UNIVERSITÀ DEGLI STUDI DI PARMA

Dottorato di ricerca in
SCIENZE E TECNOLOGIE ALIMENTARI

Ciclo XXVIII

DIFFERENT ROUTES TO PURE FOOD ALLERGENS: EXTRACTION, RECOMBINANT TECHNIQUES AND TOTAL CHEMICAL SYNTHESIS FOR OBTAINING THE PLANT-FOOD PAN-ALLERGEN LTP

Coordinatore:
Prof FURIO BRIGHENTI

Tutor:
Dott.ssa TULLIA TEDESCHI
Prof STEFANO SFORZA

Dottoranda: SOFIE BUHLER

Abstract

Pure allergens are increasingly required for diagnostic purposes, as standards for detection and quantification methods, for immunotherapy and for the study of the mechanisms of allergic reactions at a molecular level, in order to ease the development of possible cures.
In this PhD thesis, different strategies for obtaining pure non-specific Lipid Transfer Proteins (nsLTPs), which were recognized as being relevant food allergens in many commonly consumed fruits and vegetables and were defined as models for true food allergens, are described.
A novel potentially allergenic LTP was isolated from almond, while a known allergenic LTP from walnut was produced by means of recombinant DNA techniques.
Besides these classical approaches, methods for the total chemical synthesis of proteins were applied for the first time to the production of an allergen, using Pru p 3, the prototypic LTP and major allergen of peach in the Mediterranean area, as a model. Total chemical synthesis of proteins allows to completely control their sequence and to study their function at atomic resolution. Thus, its application to the production of allergens constitutes an important step forward in the research field of food allergy. The whole protein stretch of Pru p 3 was produced and only two final deprotection steps are needed to obtain the target in its native form. The experimental conditions for those deprotections were set up during the production of the peptides sPru p 3 (1-37) and sPru p 3 (38-91), composing together the entire protein.
Advanced mass spectrometry techniques were used to characterise all the obtained compounds, while their allergenicity was studied through immunological tests or *in silico* approaches.

Riassunto

La richiesta di allergeni puri è in continuo aumento per scopi diagnostici, come standard per metodi di rilevamento e di quantificazione, per l'immunoterapia e per lo studio a livello molecolare dei meccanismi delle reazioni allergiche, al fine di facilitare lo sviluppo di possibili cure.
In questa tesi di dottorato sono descritte diverse strategie per l'ottenimento di forme pure di non-specific Lipid Transfer Proteins (nsLTPs), le quali sono state riconosciute essere rilevanti allergeni alimentari in molti frutti e verdure comunemente consumati e sono state definite come modello di veri allergeni alimentari.
Una LTP potenzialmente allergenica, non nota in precedenza, è stata isolata dalle mandorle, mentre una LTP dall'allergenicità nota contenuta nelle noci è stata prodotta mediante tecniche di DNA ricombinante.
Oltre a questi approcci classici, metodi per la sintesi chimica totale di proteine sono stati applicati per la prima volta alla produzione di un allergene, utilizzando Pru p 3, la LTP prototipica e principale allergene della pesca nell'area mediterranea, come modello. La sintesi chimica totale di proteine permette di controllarne completamente la sequenza e di studiare la loro funzione a livello atomico. La sua applicazione alla produzione di allergeni costituisce perciò un importante passo avanti nel campo della ricerca sulle allergie alimentari. La proteina Pru p 3 è stata prodotta nella sua intera lunghezza e sono necessari solo due passaggi finali di deprotezione per ottenere il target nella sua forma nativa. Le condizioni sperimentali per tali deprotezioni sono state messe a punto durante la produzione dei peptidi sPru p 3 (1-37) e sPru p 3 (38-91), componenti insieme l'intera proteina.
Tecniche avanzate di spettrometria di massa sono state usate per caratterizzare tutti i composti ottenuti, mentre la loro allergenicità è stata studiata attraverso test immunologici o approcci in silico.

Contents

1 Food allergy

1.1 Adverse reactions to food

According to the European Academy of Allergy and Clinical Immunology (EAACI) adverse reactions to food can be classified as[1]:

- **Toxic:** adverse reactions that occur in any individual exposed to a high enough dose of the triggering food component
- **Nontoxic:** adverse reactions that depend on individual susceptibility to a certain food component
    - **Food allergy:** immune mediated reaction
        - **IgE mediated**
        - **Non-IgE mediated**
    - **Food intolerance:** non-immune mediated reaction that can depend on metabolic characteristics of the patient (deficiencies of enzymes involved in food metabolism e.g. milk intolerance due to lactase deficiency), on pharmacologic properties of the ingested food (e.g. vasoactive properties of some amines) or can be undefined.

1.1.1 Immune mediated reactions

Food allergies are generally caused by proteins and affect about 6-8% of young children and 3-4% of the global adult population[2].
Individuals who have a predisposition toward developing certain allergic hypersensitivity reactions are defined as **atopic**. Atopy may have a hereditary component, although contact with the allergen must occur before the hypersensitivity reaction can develop[3].
Relatively few foods are responsible for the vast majority of food allergic reactions: milk, eggs, peanuts, tree nuts, fish, shellfish, celery, sesame, lupine, soy and cereals[2]. The relative importance of these foods varies widely with the age of the patients and the geographical location and several studies have established that dietary habits play a significant role in determining specific

food allergies; for example cow's milk and egg are the most prevalent food allergies in infants and children. This reflects the worldwide consumption of these foods in this age group[4].

Symptoms of food allergy may affect:

- Skin: pruritus, urticaria, angioedema, morbilliform rashes.
- Airways: rhinoconjunctivitis, bronchospasm.
- Gastro-intestinal tract: nausea, vomiting, gastric retention, intestinal hyper-motility, abdominal pain due to colonic spasms, diarrhoea.

Anaphylaxis is the worst allergic reaction, defined by EAACI as a severe, life-threatening, generalized or systemic hypersensitivity reaction; it is due to the abrupt, massive release of mediators from mast cells and/or basophils throughout the body, inducing gastrointestinal, skin, and respiratory symptoms, in some cases associated with cardiovascular symptoms including hypotension, collapse and dysrhythmia.

Many foods can trigger anaphylaxis and this may occur upon the first known ingestion[5].

On the basis of the physical-chemical properties of the allergen responsible for the allergic reaction and the underlying immunologic mechanisms, two forms of food allergy can be distinguished:

- **Class 1:** sensitization occurs through the intestinal tract and is often caused by proteins stable to heat, acid environment and enzymatic digestion; these features allow the survival of the allergens to gastro-intestinal digestion and their passage through the mucosal barriers with intact epitopes that can trigger the allergic reaction. Proteins responsible for allergies belonging to this class are defined as true food allergens.
- **Class 2:** food allergy develops as a consequence of sensitization to airborne allergens to which the food allergens are cross-reactive (e.g. allergic reaction to food proteins which are homologous to proteins present in birch tree pollen). Class 2 food allergens are very sensitive to heat and enzymatic digestion and they tend to induce reactions limited to oral allergy symptoms (Oral Allergy Syndrome, OAS).

1.1.1.1 IgE mediated reactions

IgE-mediated reactions are the most important type of food allergy because they involve a wide variety of different foods and the symptoms can be severe in some individuals.
Also known as immediate hypersensitivity reactions, they involve the formation of IgE antibodies that specifically recognize certain allergens. Normally IgE antibodies act as protectors against parasitic infection, while individuals predisposed to the development of allergies probably produce IgE antibodies that are specific for certain environmental antigens, typically proteins.
These allergen-specific IgE antibodies can sensitize mast cells and basophilic granulocytes. When an allergen interacts with these anchored IgE molecules, cross-linking two or more of them, a mediator release from the bound cells takes place (**Figure 1.1**).

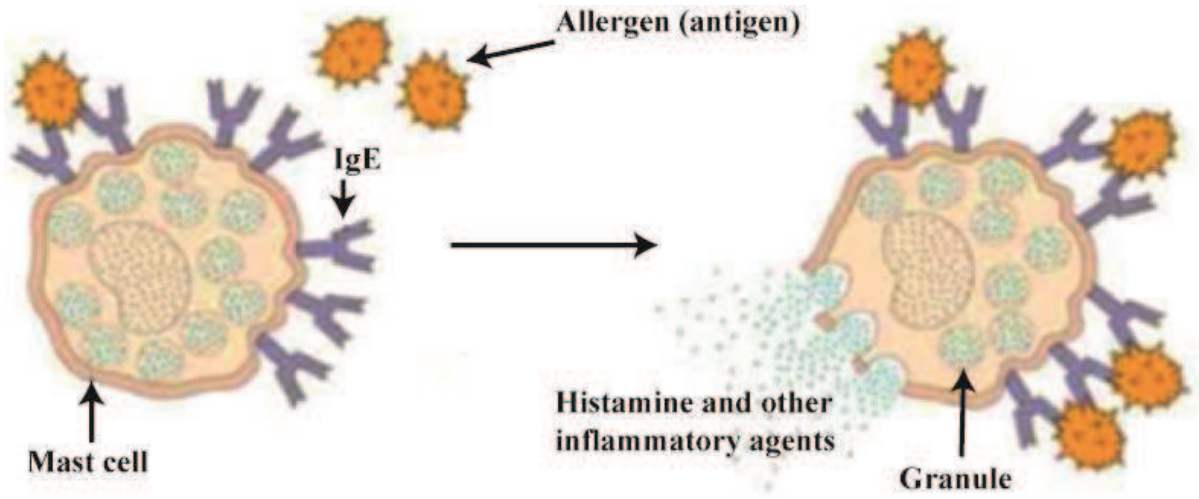


Figure 1.1 IgE mediated allergic reaction.

The amino acid residues of an allergen responsible for the interaction with the specific IgE molecules, constitute an **epitope**. Epitopes can either be linear, if the amino acids that interact with the antibody are sequential in the allergen, or conformational, if the interacting amino acids belong to different sections of the allergenic protein, but are close to each other in the tertiary structure of the latter (**Figure 1.2**).

