marker peptides, as explained below. Extraction efficiency should be improved and recovery experiments should be performed in a further step.

4.3.2 ENZYMATIC DIGESTION AND LC-ION-TRAP MS/MS

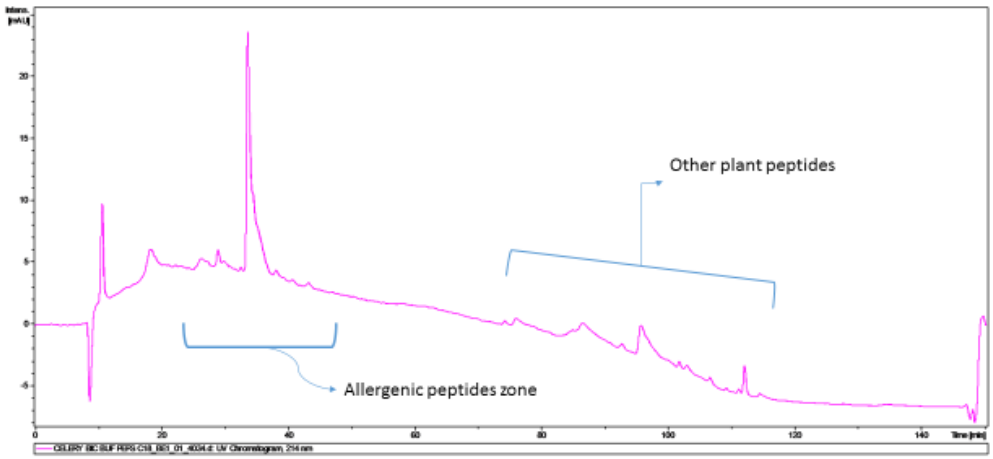
Proteins extracted from the three allergenic food were purified and digested with pepsin for four hours, according to the protocol described in the Materials and Methods section. Both trypsin and chymotrypsin were tested for this step, but pepsin was finally chosen as the default enzyme for two reasons: firstly, because the peptic peptides led to better results during the identification of allergenic sequences, and secondly, the effect of this kind of digestion on the allergenicity of proteins requires a few considerations. The human stomach in adults secretes about 10 mg/day of pepsin (in addition to other proteolytic enzymes) with a typical intake of proteins of about 75 g, corresponding to 20-30 kU of enzyme activity at 37 °C (*Jin Hur et al*, 2011), this gives pepsin a crucial role in digestion of allergens introduced with the diet. A certain number of allergens, like the LTPs, are known to be resilient to pepsin activity (*Vassilopoulou et al*, 2006), while other important allergens can bind IgE antibodies and, therefore, enhance allergenic potential after digestion (*Moreno*, 2007). The reason that gives a protein (or peptide) the ability to sensitizes an allergic individual remain unknown, nevertheless, allergen digestibility is a key factor to consider in the IgE mediated mechanism since it may affect the allergenic potential and it may

release (or destroy) allergenic epitopes. Although the physiological importance of allergen interactions on their allergenicity remains to be demonstrated, pepsin remains the key enzyme in the physiology, digestibility and, hence, in the allergenicity of food proteins (*Moreno*, 2007). For these reasons, the choice of pepsin should be considered in all allergenomic digestion protocols.

4.3.3 LIQUID CHROMATOGRAPHY

Peptides from peptic digestion were separated on a nano-LC reversed phase chromatography system ESI-IT-MS/MS analysis as described in the experimental section. Chromatographic separation was achieved with a reversed phase C_{18} nano column. Several studies were performed during LCMS analysis for optimizing separation of the allergenic peptides.

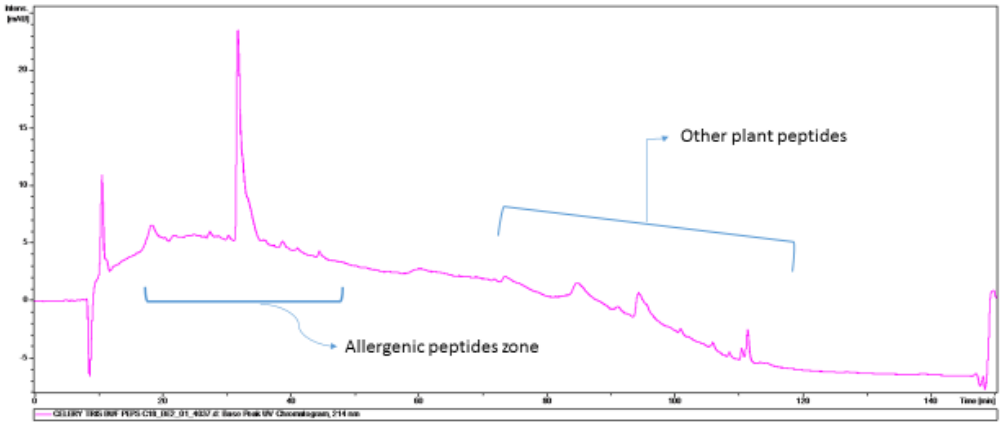
Figure 9: Uv chromatogram at 214 nm of peptic peptides from celery, extracted by carbonate buffer. All allergens from this sample eluted from 22 to 43 min, corresponding to 20% - 30% of organic mobile phase.