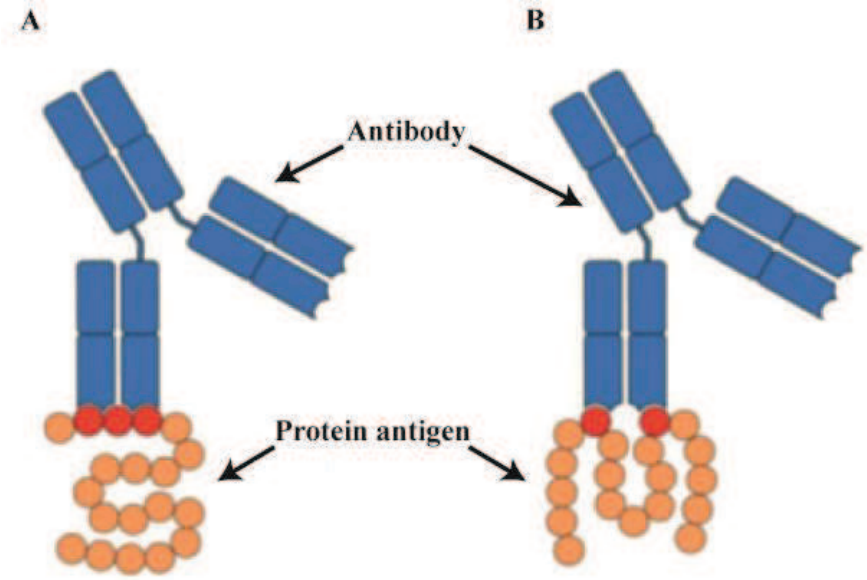


Figure 1.2 Linear (**A**) and Conformational (**B**) epitopes.

As the antigenic reactivity of conformational epitopes relies on the native conformation of the protein, they are expected to be more susceptible to processing or digestion-induced structural changes. Thus, linear epitopes are considered to be more important in class 1 food allergy and conformation epitopes in pollen-related (class 2) food allergy[6].

It is not known why only certain individuals produce specific IgE antibodies when encountering allergens. Several factors are probably involved, including: host genotype, type and concentration of the allergen, route of exposure and possible presence of agents that can enhance the sensitization process[7].

The severity of food allergic reactions may differ between individuals and vary in time for the same individual; factors contributing to these differences include genetic background, variations in diet, geographical parameters, degree of digestion of the allergen by the gastro-intestinal tract, food matrix and food processing[8, 9].

IgE-mediated allergic reactions involve two stages:

- **Sensitization phase:** first contact with an allergen in consequence of which specific IgE antibodies are produced.
- **Elicitation phase:** further expositions to the allergen; appearance of the symptoms.

1.1.1.1.1 Sensitization phase

When an allergen gets in contact for the first time with the mucous membranes of an atopic individual, antigen presenting cells such as Langerhans cells in the epithelium internalize, process and then express these allergens on their cell surface. The allergens are then presented to other cells involved in the immune response, particularly T-lymphocytes; as a consequence, B-lymphocytes are transformed into antibody secretory cells called plasma cells. In the allergic response, plasma cells produce IgE-antibodies, which are capable of binding a specific allergen through their Fab portion. Once the IgE antibodies are formed and released into the circulation, they bind, through their Fc portion, to high affinity receptors (FcεR1) on mast cells, leaving their allergen specific receptor sites available for future interaction with antigens (**Figure 1.3**).

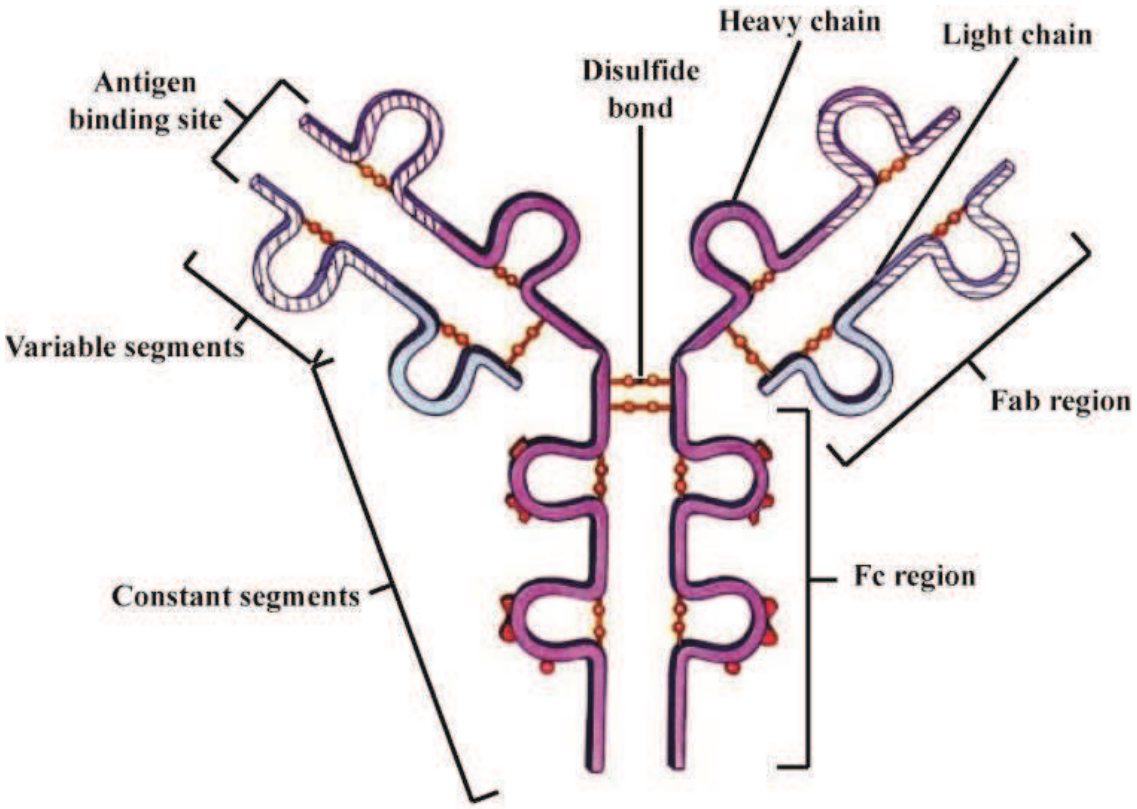


Figure 1.3 Structure of an IgE antibody.

Cells able to express high-affinity receptors for IgE include Langerhans cells, basophils and activated monocytes.

1.1.1.1.2 Elicitation phase

Upon re-exposure of a sensitized individual to an allergen, binding of the latter to IgEs induces the immune system to initiate a rapid and more aggressive memory response. Cross-linking of a sufficient number of mast cell/basophil-bound IgE antibodies by the allergen initiates a process of intra-cellular signalling, which leads to degranulation of cells, with the release of mediators of inflammation (prostaglandins, leukotriens, cytokines).
An immediate reaction, mainly due to histamine, follows a few minutes after the contact with the allergen. The release of histamine causes vasodilatation, mucous secretion and tissue fluid exudation.
A late-phase response, which begins 4–6 hours after contact with the allergen, follows the immediate reaction. This response can continue for several days and is caused by the release of chemotactic mediators, which promote the selective recruitment of inflammatory cells, mainly eosinophils and neutrophils, that infiltrate the tissue producing an inflammation.

1.1.1.2 Non-IgE mediated reactions

Cell-mediated reactions, also known as delayed hypersensitivity, develop slowly, reaching a peak at approximately 48 hours after the ingestion; the delayed onset of symptoms makes the clinical association with the offending food difficult.
Celiac disease, resulting from sensitivity to gliadin found in grains such as wheat, rye and barley, is an example of a non-IgE mediated gastrointestinal reaction to food.
Cell-mediated responses are the most uncommon form of food allergy[10] and they are classified as of small clinical relevance, when compared to other immune-mediated adverse food reactions.

1.2 Food Allergy diagnosis

The **double-blind placebo-controlled food challenge (DBPCFC)** is regarded as the gold standard for food allergy diagnosis. This test involves the delivery

of increasing amounts of the suspected allergenic food (*verum*) and a placebo food to the patient. The *verum* and placebo challenges are conducted in random order and the *verum* is disguised in a test food matrix. Both the patient and the physician are blinded for the sequence of the challenges. Even though this test allows to obtain a reliable diagnosis of food allergy, it is conducted only in a limited number of centres, because it is expensive, time-consuming and it also carries some risks on severe allergic reactions.

Several *in vivo* and *in vitro* tests allowing to detect the sensitization to an allergen have been developed.

Skin prick test (SPT) is the most commonly used *in vivo* test: the skin of the patient is pricked with a needle containing food extracts or little amounts of purified allergens; a negative control is obtained pricking the skin with a needle containing a saline solution. Hypersensitivity can be diagnosed if an immuno-response is observed: if they are present, mast-cells sensitized with specific IgEs are activated via allergen cross-linking of the latter; this activation results in release of mediators (primarily histamine) which induces a wheal and flare reaction of the skin. Food allergens that elicit wheal diameters at least 3 mm larger than the negative control are considered to be positive results. The SPT has a high negative predictive value (>95%), with a negative result essentially excluding IgE-mediated food allergy. The positive predictive value is, in contrast, less than 50%, thus, an isolated positive test is not definitive for food allergy diagnosis[11], but can rather be used to confirm a suspected food allergy in the context of a history suggestive for the latter. Larger SPT wheals have been correlated with increased likelihood of positive food challenge[12].

Serum or plasma tests can be used for the *in vitro* measurement of circulating allergen-specific IgE antibodies. These tests are preferred to skin tests in the case of patients with dermatographism, severe skin disease, and those who cannot discontinue use of antihistamines[12]. Most common in clinical routine diagnostics worldwide are the ImmunoCAP® system (Phadia AB, Uppsala, Sweden) and the Immulite system (Siemens Healthcare Diagnostics, Berlin, Germany).

ImmunoCAP is an *in vitro* quantitative assay for the measurement of allergen-specific IgE in human serum or plasma[13]. The response of the test is transformed to IgE concentration, reported as kU_A/l, using a calibration curve

based on Word Health Organization (WHO) reference preparation 75/502 for IgE[14]. The test result is considered positive (i.e. the person is sensitized) for a certain food, if specific IgE levels are higher than 0.35 kU_A/l.
The ImmunoCAP assay involves the use of a capsule with a solid phase of a cellulose derivative to which allergens are covalently bound.
The steps involved in an analysis are depicted in **Figure 1.4**.

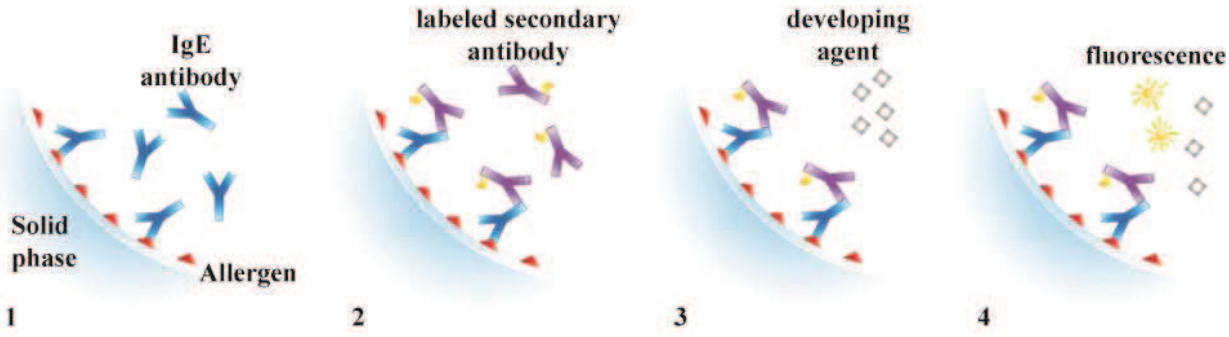


Figure 1.4 Steps of an ImmunoCAP analysis.

1. The allergen, covalently coupled to the solid phase, reacts with the specific IgE in the patient serum.
2. After washing away unbound IgE, enzyme-labelled antibodies against IgE are added to form a complex.
3. After incubation and washing the bound complex is then incubated with a developing agent.
4. After stopping the reaction, the fluorescence of the eluate is measured. The higher the fluorescence, the more specific IgE is present in the sample.

It must be noticed that the presence of specific IgE antibodies is a necessary but not sufficient requirement for allergenicity to an antigen. The latter must in fact have the capacity to cross-link IgEs, thereby causing the release of mediators from basophilic granulocytes or mast cells, to elicit an allergic reaction.
Higher levels of specific IgE can be correlated with increased likelihood of allergic reaction with exposure, however, the severity of possible allergic reactions cannot be predicted[12].

For the proper determination of the elicitation potential of allergens, cellular tests, such as histamine release, basophil activation (BAT), or the cellular allergen stimulation (CAST) test, are needed.

1.3 Food Allergy management

There is currently no cure for food allergies; once a food allergy has been diagnosed, the exclusion of the triggering food from the diet is the best way to manage this condition.
In the event of an accidental ingestion, antihistamines help to relieve symptoms in case of mild reactions, while injections of epinephrine may help in the occurrence of severe food allergies: this hormone increases heart rate, tightens the blood vessels, and opens the airways, offsetting the symptoms of the allergic reaction.
Some food allergies may disappear over time; this occasionally happens for eggs, milk, soy and wheat, while allergies to other foods tend to be lifelong (e.g. seafood, peanuts and tree nuts).
Allergen-specific immunotherapy is a type of treatment in which a patient is given increasing doses of an allergen (typically by oral or sublingual administration) with the goal of inducing immune tolerance, namely the ability of the immune system to ignore the presence of one or more food allergens while remaining responsive to unrelated proteins. The treatment is sometimes performed using an allergen that is cross-reactive to the one responsible of the allergy.
Clinical trials have established that immunotherapy can successfully lead to a desensitization, however the effect is not permanent and further studies are needed to determine the most effective dosage and time frame for treatment, degree of protection, efficacy for different ages, severity and type of food allergy responsive to treatment[15].
A recent review by Le and Burks summarizes the current state of art for the application of immunotherapy in the field of food allergy[16].

1.4 EU regulations on food labelling

Since food allergy management most often relies on patient's compliance to avoid suspected foods, clear food labelling is of crucial importance. In December 2014 a new legislation (the EU Food Information for Consumers Regulation No. 1169/2011) came into force; this requires that any of the 14 allergens that are on the regulatory list are to be emphasized on the label of a pre-packaged food, if they are used as ingredients.
Food businesses selling unpackaged products, such as catering outlets, bakeries and bars are also required to provide allergy information on sold food.
The allergens that must be reported are:

- Eggs
- Milk
- Fish
- Crustaceans (e.g. crab, lobster, crayfish, shrimp, prawn)
- Molluscs (e.g. mussels, oysters, squid)
- Peanuts
- Tree nuts (almonds, hazelnuts, walnuts, cashews, pecans, brazils, pistachios, macadamia nuts or Queensland nuts)
- Sesame seeds
- Cereals containing gluten (wheat, rye, barley, oats, or their hybridized strains).
- Soya
- Celery and celeriac
- Mustard
- Lupine
- Sulphur dioxide and sulphites (at concentration of more than ten ppm)

Information about allergenic ingredients have to be located in a single place on the label: the ingredient list; boxes (such as: 'Contains nuts') that provide a short cut to allergen information also given in the ingredients list, are no longer allowed.
The EU regulation 1169/2011 on the provision of information to consumers replaces and combines previous labelling rules deriving from Directive

2000/13/EC regarding labelling, presentation and advertising of foodstuffs and Directive 90/496/EEC on nutrition labelling of foodstuffs and other legislative acts for specific categories of foods. Food business operators have been given three years to ensure a smooth transition towards the new labelling regime for pre-packaged and non-pre-packaged foods.

1.5 Allergen nomenclature

A systematic nomenclature based on the Linnean system is used to identify an allergen: the name of the latter is formed by the first three letters of the genus and the first letter of the species of origin of the allergen, followed by a number, which indicates the allergen in the chronological order of first purification (e.g. Pru p 3 is the third allergen that was purified and identified in *Prunus persica*).
Under the auspices of the World Health Organization and International Union of Immunological Societies (WHO/IUIS), an allergen nomenclature subcommittee was formed.
To be included in the systematic nomenclature, an allergen must satisfy various criteria: the full nucleotide or amino acid sequence must be known and it must be proven that the purified allergen has allergenic activity, both *in vitro* and *in vivo*.
Natural allergens (i.e. any allergen purified from natural source material) may be denoted by the prefix (n) to distinguish them from recombinant allergens, which are indicated by the prefix (r), before the allergen name. Synthetic peptides are indicated by the prefix (s), with the particular peptide residues designated in parentheses after the allergen name (e.g. sPru p 3 (1-10) is a synthetic peptide with the same sequence of the first 10 amino acids, starting from the N-terminal, of the protein Pru p 3).
Several allergen databases can be retrieved online (**Table 1.1**).

Table 1.1 Online Databases for Allergen Nomenclature and Structural Biology[17].

Database	Locator	Remarks
WHO/IUIS Allergen Nomenclature Sub-Committee	http://www.allergen.org	Lists all allergens and isoforms that are recognized by the Subcommittee
Structural Database of Allergenic Proteins (SDAP)	http://fermi.utmb.edu/SDAP	Provides sequence information, PDB-files and programs to analyze IgE epitopes
Food Allergy Research and Resource Program (Farrp)	http://www.allergenonline.com	Focalized on food allergens; provides sequence similarity searches
Protall	http://www.ifr.bbsrc.ac.uk/protall	Focalized on food allergens; provides clinical data
ALLERbase	http://www.dadamo.com/allerbase	
Allergome	http://www.allergome.org	Provides regular updates on allergens from publications in the scientific literature
Central Science Laboratory (York, UK)	http://www.csl.gov.uk/allergen/	
AllFam	http://www.meduniwien.ac.at/allergens/allfam/	Merges the Allergome allergens database with data on protein families from the Pfam database

1.6 Plant food allergens

Food allergens of vegetable origin are mainly proteins belonging to three different super-families (groups of proteins whose members have low sequence identities but whose structures and functional features suggest a probable common evolutionary origin)[18]:

- **Prolamins:** this family includes several important types of allergens of legumes, tree nuts, cereals, fruits, and vegetables, such as the 2S albumin seed storage proteins, the non-specific lipid transfer proteins, and the cereal α-amylase and protease inhibitors.
- **Cupins:** this group comprises allergenic seed storage proteins of the vicilin and legumin type present in soybeans, peanuts, and tree nuts.
- **Pathogenesis-related proteins:** this family includes proteins produced in plants in the event of a pathogen attack[19]. Some of these proteins are antimicrobial, others also exhibit antibacterial, insecticidal or antiviral action. Many allergenic proteins found in wine are grape pathogen-related proteins[20]. Chitinases and thaumatin-like proteins are comprised in this family.

1.6.1 Non-specific Lipid Transfer Proteins (nsLTPs)

Non-specific Lipid Transfer Proteins (nsLTPs) belong to the large superfamily of prolamins, which includes Alpha-Amylase Inhibitors (AAI), Lipid Transfer (LT) and Seed Storage (SS) Proteins. Members of this super-family are highly conserved and widely distributed throughout the plant kingdom.

The most abundant non-specific lipid transfer proteins are extracellular proteins, associated with cell walls, and mainly accumulated in epidermal tissues surrounding aerial organs like fruits and leaves[21, 22, 23].

According to their molecular weight, non-specific lipid transfer proteins can be divided into two groups: $nsLTP_1$ (9 kDa) and $nsLTP_2$ (7 kDa). The $nsLTP_1$ family includes basic polypeptides (pI 8.5–10) with 90–95 amino acid residues, lacking Trp.

These proteins derive their name from their broad lipid-binding capacity observed in *in vitro* assays (the designation “non-specific” stands for the fact

that many different lipids can be bound). This feature is closely related to the peculiar characteristics of the nsLTP fold: a compact domain composed of 4 α-helixes connected by short loops and a non-structured C-terminal tail (**Figure 1.5**). A large internal cavity following the long axis of the molecule is delimited by the 4 α-helixes; the surface of this cavity is coated with the side chain of hydrophobic residues, enabling the lipid binding properties.

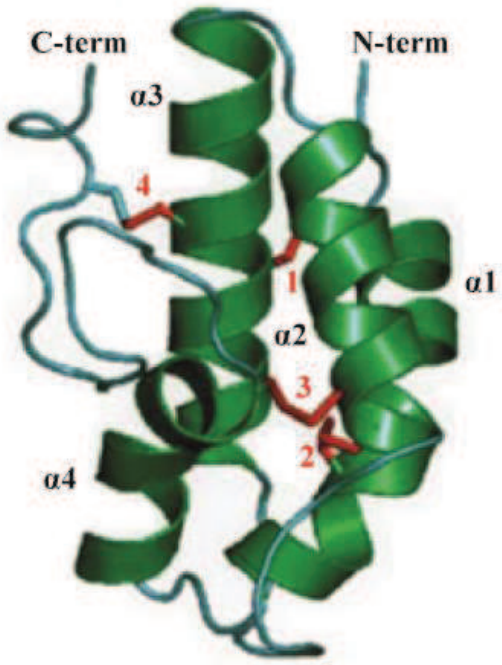


Figure 1.5 3D structure of an nsLTP.