The best results in terms of good resolution of peaks and convenient analysis time were obtained using a 150 min Water/Formic Acid/Acetonitrile multi step elution gradient; this consisted of a nonlinear gradient (curve 7) in which the percentage of aqueous mobile phase decreased from 96% to 10% before the final equilibration step (see the experimental section). Curve no. 7 gave the best results for the early elution of allergenic peptides.

No allergens were found in samples from carrot and fennel, but it was interesting to note that proteins extracted from celery revealed several allergenic peptides. For this reason, only the results from celery extraction will be shown only. Figs. 9 and 10 show two representative examples of the chromatographic separation (UV, 214 nm) of peptic solution of celery under the optimized elution gradient.

Figure 10: Uv chromatogram at 214 nm of peptic peptides from celery, extracted by TRIS HCl buffer. All allergens from this sample eluted from 18 to 40 min, corresponding to around 15% - 25% of organic mobile phase.



Both chromatographic separations from the buffers Carb and Tris revealed the elution of allergenic peptic peptides in the first section of the gradient, corresponding to lower concentration of the organic mobile phase.

In particular, all allergenic peptides in the Carb extraction eluted from 22.3 minutes to 42.7 minutes, corresponding to Acetonitrile concentrations approximately from 20% to 30%. The allergens from Tris extraction showed a similar behavior, as they eluted from 18.8 to 40 minutes, corresponding to organic concentrations from approximately 15% to 25%. These result seem to be compatible with the mostly hydrophilic nature of the allergens found, that are listed below. The second section of chromatographic separation revealed the presence of some other plant peptides that, at the current state of the art, are not relevant from the allergenomic point of view, but could be useful for further investigations. The allergenic sequences found are described in the following sections.

4.3.4 MASS SPECTROMETRY ANALYSIS

The allergenic pattern of carrot, fennel and celery extracts and the characterization of the resultant allergenic peptides were analyzed by nanoliquid chromatography, electrospray ionization and ion trap mass spectrometry (MS). A specific bioinformatic approach based on scoring distribution and Allergome® (*Mari,* 2015) database search was applied, based on MASCOT® scores (*The Matrix Science*, 2014) higher than an arbitrary cut-off level of 20 in order to focus on the most significant peptides and minimize the risk of false-positives. A fundamental step in the development of MS-based methods for food allergen detection is the selection of peptides with specific requirements to be considered consistent markers for the target allergens. This strategy is used by several scientists (*Monaci et al*, 2009); another strategy is to find the highest number of allergens from a matrix (*Ansari et al*, 2011). Thus, a preliminary untargeted MS analysis was carried out on the plant protein extracts to identify the most abundant peptides attributed to each allergenic food under investigation, in order to find one or more candidate markers. To this aim, the peptides eluted from nano-HPLC were analyzed by mass spectrometry. A full-scan MS acquisition in the 300–1500 m/z range was performed with the acquisition of Base Peak Chromatogram (BPC) and Total Ion Current (TIC) profile by using a high capacity ion trap, coupled to a nano-ESI source. MS/MS analysis by collision-induced dissociation (CID) was performed using unattended data-dependant acquisition (DDA) with the criteria described in the experimental section. The acquired MS and MS/MS data were submitted to NCBI-nr database search using Biotools™ 3.2 software and the MASCOT® platform as a search engine. The alignment of allergenic sequences was performed on the Allergome® database using the algorithm NCBI blastp (2.2.18) with specific parameters as listed in the experimental section. This bioinformatic approach was used for peptide identification, applying the strategy optimized in the second year of this PhD project. DDA data were processed to identify plants peptic peptides and to assess the sequence coverage for each allergen.

The entire chromatographic runs (BPC of all MS^n spectra), of celery peptides, extracted with Carb and Tris are shown below, in Figs. 11-12.

Figure 11: Base peaks chromatogram (BPC, all Ms^n) of celery peptic peptides, extracted by Carb buffer and separated by nano-RP-HPLC