However, the extracellular localization of nsLTPs has led to consider very unlikely an *in vivo* role in intracellular trafficking of membrane lipids.
The possible functions of nsLTPs in plants have been extensively debated in the recent years, and their role in plant defence mechanisms against phytopathogens (bacteria and fungi) seems to be established[24]. This is strongly supported by the fact that the highest expression levels for these proteins have been found in peripheral cell layers surrounding aerial organs, associated with cell wall and cuticle of epidermal tissues. Moreover, *in vitro* activity against phytopathogens, permeabilization of fungal spores and induction by bacterial and fungal infection, were observed[25].

NsLTPs were shown to be among the most important allergens in many fruits and vegetables, mostly affecting sensitized individuals who live in the Mediterranean area[26].

Especially in fruit species belonging to the Rosaceae family several allergenic nsLTPs have been identified and characterized; these proteins have been associated with a variety of adverse reactions with allergic feature, including urticaria, facial angioedema, glottis edema, tongue edema, vomiting, asthma and even anaphylaxis[27].

Actually, 87 allergenic nsLTPs originating from different plant sources are listed in the Allergome database[28] (**Table 1.2**).

Table 1.2 Allergenic nsLTPs present in the plant kingdom.

Botanical order	Plant species	Allergen
Apiales	*Apium graveolens* (celery)	Api g 2 Api g 6
	Daucus carota (carrot)	Dau c 3
	Foeniculum vulgare (fennel)	Foe v 3
	Petroselinum crispum (parsley)	Pet c 3
Asparagales	*Allium ascalonicum* (shallot)	All a 3
	Allium cepa (onion)	All c 3
	Asparagus officinalis (asparagus)	Aspa o 1
	Crocus sativus (crocus)	Cro s 3
Asterales	*Ambrosia artemisiifolia* (short ragweed)	Amb a 6
	Artemisia vulgaris (English mugwort)	Art v 3
	Lactuca sativa (garden lettuce)	Lac s 1
	Helianthus annuus (sunflower)	Hel a 3
Brassicales	*Arabidopsis thaliana* (Thale cress)	Ara t 3
	Brassica oleracea (cauliflower)	Bra o 3
	Brassica rapa (field mustard)	Bra r 3
	Sinapis alba (white mustard)	Sin a 3

Caryophyllales	*Hylocereus undatus* (dragonfruit)	Hyl un LTP
Cucurbitales	*Cucumis melo* (muskmelon)	Cuc m LTP
Ericales	*Actinidia chinensis* (gold kiwi) *Actinidia deliciosa* (green kiwi) *Vaccinium myrtillus* (whortleberry)	Act c 10 Act d 10 Vac m 3
Fabales	*Arachis hypogaea* (peanut) *Cicer arietinum* (chickpea) *Lens culinaris* (lentil) *Phaseolus vulgaris* (kidney bean)	Ara h 9 Ara h 16 Ara h 17 Cic a 3 Len c 3 Pha v 3
Fagales	*Castanea sativa* (chestnut) *Corylus avellana* (hazel) *Juglans regia* (walnut)	Cas s 8 Cor a 8 Jug r 3
Lamiales	*Fraxinus excelsior* (European ash) *Olea europea* (olive tree) *Sesamum indicum* (sesame)	Fra e 7 Ole e 7 Ses i LTP
Malpighiales	*Hevea brasiliensis* (latex)	Hev b 12
Myrtales	*Punica granatum* (pomegranate)	Pun g 1
Pinales	*Cryptomeria japonica* (cypress)	Cry j LTP
Poales	*Hordeum vulgare* (barley) *Oryza sativa* (rice) *Triticum aestivum* (wheat) *Triticum spelta* (spelt) *Triticum turgidum* (durum)	Hor v 14 Hor v 7k-LTP Ory s 14 Ory s 7k-LTP Tri a 14 Tri a 7k-LTP Tri s 14 Tri td 14 Tri td 7k-LTP

	Zea mays (maize)	Zea m 14
Proteales	*Platanus acerifolia* (plane tree)	Pla a 3
	Platanus orientalis (oriental plane tree)	Pla or 3
Rosales	*Cotoneaster lacteus* (clusterberry)	Cot l 3
	Fragaria ananassa (strawberry)	Fra a 3
	Fragaria vesca (wild strawberry)	Fra v 3
	Malus domestica (apple)	Mal d 3
	Morus alba (white mulberry)	Mor a 3
	Morus nigra (black mulberry)	Mor n 3
	Parietaria judaica (pellitory)	Par j 1
		Par j 2
	Parietaria mauritanica (pellitory)	Par m 1
	Parietaria officinalis (pellitory)	Par o 1
	Prunus armeniaca (apricot)	Pru ar 3
		Pru ar 7k-LTP
	Prunus avium (cherry)	Pru av 3
	Prunus davidiana (David's Peach)	Pru da 3
	Prunus domestica (plum)	Pru d 3
	Prunus dulcis (almond)	Pru du 3
	Prunus kansuensis (Chinese wild peach)	Pru ka 3
	Prunus mira (Tibetan peach)	Pru mi 3
	Prunus pérsica (peach)	Pru p 3
	Prunus sargentii (Sargent's Cherry)	Pru sa 3
	Pyrus communis (pear)	Pyr c 3
	Rosa rugosa (rose)	Ros r 3
	Rubus idaeus (raspberry)	Rub i 3
Sapindales	*Citrus clementina* (clementine)	Cit cl 3
	Citrus limon (lemon)	Cit l 3
	Citrus reticulata (tangerine)	Cit r 3
	Citrus sinensis (orange)	Cit s 3

Saxifragales	*Ribes rubrum* (redcurrant)	Rib r 3
Solanales	*Lycium barbarum* (wolfberry) *Lycopersicon esculentum* (tomato)	Lyc ba 3 Sola l 3 Sola l 6 Sola l 7 Sola l 7k-LTP
Vitales	*Vitis aestivalis* (American grape) *Vitis pseudoreticulata* (Chinese Wild Grape) *Vitis vinifera* (grape)	Vit ae 1 Vit ps 1 Vit v 1
Urticales	*Cannabis sativa* (hemp)	Can s 3
Zingiberales	*Musa acuminata* (banana)	Mus a 3

The nsLTP family presents highly conserved sequences and tridimensional structures that enable IgE recognition, promoting cross-reactivity among this type of proteins[29]. Even severe adverse reactions due to immunological cross-reactivity between LTPs derived from botanically unrelated foods have been observed; in particular, in a study conducted by the group of Asero[30], it was shown that most of the investigated patients sensitized to LTPs contained in Rosaceae fruits, were allergic also to nuts, peanut, legumes, celery, rice and corn.

The fact that multiple plant-food sensitization is common for LTP-allergic patients has led some researchers to define an LTP-syndrome[31, 32].

The molecular features of LTPs, such as the characteristic pattern of eight cysteine residues forming four disulphide bridges[33], confer a peculiar compact structure to these proteins, decreasing the probability of degradation due to cooking, industrial processing or digestion, thereby increasing the probability of systemic absorption and severe allergic reactions[27].

The resistance of these proteins to gastroduodenal proteolysis has been proven and explained at molecular level in a study of 2010[34]: the LTPs from peach and barley were subjected to *in vitro* gastrointestinal digestion and they turned out to be resistant to gastric pepsinolysis and were only slowly digested at 1 to 2 out of 14 potential tryptic and chymotryptic cleavage sites under duodenal

conditions. The observed resistance was related with the lack of backbone flexibility in the folded protein, which hinders the interaction with the digestive enzymes.

Stability to digestion may allow a protein to be presented in an immunologically active form to the gastrointestinal immune system[35] and for this reason LTPs are tough to be able to sensitize via the gastrointestinal tract[36].

The fact that they are probably primary sensitizers by ingestion and their association with systemic and severe clinical symptoms, have led to propose LTPs as a model of true food allergens[25].

1.6.1.1 Pru p 3

Pru p 3, the protein accounting for sensitization in more than 90% of patients allergic to peach in the Mediterranean area[37], is one of the most studied and best characterized LTPs and it is considered as the prototypic LTP allergen[38, 39]. Allergenic LTPs identified in other fruits and vegetables, show amino acid sequence identities from 92% to 45% to Pru p 3[40] and this similarity provides the molecular basis for the wide cross-reactivity found among most LTP allergens from plant foods[41]. In many studies, it was shown, through competitive *in vitro* assays, that Pru p 3 is the most potent inhibitory LTP, while inhibition to Pru p 3 by other members of this protein family can only be partially achieved. Thus, it seems that this molecule possesses more epitopes and/or epitopes with higher IgE-binding affinity compared with other LTPs[42].

The sequence of Pru p 3 is constituted by 91 amino acids and it has molecular weight of 9136 Da.

Its three-dimensional crystal structure was described by Pasquato and co-workers in 2005[43]. As for other known structures of LTPs, the main motif of Pru p 3 is the helical compact domain made up of four α-helices. Eight cysteines form four disulphide bridges that contribute to the structural rigidity of the protein by connecting helices H1 to H3 (Cys 3-Cys 50), H1 to H2 (Cys 13-Cys 27), H2 to H4 (Cys 28-Cys 73), and helix H3 to the C-terminal coil (Cys 48-Cys 87) (**Figure 1.6**).

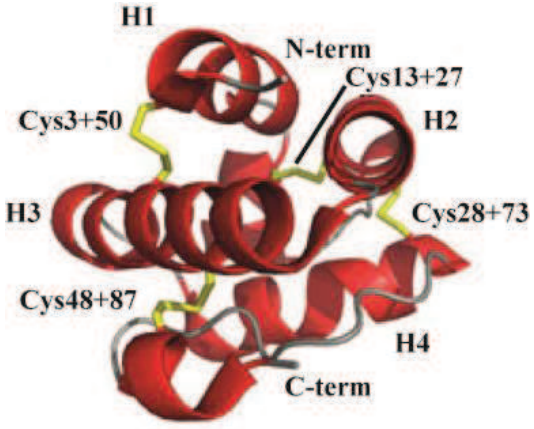


Figure 1.6 3D-structure of Pru p 3.

Pru p 3 has been studied extensively at the biochemical, immunological and clinical level; its IgE-epitopes have been identified by an experimental and modelling based approach.

García-Casado and co-workers screened the specific IgE-binding ability of a library of 10-mer synthetic peptides, overlapping five amino acids, which covered the entire sequence of the protein, by using a serum pool from patients allergic to peach. The main responses were observed for the segments 11-25, 31-45, and 71-80[44].

In 2012, Garino et al. showed that also a particularly effective sequential combination of microwaves/ultrasounds treatments applied on peach juices rich in LTP appeared to be insufficient for the reduction of IgE binding to Pru p 3. In fact, the incubation with specific anti-Pru p 3 serum proved how treating peach peel with microwave at 140 °C and with ultrasound does not eliminate Pru p 3 IgE binding properties[45].

Regarding the animal digestive process, it has been demonstrated that peach LTP is not only resistant to simulated gastrointestinal digestion, with almost one third of protein still intact after few hours, but also that the digested products mostly consist of large polypeptides still linked by the disulphide bridges that keep intact their allergenic properties[46]. Apricot LTP, on the other side, being more susceptible to gastrointestinal digestion, is also less allergenic[47].

1.7 References

[1] Bruijnzeel-Koomen C., Ortolani C., Aas K., Bindslev-Jensen C., Björkstén B., Moneret-Vautrin D., Wüthrich B. Adverse reactions to food. European Academy of Allergology and Clinical Immunology Subcommittee. Allergy. 50(8), 623-635, **1995**.

[2] Husain Z., Schwartz R.A. Food allergy update: more than a peanut of a problem. Inter. J. Dermatol. 52, 286–294, **2013**.

[3] Mosby's Medical Dictionary: atopy

[4] Mills C., Shewry P.R. Plant Food Allergens. Edited by Blackwell Science Ltd, **2004**.

[5] Lee J.K., Vadas P. Anaphylaxis: mechanisms and management. Clin. Exp. Allergy. 41(7), 923-938, Jul **2011**.

[6] Bannon G. A., Ogawa T. Evaluation of available IgEbinding epitope data and its utility in bioinformatics. Mol. Nutr. Food Res. 50, 638–644, **2006**.

[7] Rindsjő E., Scheynius A. Mechanisms of IgE-mediated allergy. Exp. Cell. Res. 316, 1384-9, **2010**.

[8] Taylor S.L., Moneret-Vautrin D.A., Crevel R.V.R., Sheffield D., Morisset M., Dumont P., Remington B.C., Baumert J.L. Threshold dose for peanut: risk characterization based upon diagnostic oral challenge of a series of 286 peanut-allergic individuals. Food Chem. Toxicol. 48, 814-9, **2010**.

[9] Summers C.W., Pumphrey R.S., Woods C.N., McDowel G., Pemperton P.W., Arkwright P.D. Factors predicting anaphylaxis to peanut and tree nuts in patiens referred to a specialist center. J. Allergy Clin. Immunol. 121, 632-8, **2008**.

[10] Burks A.W., Tang M., Sicherer S., Muraro A., Eigenmann P.A., Ebisawa M., Fiocchi A., Chiang W., Beyer K., Wood R., Hourihane J., Jones S.M., Lack G., Sampson H.A. ICON: food allergy. J. Allergy Clin. Immunol. 129(4):906-20, Apr **2012**.

[11] Sampson H.A., Albergo R. Comparison of results of skin tests, RAST, and double-blind, placebo-controlled food challenges in children with atopic dermatitis. J. Allergy Clin. Immunol. 74(1), 26-33, Jul **1984**.

[12] Wang J. Management of the Patient with Multiple Food Allergies. Curr Allergy Asthma Rep. 10(4), 271–277, Jul **2010**.

[13] Yman L. Standardization of IgE antibody assays. J. Int. Fed. Clin. Chem. 3, 198-203, **1991**.

[14] Yman L. Allergy in: The immunoassay handbook (2nd edition) Wild D., Nature Publishing Group, London, UK, 664-680, **2001**.

[15] Nowak-Węgrzyn A., Fiocchi A. Is Oral Immunotherapy the Cure for Food Allergies? Curr. Opin. Allergy Clin. Immunol 3, 214-219, Jun **2010**.

[16] Le U.H., Burks A.W. Oral and sublingual immunotherapy for food allergy. World Allergy Organization Journal 7, 35, **2014**.

[17] Chapman M.D. Allergen nomenclature. In: Lockey R.F., Bukantz S.C., Bousquet J. (Eds.) Allergens and allergen immunotherapy. Marcel Decker, New York. 51–64, **2003**.

[18] Breiteneder H., Radauer C. A classification of plant food allergens. J. Allergy Clin. Immunol. 113 (5), 821-830 May **2004**.

[19] Loon L.C. Pathogenesis-related proteins. Plant Molecular Biology 4 (2-3), 111–116, **1985**.

[20] Waters E.J., Shirley N.J., Williams P.J. Nuisance Proteins of Wine Are Grape Pathogenesis-Related Proteins. J. Agric. Food Chem. 44 (1), 3–5, **1996**.

[21] García-Olmedo F., Molina A., Segura A., Moreno M. The defensive role of non-specific lipid transfer proteins in plants. Trends Microbiol. 3, 72-74, **1995**.

[22] Marzban G., Puehringer H., Dey R., Brynda S., Ma Y., Martinelli A., Zaccarini M., Van Der Weg E., Housley Z., Kolarich D.E.A. Localization and distribution of the major allergens in apple fruits. Plant Sci. 169, 387-394, **2005**.

[23] Ahrazem O., Jimeno L., López-Torrejón G., Herrero M., Espada J.L., Sánchez-Monge R., Duffort O., Barber D., Salcedo G. Assessing allergen levels in peach and nectarine cultivars. Ann. Allergy Asthma Immunol. 99, 42-47, **2007**.

[24] Carvalho A.O., Gomes V.M. Role of plant lipid transfer proteins in plant cell physiology—A concise review. Peptides 28, 1144–1153, **2007**.

[25] Salcedo G., Sánchez-Monge R., Díaz-Perales A., Pacios L.F. Review. Plant food allergens: peach non-specific lipid transfer protein Pru p 3 as a model. Span. J. Agric. Res. 6 (Special issue J. M. Malpica), 30-37, **2008**.

[26] Borghesan F., Mistrello G., Roncarolo D., Amato S., Plebani M., Asero R. Respiratory allergy to lipid transfer protein. Int. Arch. Allergy Immunol. 147(2), 161-165, **2008**.

[27] Van Winkle R.C., Chang C. The Biochemical Basis and Clinical Evidence of Food Allergy due to Lipid Transfer Proteins: A Comprehensive Review. Clinic. Rev. Allerg. Immunol. 46, 211–224, **2014**.

[28] http://www.allergome.org/

[29] Hauser M., Roulias A., Ferreira F., Egger M. Panallergens and their impact on the allergic patient. Allergy, Asthma, Clin. Immunol. 6, 1−14, **2010**.

[30] Asero R., Mistrello G., Roncarolo D., Amato S., Caldironi G., Barocci F., van Ree R. Immunological cross-reactivity between lipid transfer proteins from botanically unrelated plant-derived foods: a clinical study. Allergy. 57(10), 900-906, Oct **2002**.

[31] Pastorello E.A., Robino A.M. Clinical role of lipid transfer proteins in food allergy. Mol. Nutr. Food Res. 48(5), 356-362, Oct **2004**.

[32] Palacin A., Bartra J., Munoz R., Diaz-Perales A., Valero A., Salcedo G. Anaphylaxis to wheat flour-derived foodstuffs and the lipid transfer protein syndrome: a potential role of wheat lipid transfer protein Tri a 14. Int. Arch. Allergy Immunol. 152, 178–183, **2010**.

[33] Zuidmeer L., van Ree R. Lipid transfer protein allergy: primary food allergy or pollen/food syndrome in some cases. Curr. Opin. Allergy Clin. Immunol. 7(3), 269–273, **2007**.

[34] Wijesinha-Bettoni R., Alexeev Y., Johnson P., Marsh J., Sancho A.I., Abdullah S.U., Mackie A.R., Shewry P.R., Smith L.J., Mills E.N. The structural characteristics of nonspecific lipid transfer proteins explain their resistance to gastroduodenal proteolysis. Biochemistry 49(10), 2130-2139, Mar **2010**.

[35] Mills E.N., Jenkins J.A., Alcocer M.J., Shewry P.R. Structural, biological, and evolutionary relationships of plant food allergens sensitizing via the gastrointestinal tract. Crit. Rev. Food Sci. Nutr. 44(5), 379-407, **2004**.

[36] Asero R., Mistrello G., Roncarolo D., de Vries S.C., Gautier M.F., Ciurana L.F., Verbeek E., Mohammadi T., Knul-Brettlova V., Akkerdaas J.H., Bulder I., Aalberse R.C. van Ree R. Lipid transfer protein: A pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. Int. Arch. Allergy Immunol. 122, 20–32, **2000**.

[37] Wijesinha-Bettoni R., Alexeev Y., Johnson P., Marsh J., Sancho A.I., Abdullah S.U., Mackie A.R., Shewry P.R., Smith L.J., Mills E.N. The structural characteristics of nonspecific lipid transfer proteins explain their resistance to gastroduodenal proteolysis. Biochemistry 49(10), 2130-2139, Mar **2010**.

[38] Pastorello E.A., Farioli L., Pravettoni V., Ortolani C., Ispano M., Monza M. The major allergen of peach (Prunus Persica) is a lipid transfer protein. J. Allergy Clin. Immunol. 103, 520-526, **1999**.

[39] Fernandez-Rivas M., Gonzalez-Mancebo E., Rodriguez-Perez R., Benito C., Sanchez-Monge R., Salcedo G., Alonso D., Rosado A., Tejedor M.A., Vila C., Casas M.L. Clinically relevant peach allergy is related to peach lipid transfer protein, Pru p 3, in the Spanish population. J. Allergy Clin. Immunol. 112, 789-795, **2003**.