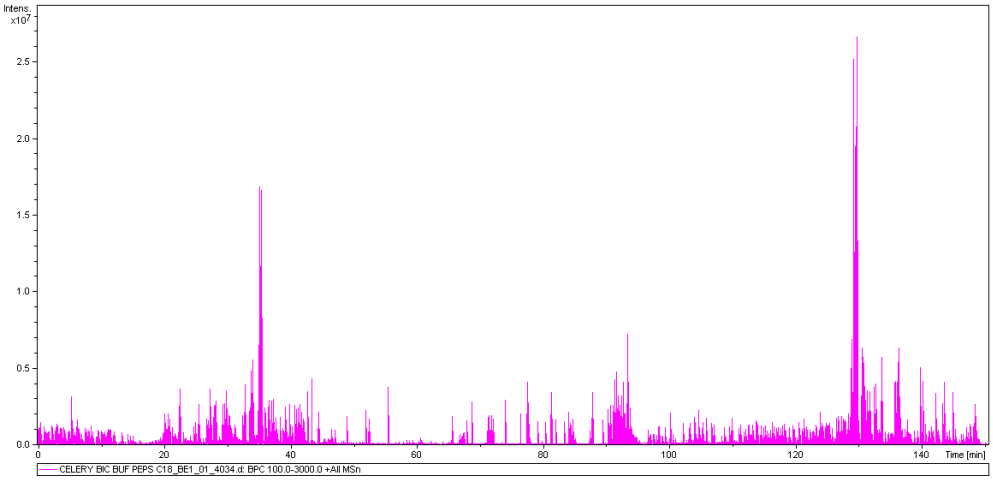
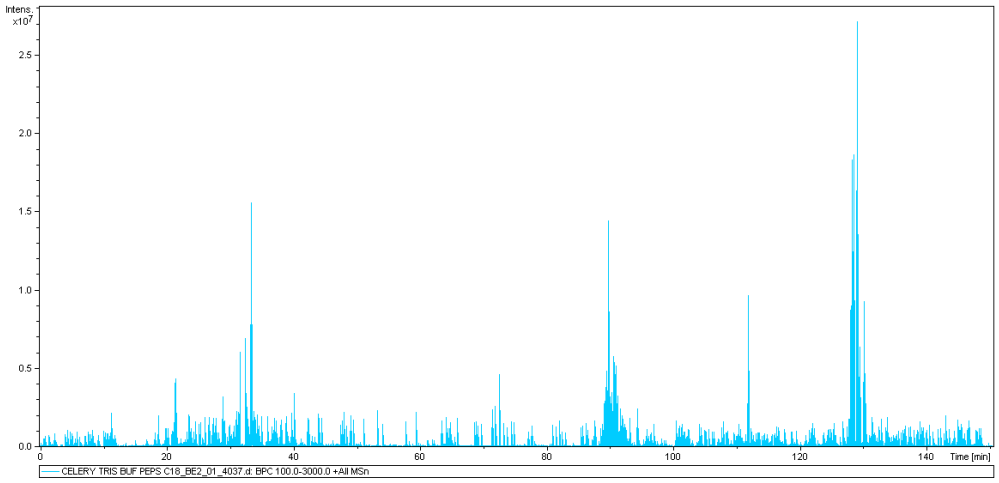


Figure 12: Base peaks chromatogram (BPC, all Msn) of celery peptic peptides, extracted by Tris buffer and separated by nano-RP-HPLC



4.3.5 ALLERGEN IDENTIFICATION AND POTENTIAL MARKER PEPTIDES SELECTION

The most intense ions identified by the Mascot® software (see the experimental section of this chapter) and by the Allergome® database of the two celery extracts (from Carb and Tris buffer) peptic digests are reported in Table 5. The investigation on the Allergome database, coupled with selective data elaboration (as explained above) proved to be the most suitable method for allergenomic purposes, since it allowed automatic identification of all allergenic sequences coming from MS/MS analysis. A total of 15 allergenic peptides were identified with this approach. In particular, Carb buffer extracted 11 allergens, and Tris buffer 4 allergens, most of which came from the Ribulose bisphosphate carboxylase. Regarding the presence of food allergens in the extracts, Carb buffer demonstrates a higher ability to extract allergens from several proteins, with most of them being very important in food immunology, such as the Non-Specific Lipid Transfer Protein (NSLTP), the Olee-1 like protein, and the Lecitin seed storage protein. In the case of celery, Carb buffer seems to be a valid sample extraction buffer for multi-allergen analysis purposes, because it can extract more allergens than Tris buffer. It should be noticed that Carb extraction data are less significant, as their Mascot score distributions were borderline to a p-value of 0.05. Recovery experiments are required, in parallel with method validation. On the contrary, the information about Rubisco peptides from Tris extracts could be useful for selection of potential marker peptides and for fast screening purposes, as described in the next section. Interpretation of the results from an LC-MS/MS search can be very complex for the analysis of low abundant proteins such as most plant allergens, which also influence their extraction. Sometimes it is not always clear which peptide "belongs" to which protein. Either peptide masses or MS/MS fragment ion masses matches are always handled on a probabilistic basis, but reporting probabilities directly can be confusing because they encompass a very wide range of magnitudes, and also because a "high" score is a "low" probability, which can be ambiguous! The scoring from the MS tools (in addition to many other parameters such as the taxonomy selection) help scientist to select the significant sequence among all that are reported at the end of the analysis. In the Mascot MS/MS database, the Mascot

Probability Based Scoring (or “Mascot score”) is a simple parameter to judge whether a result is significant or not. The total score is the probability that the observed match is a random event. The significance level is always given from the Mascot search in the “protein summary” page and it changes on each analysis. As a typical example, the Mascot scores greater than 50 are significant with $p<0.05$. In addition, the use of red and bold typefaces is intended to highlight the most logical assignment of peptides to proteins. The first time a match to a spectrum appears in the report, it is shown in bold face. Whenever the top scoring peptide match for a spectrum appears, it is shown in red, this means that peptide matches which are both bold and red are the most likely assignments of the best matches.