[40] Salcedo G., Sánchez-Monge R., Díaz-Perales A., Pacios L.F. Review. Plant food allergens: peach non-specific lipid transfer protein Pru p 3 as a model. Span. J. Agric. Res. 6, 30-37, **2008**.

[41] Dìaz-Perales A., Lombardero M., Sanchez-Monge R., García-Selles F.J., Pernas M., Fernandez-Rivas M., Barber D., Salcedo G., Lipid-transfer proteins as potential plant panallergens: cross-reactivity among proteins of Artemisia pollen, Castanea nut and Rosaceae fruit, with different IgE binding capacities. Clin. Exp. Allergy 30, 1403-1410, **2000**.

[42] Zuidmeer L., Salentijn E., Rivas M.F., Mancebo E.G., Asero R., Matos C.I., Pelgrom K.T., Gilissen L.J., van Ree R. The role of profilin and lipid transfer protein in strawberry allergy in the Mediterranean area. Clin. Exp. Allergy 36(5), 666-675, May **2006**.

[43] Pasquato N., Berni R., Folli C., Folloni S., Cianci M., Pantano S., Helliwell J.R., Zanotti G. Crystal structure of peach Pru p 3, the prototypic

member of the family of plant non-specific lipid transfer protein pan-allergens. J. Mol. Biol. 356, 3,684–694, **2006**.

[44] García-Casado G., Pacios L.F., Díaz-Perales A., Sánchez-Monge R., Lombardero M., García-selles F.J., Polo F., Barber D., Salcedo G. Identification of IgE-binding epitopes of the major peach allergen Pru p 3. J. Allergy Clin. Immunol. 112, 599-605, **2003**.

[45] Garino C., Zitelli F., Travaglia F., Colsson J.D., Cravotto G., Arlorio M. Evaluation of the impact of sequential microwave/ultrasound processing on the IgE binding properties of Pru p 3 in treated peach juice. J. Agric. Food Chem. 60(35), 8755-8762, **2012**.

[46] Cavatorta V., Sforza S., Aquino G., Galaverna G., Dossena A., Pastorello E.A., Marchelli R. In vitro gastrointestinal digestion of the major peach allergen Pru p 3, a lipid transfer protein: Molecular characterization of the products and assessment of their IgE binding abilities. Mol. Nutr. Food Res. 54, 1452–1457, **2010**.

[47] Prandi B., Farioli L., Tedeschi T., Pastorello E.A., Sforza S. Simulated gastrointestinal digestion of Pru ar 3 apricot allergen: assessment of allergen resistance and characterization of the peptides by ultra-performance liquid chromatography/electrospray ionisation mass spectrometry. Rapid Commun. Mass Spectrom. 26(24), 2905-2912, **2012**.

2 Food allergens analysis

Food allergies are generally caused by proteins, so for their characterisation at molecular level, methods for protein purification and analysis are often required.

Since a cell contains tens of thousands of different proteins, highly efficient separation methods and also bioinformatic tools that allow the handling of large amounts of data are needed.

Moreover, the concentration range of different proteins may vary enormously: most biological samples contain highly-abundant proteins, which are often not responsible for allergies, while proteins with relevant immunological activity may be orders-of-magnitude less abundant, thus, the methods used for allergen isolation and characterisation need also to be highly sensitive.

2.1 Gel electrophoresis in protein separation

Gel electrophoresis is a commonly used technique for protein separation. In electrophoresis, all molecules are forced to move through the same matrix, driven by an electric field. The ability of a protein to move in the gel (usually constituted by polyacrylamide, cross-linked by methylenebisacrylamide) depends on its size, shape and net charge (**Figure 2.1**).

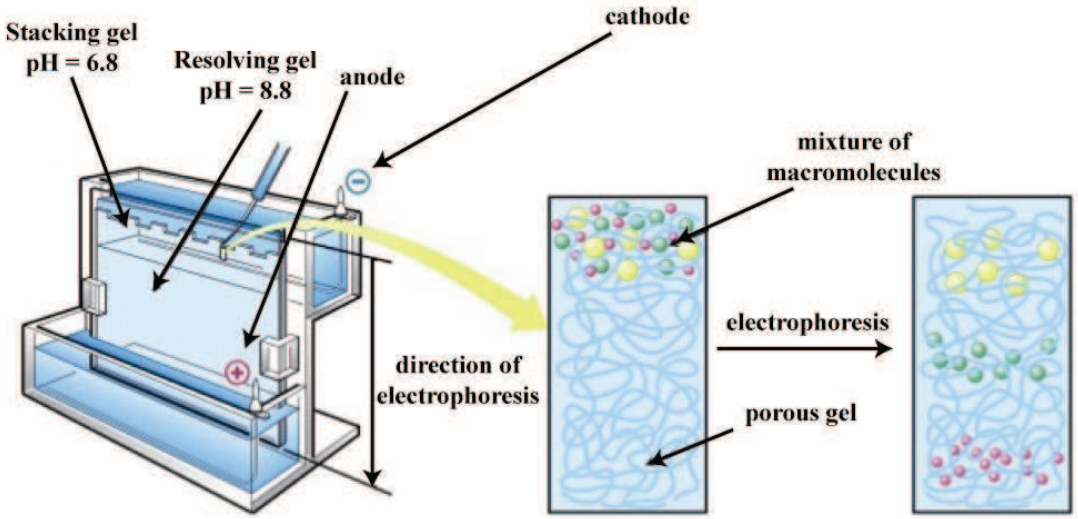


Figure 2.1 Separation of macromolecules by gel electrophoresis.

In SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) prior to the electrophoresis step, the mixture of proteins is dissolved in a solution of sodium dodecyl sulphate (SDS), an anionic detergent that disrupts nearly all non-covalent interactions in the native proteins. Mercaptoethanol or dithiothreitol are also added to reduce disulphide bonds. Anions of SDS bind to the protein at a ratio of about one SDS anion for every two amino acid residues. The complex constituted by SDS and a denatured protein has a large net negative charge that is roughly proportional to the mass of the protein, as the negative charge acquired on binding SDS is usually much greater than the charge on the native protein, which is thus almost insignificant. In addition, the native conformation of proteins is altered when SDS is bound, and most of them assume a similar shape. Proteins can therefore be separated on the basis of their molecular weight only, under these conditions.

After the electrophoretic separation, protein spots can be detected by staining with Coomassie Brilliant Blue, silver, fluorescence dyes, or radio labelling.

To identify the protein corresponding to a band of interest, Edman degradation or, more commonly, mass spectrometry after enzymatic digestion, can be used.

SDS-PAGE is a fast and simple technique, however it is able to resolve only fewer than 100 proteins. Therefore, two-dimensional electrophoresis techniques were developed which allow the resolution of 2000 or even more proteins. Isoelectric focusing (IEF), allowing to separate proteins by their charge, is commonly used in the first dimension. For this purpose, the sample is loaded on a strip containing an immobilized pH gradient (IPG). Then an electric field is connected, leading the proteins to migrate in the IPG until they reach the pH equivalent to their isoelectric point, where their net charge is zero.

In the second dimension, the IPG strip is treated with SDS and applied to SDS-PAGE, where the proteins are further separated by their apparent molecular weight (**Figure 2.2**).

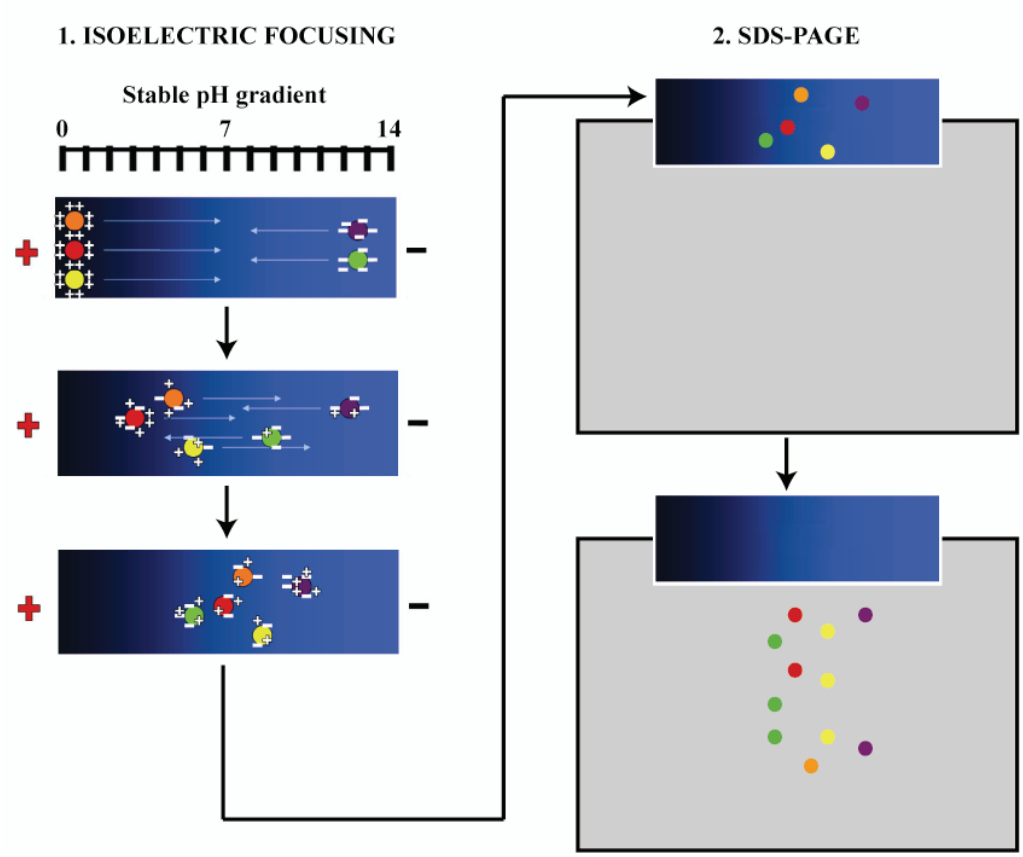


Figure 2.2 Protein separation by 2D-PAGE.

Despite their widespread use, gel based methods still have some limitations. Hydrophobic and poorly soluble proteins as well as basic proteins are not readily accessible by gel electrophoresis and reproducibility, sensitivity and dynamic range for quantification are rather low[1]. Moreover, quantitation is impeded by the limited dynamic range of the available detection methods. Finally, highly-abundant proteins may hinder the analysis of proteins present in low-abundance and often have to be removed from the samples prior to analysis.