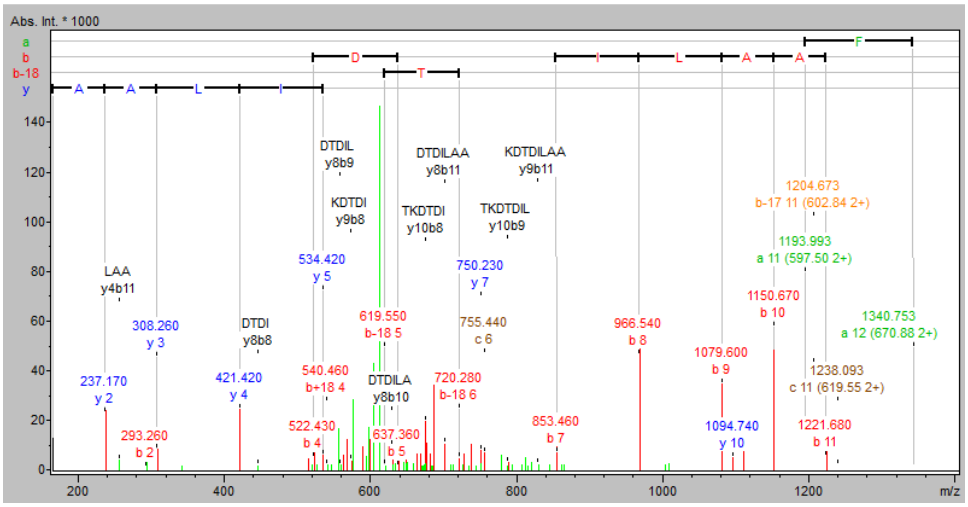
Table 5: Protein hits, peptides detected and relative sequence assigned by interrogating both the NCBI and the Allergome databases with MS data on to two celery samples extracted by Carb and Tris buffer and digested with pepsin.

#	Allergome ID	Allergenic peptide (z)	Theoretical pI/Mw	Observed m/z	Allergen origin	Mascot Score	Allergen Identity	Buffer	Rt (min)	Protein origin
1	1059	KLDHDACGNIDATIAPCF (+2)	4.41 / 1904.14	635,620	**Tri a 14** (*Triticum aestivum*)	26	53%	Carb	22,3	Non-specific lipid transfer protein (NSLTP)
2	3334	KVYGGYDEDDL (+2)	3.84 / 1273.32	637,780	**Jug r 1.0101** (*Juglans regia*)	25	85%	Carb	23,3	Albumin seed storage protein
3	6326	PYEYVEENLGDKSDL (+2)	3.83 / 1770.87	885,500	**Tri a GST** (*Triticum aestivum*)	25	66%	Carb	23,4	Glutathione transferase
4	3814	SGGDHIPSGTVVGKLEGEREITL (+2)	4.83 / 2351.60	784,400	**Spi o RuBisCo** (*Spinacia oleracea*)	25	91%	Carb	29,6	Ribulose bisphosphate carboxylase
5	1188	LAAQAIRASALSVERKNRSQDFSDYDDASAKSGFW (+3)	6.20 / 3862.19	1188,170	**Gly m 39kD** (*Glycine max*)	19	61%	Carb	30,4	Lecitin seed strorage protein
6	9598	NKQVVNGWKDMNEECLKPTQVPMPL (+3)	6.17 / 2899.39	966,860	**Tri a Tritin** (*Triticum aestivum*)	28	56%	Carb	30,8	rRNA N-glycosidase
7	895	L.NSHEQDGFHKVGGSSHPLA (+2)	6.26 / 2117.27	966,870	**Cup a 1.02** (*Cupressus arizonica*)	30	41%	Carb	30,8	Pectate lyase
8	9881	TGKVVRSGDYELIMDGVHENIL (+2)	4.75 / 2445.77	815,530	**Bet v BB18** (*Betula pendula*)	27	50%	Carb	31,1	Olee1-like protein
9	3814	EYEGTSADILAAFR (+2)	4.14 / 1542.67	693,970	**Spi o RuBisCo** (*Spinacia oleracea*)	31	71%	Carb	32,1	Ribulose bisphosphate carboxylase
10	3814	FEFQAMDTI (+1)	3.67 / 1101.24	551,340	**Spi o RuBisCo** (*Spinacia oleracea*)	35	77%	Carb	39,5	Ribulose bisphosphate carboxylase
11	9502	VPTPNVSVVDLTVRLGK (+2)	8.72 / 1794.12	897,580	**Tri a 34.0101** (*Triticum aestivum*)	30	85%	Carb	42,7	Glyceraldehyde-3-phosphate
12	3814	FRVSPQPGVPPEEAGAAVA +2)	4.53 / 1879.10	866,480	**Spi o RuBisCo** (*Spinacia oleracea*)	33	100%	Tris	18,8	Ribulose bisphosphate carboxylase
13	3814	FYETKDTDILAAFR (+2)	4.56 / 1689.89	693,970	**Spi o RuBisCo** (*Spinacia oleracea*)	53	92%	Tris	30,1	Ribulose bisphosphate carboxylase
14	1534	AVDAGCPKPSDVVEAGVEGGDESVVTVTL (+3)	3.71 / 2801.07	934,340	**Fag t 1** (*Fagopyrum tataricum*)	30	39%	Tris	31,5	Allergenic protein
15	3814	YTPEYETQDTDILAAFR (+2)	3.92 / 2033.18	939,000	**Spi o RuBisCo** (*Spinacia oleracea*)	35	94%	Tris	33,9	Ribulose bisphosphate carboxylase

Particular attention was given to the separation of a valid candidate to be used as an allergenic marker peptide. Rapid peptic peptide separation is very important for the development of a reliable LC-ESI-MS/MS method for fast screening analysis in food control quality. MS full scan experiments of two celery extracts were performed in the mass range of 300–1500 m/z. MS signals of putative singly, doubly and triply charged peptic peptides ions, calculated from the data of the corresponding allergenic proteins, were searched in the MS spectra. The obtained peptide will serve as a candidate for further MS experiments. Gradient optimization, involving acetonitrile as an organic-phase modifier in water, allowed allergenic peptide elution within 50 min (see Figure 9 and Figure 10). In order to select a peptide able to target investigated *Apiaceae* allergens, occurring in peptic digest and satisfying stringent criteria such as good ESI sensitivity, no post-translational modification sites, and sequence specificity, a few peptides from the Tris extract was chosen. In addition to these criteria, the potential markers were selected among those showed in bold red, which had the highest mascot score and the lowest level of similarity. Table 6 shows the BLAST analysis of the 15 Apiaceae allergens versus in the "green plants" database (tax ID 33090). In table 6 are reported (where found) the sequences with a high level of similarity (> 95%). The Basic Local Alignment Search Tool (BLAST) is a powerful tool for the comparison of nucleotide or protein sequences from the same or different organisms. The BLAST is based on a heuristic algorithm, which means that it performs "local" alignments on some smart shortcuts in order to search faster. Most proteins are modular in nature, with functional domains often being repeated within the same protein as well as across different proteins from different species. The BLAST algorithm is tuned to find these domains or shorter stretches of sequence similarity. By finding similarities between sequences, scientists can infer the function of newly sequenced genes, predict new members of gene families, and explore evolutionary relationships. Among the 15 allergens found in the Apiaceae samples (see table 5), the peptides nos. 3, 10, 12 and 15 should be excluded as they have more than three sequences with high similarity. The peptides nos. 1, 2, 4, 5, 6, 7, 8, 13 and 14 presented just one significant sequence each.

Among those, the peptide no. 13 (FYETKDTDILAAFR) was the most significant and should be proposed as a potential marker for the fast detection of *Apiaceae* allergens through ESI-MS^2. In this case, other aspects regarding this allergen should be investigated, such as its behavior under food processing and its role in the allergenic reactions (epitope mapping), and a good label free quantification method should be developed. Figure 13 shows the MS/MS product ion spectrum of the $[M+2H]^{2+}$ precursor ion at m/z 693.970, of which the entire amino acid sequence of the peptic peptide can be deduced.

Figure 13. MS/MS spectrum of the peptic peptide FYETKDTDILAAFR from the allergen Spi o Rubisco.



Rubisco is an enzyme involved in the first major step of carbon fixation, a process by which atmospheric carbon dioxide is converted by plants into energy-rich molecules such as glucose. In chemical terms, it catalyzes the carboxylation of ribulose-1,5-bisphosphate (also known as RuBP), a very complex protein (*Spreitzer*, 1999) whose allergenic potential has been discovered in allergenic reactions coming to spinach and tomato (*Herrera et al, 2002; Foti et al*, 2012).

Table 6: BLAST analysis of *Apiaceae* allergens vs. the green plants database (TAX ID 33090)

#	Allergenic peptides	Sequences producing >96% significant alignments	Query cover %	Identity %
1	KLDHDACGNIDATIAPCF	PREDICTED: 125 kDa kinesin-related protein [*Amborella trichopoda*]	100	100
2	KVYGGYDEDDL	PREDICTED: probable disease resistance protein At4g27220 [*Fragaria vesca*]	100	100
3	PYEYVEENLGDKSDL	Os10g0530900 [*Oryza sativa* Japonica Group]	100	100
		Hypothetical protein OsJ_32247 [*Oryza sativa* Japonica Group]	100	100
		Unnamed protein product [*Oryza sativa* Japonica Group]	100	100
4	SGGDHIPSGTVVGKLEGEREITL	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, [*Nesaea myrtifolia*]	100	100
5	LAAQAIRASALSVERKNRSQDFSDYDDASAKSGFW	PREDICTED: uncharacterized protein LOC103450787 [*Malus domestica*]	100	97
6	NKQVVNGWKDMNEECLKPTQVPMPL	PREDICTED: (-)-germacrene D synthase [*Vitis vinifera*]	100	100
7	L.NSHEQDGFHKVGGSSHPLA	Predicted protein [*Physcomitrella patens*]	100	100
8	TGKVVRSGDYELIMDGVHENIL	Chromosome transmission fidelity protein 18-like protein [*Morus notabilis*]	100	100
9	EYEGTSADILAAFR	*none*	-	-
10	FEFQAMDTI	> 10	-	-
11	VPTPNVSVVDLTVRLGK	*none*	-	-
12	FRVSPQPGVPPEEAGAAVA	> 10	-	-
13	FYETKDTDILAAFR	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, [*Lachnostylis bilocularis*]	100	100
14	AVDAGCPKPSDVVEAGVEGGDESVVTVTL	predicted protein [*Micromonas* sp. RCC299]	100	100
15	YTPEYETQDTDILAAFR	>10	-	-

The allergens no. 9 and no. 11 (see table 5) from ribulose bisphosphate carboxylase and glyceraldehyde-3-phosphate respectively, showed the lowest similarity (no significant sequences above 95% of alignments) but their identification should be improved in further analysis as they presented a less significant score.