2.2 Liquid chromatography in protein separation

Liquid chromatography (LC) is another popular separation technique used for proteins; this method allows the resolution of mixtures of molecules as a consequence of a differential interaction between the latter and a system formed by a mobile and a stationary phase. Depending on the nature of these two

phases, various physical or chemical properties of the analytes can be exploited to achieve their separation.
In **Table 2.1** the chromatographic techniques that are most frequently used in the separation of proteins are summarized.

Table 2.1 Common LC techniques used for proteins or peptides.

Method	Remarks
Reversed phase chromatography (RP-LC)	Separation based on hydrophobicity; more hydrophobic species elute later as they better interact with the stationary phase.
Ion-exchange chromatography (IEC)	Separation based on affinity to the ion exchanger.
Size-exclusion chromatography (SEC)	Separation based on molecular size; larger molecules elute more rapidly than smaller ones.
Affinity chromatography (AC)	A bioselective ligand chemically bound to an inert matrix retains the target component, which has selective affinity to the ligand.

2.3 Mass spectrometry (MS)

Mass spectrometry allows to obtain information about the molecular weight of analytes with very high sensitivity. Over the years several different types of mass spectrometers have been developed, but they all share the same main components (**Figure 2.3**).

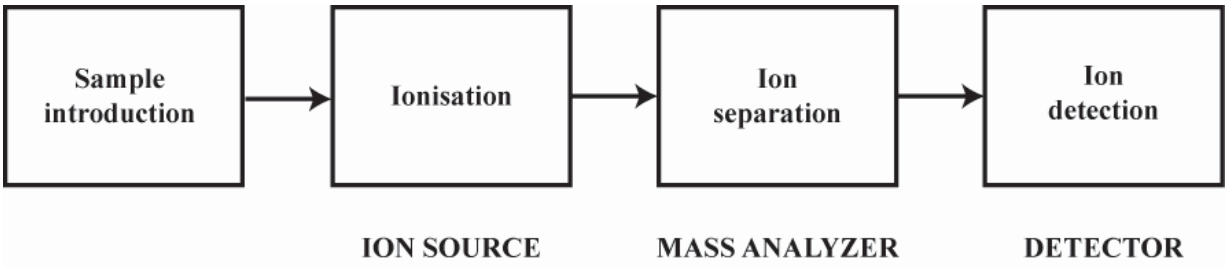


Figure 2.3 Components of a mass spectrometer.

To be visualized through mass spectrometry, a molecule has to be transformed into an ion and to be transferred to the gas-phase; these processes take place in the ion source. The formed charged species are then separated according to their mass to charge ratio (m/z) in the mass analyser and finally detected.
For a long time the use of MS has been restricted to the analysis of small and thermo-stable compounds. The development in the late 1980s of soft ionization techniques like ElectroSpray Ionization (ESI)[2] and Matrix Assisted Laser Desorption Ionization (MALDI)[3] has extended the scope of mass spectrometry to large biomolecules like proteins.
In **ESI** sources the sample is introduced in solution (e.g. in water) and then injected into a capillary held at a potential of 2-5 kV relative to the wall of the surrounding chamber.
As a result, a very fine spray of solution droplets with the same polarity as the capillary is produced; the size of these droplets decreases due to the evaporation of the solvent and when the charge on their surface becomes too high (Rayleigh limit) smaller droplets are formed. Finally charged gas phase ions are released (**Figure 2.4**).

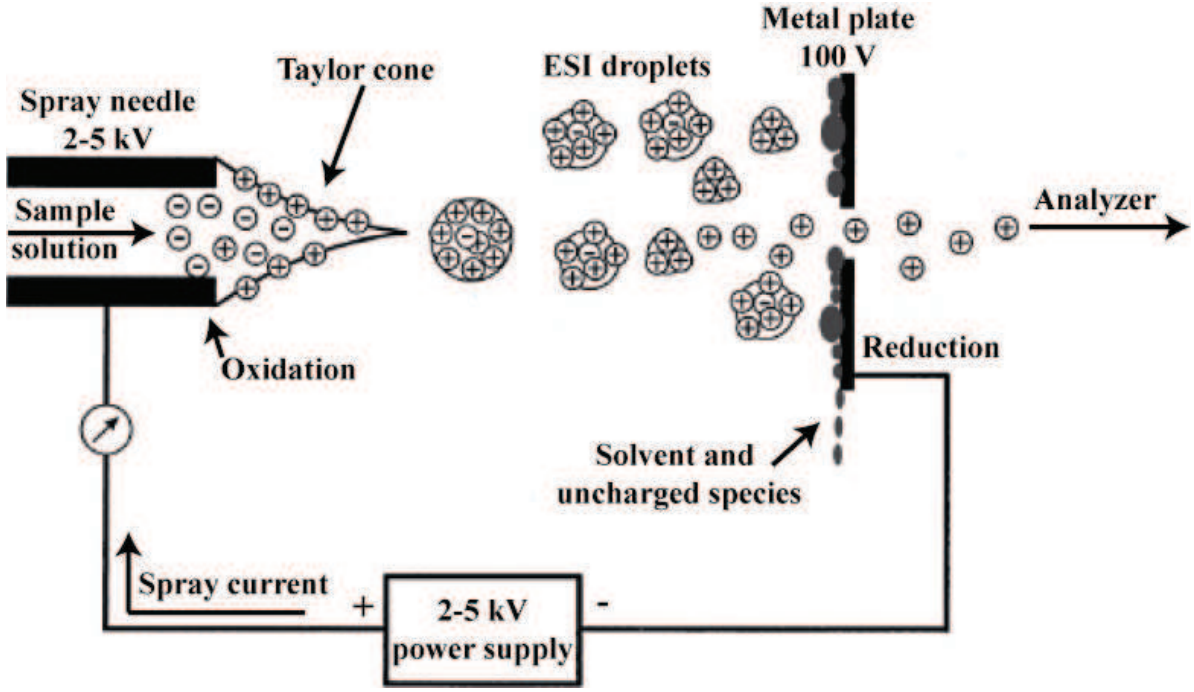


Figure 2.4 Ion formation in an ESI source.

In positive ionization mode, a trace of formic acid is often added to aid protonation of the sample molecules. Depending on their chemical properties and size, these ions can have one or more charges: typically, a protein will approximately carry one charge per thousand Dalton; in a spectrum of a big protein, signals with a bell shaped distribution, relative to several different charge states, can be found.

The ability of ESI sources to generate multiply charged ions allows the detection of proteins also using analysers with limited mass range, like quadrupoles. Another relevant advantage of ESI sources is that they can be easily coupled on-line with high-performance separation techniques such as capillary electrophoresis and HPLC.

Different physical principles can be employed to obtain the ion separation in the analyser. In a **quadrupole** oscillating electrical fields, applied to four parallel rods, are used to selectively stabilize or destabilize ions inside the analyser; for a certain field intensity, only ions within a particular mass range, exhibiting oscillations of constant amplitude, can reach the detector (**Figure 2.5**). The applied field is varied in time to perform a scan of the different ions formed by the analysed sample.

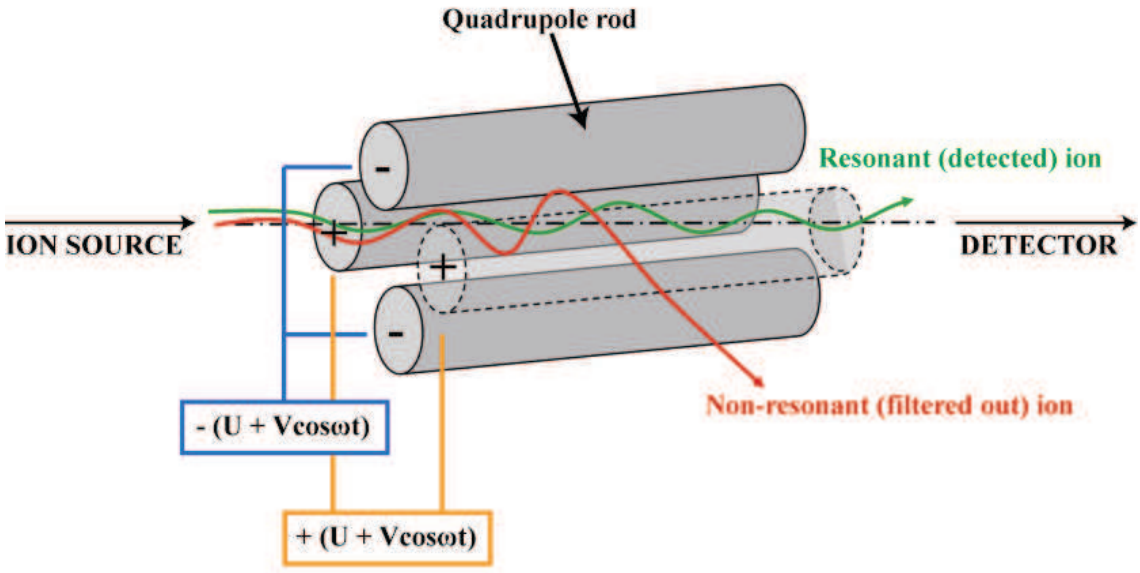


Figure 2.5 Quadrupole mass analyser.

In an **Orbitrap** analyser ions are forced to move in circular paths around an electrode with a frequency depending on their m/z value; the produced image current is recorded on the outer split electrodes (**Figure 2.6**). The signals are amplified and a fast Fourier Transformation is used to convert the registered frequencies into a mass spectrum.