4.4 CONCLUSION

In the third year's research, the allergenomic platform developed during the second year of this PhD thesis, was applied on *Apiaceae* samples in order to find a LCMS method for detection and characterization of *Apiaceae* allergens. For this purpose, carrot, celery and fennel samples underwent different eco-compatible extraction protocols based on saline buffers Carb, HSPBA and Tris. HSPBA buffer was unable to extract allergens from all samples; on the contrary the Tris and Carb buffers were able to extract a total of 15 allergens from celery. Most allergens were not directly related to *Apiaceae* allergens, basing on the data from IUIS official database (IUIS, 2015), in fact, other allergens from important families were found, like a NSLTP-like allergen and a Bet v-related allergen. The best chromatographic separation was achieved by a 150 minutes complex gradient of water and acetonitrile, which allowed a valid separation of either allergenic peptides or other plant peptides, whose study could be useful in the next. In addition, the hydrophilic peptide FYETKDTDILAAFR from Rubisco was proposed as a potential marker for routine detection of *Apiaceae* allergens in food.

REFERENCES

Aebersold R, Mann M, 2003. Mass spectrometry-based proteomics. Nature 422 (6928): 198–207

Ahamed, T., Chilamkurthi, S., & Nfor, B., 2008. Selection of pH-related parameters in ion-exchange chromatography using pH-gradient operations. Journal of Chromatography A 1194, 22-29.

Akagawa M, Handoyo T, Ishii T, Kumazawa S, Morita N, Suyama K, 2007. Proteomic analysis of wheat flour allergens. J Agric Food Chem 55:6863-6870

Ansari, P., Stoppacher, N., & Baumgartner, S, 2011. Marker peptide selection for the determination of hazelnut by LC–MS/MS and occurrence in other nuts. Analytical and Bioanalytical Chemistry, 402:2607–2615.

Andjelkovic U, Martinovic T, Josic D, 2015. Foodomics investigations of food allergies. Current Opinion in Food Science, 4:92–98.

Ansari, P., Stoppacher, N., Rudolf, J., Schumacher, R., & Baumgartner, S., 2010. Selection of possible marker peptides for the detection of major ruminant milk proteins in food by liquid chromatography-tandem mass spectrometry. Analytical and Bioanalytical Chemistry(399), pp. 1105–1115.

Asero R, Antonicelli L, Arena A, Bommarito L, 2009. Causes of Food-Induced Anaphylaxis in Italian Adults: a Multi Centre Study. International Archives of Allergy and Immunology, 150:271-277.

Australia New Zealand Standard 1.2.3, 2013 – Mandatory Warning and Advisory Statements and Declarations (Australia New Zealand Standard).

Bellotti, A., Caiaffa, M., & De Serio, A., 2013. Studi "in vivo" sulla cross-reattività allergenica tra polline di Compositae anemofile e Compositae edibili. Incontro delle Scuole di Specializzazione in Allergologia ed Immunologia Clinica, (p. 63-65). Montesilvano

Bignardi, C., Elviri, L., & Penna, A., 2010. Particle-packed column versus silica-based monolithic column for liquid chromatography–electrospray-linear ion trap-tandem mass spectrometry multiallergen trace analysis in foods. Journal of Chromatography A(1217), pp. 7579-7585.

Bignardi, C., Mattarozzi, M., & Penna, A., 2013. A Rapid Size-Exclusion Solid-Phase Extraction Step for Enhanced Sensitivity in Multi-Allergen Determination in Dark Chocolate and Biscuits by Liquid Chromatography-Tandem Mass Spectrometry. Food Analythical Methods 6, 1144-1152.

Björkstén F, Halmepuro L, Hannuksela M, Lahti A, 1980. Extraction and properties of apple allergens. Allergy, 35 (8):671-678.

Boye, J. I., & Godefroy, S. B., 2010. Allergen Management in the Food Industry. Hoboken, New Jersey: John Wiley & Sons, Inc.

Bradford MM, 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72:248–54

Breiteneder, H., & Mills, E., 2008. Food Allergens: Molecular and Immunological Characteristics. In D. Metcalfe, H. Samson, & R. Simon, Food Allergy: Adverse Reactions to Foods and Food Additives (4th ed., pp. 46-61). Oxford: Blackwell Publishing.

Caiaffa, M., Damiani, E., & Giliberti, A., 2011. Allergeni alimentari emergenti dalla dieta mediterranea. Incontro Nazionale di Aggiornamento delle Scuole di Specializzazione in Allergologia ed Immunologia Clinica, (p. 21-28). Montesilvano.

Ciardiello, M., Tamburrini, M., & Liso, M., 2013. Food allergen profiling: A big challenge. Food Research International 54:1033-1041.

Codex General Standard for the Labeling of Prepackaged Foods, 2010. (CODEX STAN 1-1985, Rev. 2010).

Cruz-Romero, M., Murphy, T., & Morris, M., 2013. Antimicobial activity of chitosan, organic acids and nano- sized solubilisates for potential use in smart antimicrobially-active packaging for potential food applications. Food Control 34, 393-397.

Cucua, T., De Meulenaera, B., & Devrees, B., 2011. MALDI based identification of soybean protein markers – Possible analytical targets for allergen detection in processed foods. Peptides 33:187-196.

D'Alessandro, A., & Zolla, L., 2012. We Are What We Eat: Food Safety and Proteomics. Journal of Proteome Research, 11(1):26-36.

European Commission Directive 2003/89/EC and Annex IIIa (European Union 2003).

European Commission Directive 2007/68/EC and Annex IIIa (European Union 2007).

Food Allergen Labeling and Consumer Protection Act of 2004, United States 2004.

Foti C, Damiani E, Zambonin CG, Cassano N, 2012. Urticaria and angioedema to rubisco allergen in spinach and tomato. Annals of Allergy Asthma Immunology, 108:60–67.

Fuchs M, Cichna-Markl M, Hochegger R, 2013. Development and validation of a duplex real- time PCR method for the simultaneous detection of celery and white mustard in food. Food Chemistry, Analytical Methods, 141:229-235

Gendel, S. M., 2012. Comparison of international food allergen labeling regulations. Regulatory Toxicology and Pharmacology, 63:279-285.

Halmepuro L, Løswenstein H, 1985. Immunological investigation of possible structural similarities between pollen antigens and antigens in apple, carrot and celery tuber. Allergy , 40:264-272.

Han, X., Aslanian, A., & Yates III, J., 2008. Current Opinion in Chemical Biology, 12:483-490

Health Canada Amendments to the Food Allergen Labelling Regulations, Canada Gazette, Part II (CGII), 2011.

Herrera I, Moneo I, Caballero ML, De Paz S, 2002. Food allergy to spinach and mushroom. Allergy, 57:261–262.

Hochwallner, H., Schulmeister, U., Swoboda, I., Spitzauer, S., & Valenta, R., 2014. Cow's milk allergy: From allergens to new forms of diagnosis, therapy and prevention. Methods, 66:22-33.

Hur SJ, Lim BO, Decker EA, McClements DJ, 2011. Review: In vitro human digestion models for food applications. Food Chemistry, 125:1-12.

IUIS, 2015. International Union of Immunological Societies Allergen Nomenclature: IUIS official list. [Online], http://www.allergen.org/List [Accessed 2015].

Jasim, A., & Shafiur Rahman, M. 2012. Handbook of Food Process Design. Oxford: Blackwell Publishing Ltd.

King, T., Hoffman, D., Løwenstein H, H., Marsh, D., & Platts., 1995. Allergen nomenclature. Allergy, 50(270):765-774.

Kirsch, s., Fourdrilis, S., Dobson, R., Scippo, M. L., Maghuin-, G., & De Pauw, E., 2009. Quantitative methods for food allergens: a review. Analytical and Bioanalytical Chemistry , 395:57-67.

Koppelman, S. J., & Hefle, S. L., 2006. Detecting allergens in food. Cambridge: Woodhead Publishing Ltd.

Krska, R., Becalski, A., Braekevelt, E., Koerner, T., Dabeka , R., Godefroy, S., Forsyth, D., 2011. Challenges and trends in the determination of selected contaminants and allergens in food. Analytical and Bioanalytical Chemistry(402), pp. 402:139-162.

Kruse Faeste, C., Thorsen Rønning, H., Christians, U., & Granum, P., 2011. Liquid Chromatography and Mass Spectrometry in Food Allergen Detection. Journal of Food Protection, 74(2):316–345.

Kunda, P., Benavente, F., & Català-Clariana, S., 2012. Identification of bioactive peptides in a functional yogurt by micro liquid chromatography time-of-flight mass spectrometry assisted by retention time prediction. Journal of Chromatography A, 1129:121-128.

L'Hocine L, Pitre M, 2015. Quantitative and qualitative optimization of allergen extraction from peanut and selected tree nuts. Part 1. Screening of optimal extraction conditions using a D- optimal experimental design. Food Chemistry, 194:780–786

Lippe, G., Conte, L., Foti, S., & Cusolo, V., 2011. Application of proteomics in food science. Journal of Scientific & Industrial Research 70:905-908

Losito, I., Mavelli, F., Demarinis Loiotile, A., & Palmisano, F., 2012. A support for the identification on non- tryptic petides based on low resolution tandem and sequential mass spectrometry data: the INSPIRE software. Analytica Chimica Acta, 718:70-77.

Mamone, G., Addeo, F., Ferranti, P., & Picariello, G., 2011. The frontiers of mass spectrometry-based techniques in food allergenomics. Journal of Chromatography A,1218:7386– 7398.

Mari, C. 2003- 2015. The Allergome Database, www.allergome.org [Online] [Retrieved 2015]

Mills, C. - The Institute of Food Research. (n.d.). Retrieved July 2012, from The Informall Database: http://foodallergens.ifr.ac.uk/

Mills, C., Wichers, H., & Hoffmann-Sommergruber, K. 2007. Managing allergens in food. Boca Raton: CRC Press.

Mills, E., Johnson, P., & Alexeev, Y. 2012. Food Antigens. In J. M. James, W. Burks, & P. Eigenmann, Food Allergy. Edinburgh: Elsevier Inc.

Mills ENC, Breiteneider H, 2005. Molecular properties of food allergens. Journal of Allergy and Clinical Immunology, 115 (1):14-23.

Monaci, L., Palmisano, F., & Visconti , A., 2010. Identification of allergenic milk proteins markers in fined white wines by capillary liquid chromatography–electrospray ionization-tandem mass spectrometry. Journal of Chromatography A, 1217:4300–4305.

Monaci, L., & van Hengel, A., 2008. Development of a method for the quantification of whey allergen traces in mixed-fruit juices based on liquid chromatography with mass spectrometric detection. Journal of Chromatography A 1192, 113-120.

Monaci L, Visconti A , 2009. Mass spectrometry based proteomics methods for analysis of food allergens. Trends in Analytical Chemistry, 28(5):581-591.

Monaci L, Pilolli R, De Angelis E, Godula M, 2014. Multi-allergen detection in food by micro high-performance liquid chromatography coupled to a dual cell linear ion trap mass spectrometry. Journal of Chromatography A, 1358:136-144.

Moreno FJ, 2007. Gastrointestinal digestion of food allergens: Effect on their allergenicity. Biomedicine & Pharmacotherapy, 61:50-60.

Ng, P., He, J., & Snyer, M., 2009. Separation of protein mixtures using pH-gradient cation-exchange chromatography. Journal of Chromatography A 1216, 1372-1376.

Nollet, L. M., & Toldrà, F., 2007. Advances in Food Diagostics. Oxford: Blackwell Publishing.

Ötles, S., 2012. Methods of Analysis of Food Components and Additives (Second ed.). New York: CRC Press.

Pastorello EA, Trambaioli C, 2001. Review: Isolation of food allergens. Journal of Chromatography B, 756:71–84.

Pastorello, E., Farioli, L., & Stafylaraki, C., 2013. Fennel Allergy Is a Lipid-Transfer Protein (LTP)-Related Food Hypersensitivity Associated with Peach Allergy. Journal of Agricultural and Food Chemistry, 61:740-746.

Pedreschi, R., Norgaard, J., & Maquet, A., 2012. Current Challeges in Detecting Food Allergens by Shotgun and Targeted Proteomic Approaches: a Case Study on Traces of Peanut Allergens in Baked Cookies. Nutrients, 4(2):132-150.

Regulation No. 1169/2011 and Annex II (European Union 2011).

Resta, D., Brambilla, F., & Arnoldi, A., 2011. HPLC-Chip-Multiple Reaction Monitoring (MRM) method for the label-free absolute quantification of c-conglutin in lupin: Proteotypic peptides and standard addition method. Food Chemistry(131), pp. 126-133.

Rudolf J, Ansari P, Kern C, Ludwig T, 2012. Effects of different extraction buffers on peanut protein detectability and lateral flow device (LFD) performance. Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment , 29(1):1- 11.

Sathe, S., Teuber, S., & Roux, K., 2005. Effects of Food Processing on the Stability of Food Allergens. Biotechnology Advances, 23:423-429.

Sathe SK, Sharma GM, 2009. Effects of food processing on food allergens. Molecular nutrition and food research, 53(8):970-978

Schwager C, Kull S, Krause S, Schocker F, Petersen A, 2015. Development of a novel strategy to isolate lipophilic allergens (oleosins) from peanuts. PLoS ONE, 10 (4):1-15.

Simpson, B. K., 2012. Food Biochemistry and Food Processing. Oxford: Wiley-Blackwell.

Spreitzer RJ, 1999. Questions about the complexity of chloroplast Ribulose-1,5-bisphosphate. 60:29–42.

Japan Standards and Evaluation Division, 2013. Department of Food Sanitation, Pharmaceutical and Medical Safety Bureau, Ministry of Health, Labour e Food Labelling Act (Act No. 70 of 2013).

Stevenson, S., Ozias-akins, P., & Thelen, J., 2009. Validation of gel-free, label-free, quantitative proteomics approaches: Applications for seed allergen profiling. Journal of Proteomics, 72:555-566.

The Matrix Science, 2014. [Online] Retrieved [2015] from http://www.matrixscience.com/help/scoring_help.html.

University of Nebraska - Lincoln., 2010. Retrieved August 2012, from Allergene Online: http://www.allergenonline.org/

Vassilopoulou E, Rigby N, Moreno FJ, 2006. Effect of in vitro gastric and duodenal digestion on the allergenicity of grape lipid transfer protein. Journal of Allergy and Clinical Immunology, 118(2):473–480.

Van Hengel, A. J., 2007. Food allergen detection methods and the challenge to protect food-allergic consumers. Analytical and Bioanalytical Chemistry, 389:111-118.

Verhoeckx KCM, Vissers YM, Baumert JL, 2015. Food processing and allergenicity. Food and Chemical Toxicology, 80:223–240.

Westphal CD, Pereira MR, 2004. Evaluation of extraction buffers using the current approach of detecting multiple allergenic and nonallergenic proteins in food. Journal of AOAC International, 87(6):1458-1465.

Wienkoop S, Glinski M, Tanaka N, 2014. Linking protein fractionation with multidimensional monolithic reversed-phase peptide chromatography/mass spectrometry enhances protein identification from complex mixtures even in the presence of abundant proteins. Rapid Communications in Mass Spectrometry, 18(6):643–650.

Zahradnik C, Martzy R, Mach RL, Krska R, 2014. Detection of the food allergen celery via loop-mediated isothermal amplification technique. Analytical and Bioanalytical Chemistry, 406(27):6827-6833.