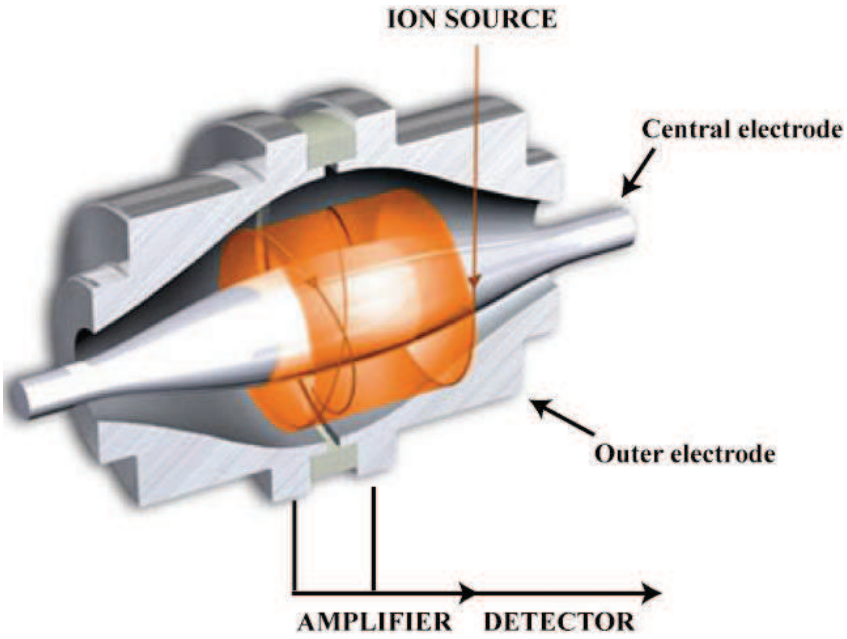


Figure 2.6 Orbitrap mass analyser.

2.3.1 Molecular weight determination through MS

Depending on the resolution of the used mass analyser, the Monoisotopic or the Average masses of the analytes can be determined:

- **Monoisotopic mass:** the sum of the exact or accurate masses of the lightest stable isotope of the atoms in a molecule. This value is obtained from spectra acquired using high resolution mass analysers (e.g. LTQ-Orbitrap).
- **Average mass:** the sum of the average of the isotopic masses of the atoms in a molecule. Equivalent to taking the centroid of the isotopic distribution for all the atoms constituting the analysed molecule. This value is obtained from spectra acquired using low resolution mass analysers (e.g. quadrupoles).

Given the m/z value for an n-fold protonated ion, the corresponding molecular weight is determined as:

$$MW = (m/z * n) - nH^+$$

If n is not known, a mathematical system involving the m/z values of two consequent ions formed by the same analyte, must be applied:

$$MW = ((m/z)_a * n) - n_aH^+$$
$$MW = ((m/z)_b * n) - n_bH^+$$
$$a = b+1$$

Various computational software perform this calculation using the m/z values relative to all the multiple charged ions formed by the same macromolecule, to yield the molecular weight of the uncharged specie. This kind of transformation is called deconvolution.

2.3.2 Tandem mass spectrometry (MS/MS)

Soft ionization techniques are capable to ionize polar and non-volatile compounds without significant analyte fragmentation, therefore only

information concerning the molecular weight of the entire molecule are usually obtained. Fragmentation of the analyte is often of crucial importance for the identification of large biomolecules like proteins, as the obtained data give structural insights. Tandem mass spectrometry allows to obtain the desired fragmentation inside the spectrometer, after a first analysis of the entire analytes. Collision-induced dissociation (**CID**) is the most widely applied fragmentation method for MS/MS analysis: it involves the collision, inside a chamber located after the first mass analyser, of the precursor, selected through the latter, with an inert gas, such as xenon or argon; the produced fragments are then separated in a second mass analyser and detected. The obtained data provide structural information about the precursor ion (**Figure 2.7**).

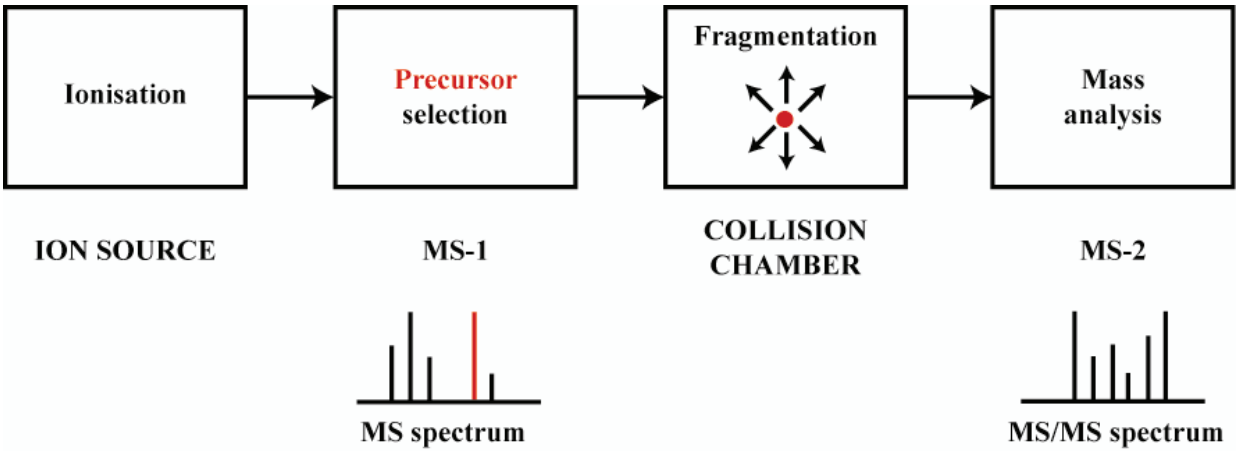


Figure 2.7 Flowchart for a tandem MS analysis.

Furthermore, the recent development of the new "soft fragmentation" techniques of electron-capture dissociation (**ECD**) and electron-transfer dissociation (**ETD**) and higher-energy C-trap dissociation (**HCD**) has greatly enlarged the capabilities of MS/MS strategies[4].

2.3.3 Peptide analysis and protein identification by mass spectrometry

Mass spectrometry is a technique commonly used to identify proteins after gel electrophoresis or liquid chromatography.

A **peptide-mass fingerprint (PMF)** is obtained after digestion of the protein of interest with specific proteases such as trypsin and subsequent MS analysis of the so produced peptide mixture. Prior to the digestion, any disulphide bridges are generally reduced and, to prevent them from forming again, the free thiols are alkylated.

Proteins with known sequences can be identified, starting from their peptide mass fingerprint, by means of computational tools that compare the obtained experimental data with those present in protein or DNA databases. The premise of peptide mass fingerprinting is that every protein will yield a unique set of peptides after digestion and hence unique peptide masses. Identification is accomplished by matching, at high mass accuracy, at least five of the observed peptide masses to the theoretical masses derived from the sequence database. To give an estimation of the likelihood of a protein identification being correct, bioinformatics tools provide a statistical approach to operating the appropriate attribution expressed by the probability-based matching. The calculated score reflects the statistical significance of the match between experimental and theoretical data[5]. If the analysed protein is not present in the database, the best match will probably be the entry with the closest homology, usually an analogue protein from a related species.

A **peptide-fragmentation fingerprinting (PFF)** is obtained by means of MS/MS technologies, fragmenting the intact peptides produced by enzymatic digestion, after the first mass analysis, in a collision cell; the produced fragments are then characterized by a second mass analyser. The resulting molecular masses and fragmentation patterns of the peptides are very specific and allow their characterization. Identification of the peptides, and hence of the originating protein, is also in this case obtained by comparison with sequence databases and the theoretical spectra of peptides producible by the enzymatic digestion of the listed proteins.

When the objective is to study the highest possible number of proteins present in a sample, data-dependent analysis is the most popular acquisition mode. This method consists of an automatic MS/MS fragmentation of the most abundant ions detected during a survey MS scan[6].

For the complete characterisation of proteins not included in databases, *de novo* sequencing, using the data produced by two-dimensional mass spectrometry,

must be attempted; this can be done manually or using sequencing algorithms (e.g. PEAKS, PepNovo).

Peptides fragment in a well-documented manner in the collision cell of a mass spectrometer[7, 8]: respectively in correspondence of the NH-CH, CH-CO, and CO-NH bonds of the peptide backbone (**Figure 2.8**).

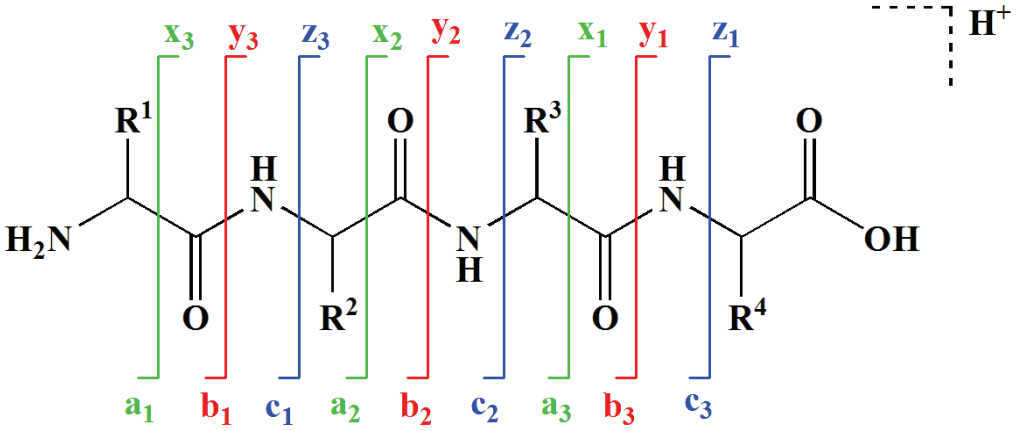


Figure 2.8 Main fragmentation patterns observed in an MS/MS analysis of a peptide.

A neutral and a charged fragment are formed as a consequence of each possible bond cleavage; the charged species can be detected by the mass spectrometer, which measures their mass to charge ratio (m/z). Depending on the chemistry and the proton affinity of the two fragments, the charge can stay on either of them. Hence, six different ions can be formed consequently to cleavages in correspondence of every single amino acid forming the peptide. If the charge is retained on the N-terminal fragment a, b and c type ions can be formed, while x, y and z ions are produced if the C-terminal part constitutes the charged fragment.

The number of residues present in a fragment is reported as a subscript.

The mass difference between two adjacent ions of the same type is indicative for a specific amino acid.

The use of CID fragmentation usually results in cleavage of amide bonds to yield b- and y-type ions, while ECD and ETD fragmentations predominantly produce c- and z-type ions.

2.4 Bottom-up and Top-down approaches in the characterisation of proteins

The choice of the separation method or, more generally, of the strategy applied to the experiments, depends on the scope of the research. Therefore, two complementary approaches are currently used in the study of proteins: the so-called "bottom-up" and "top-down" methods.
The first approach uses one or more proteases to digest the proteins into peptides for subsequent analyses. Instead, the "top-down" strategy directly analyses intact proteins without prior digestion (**Figure 2.9**).

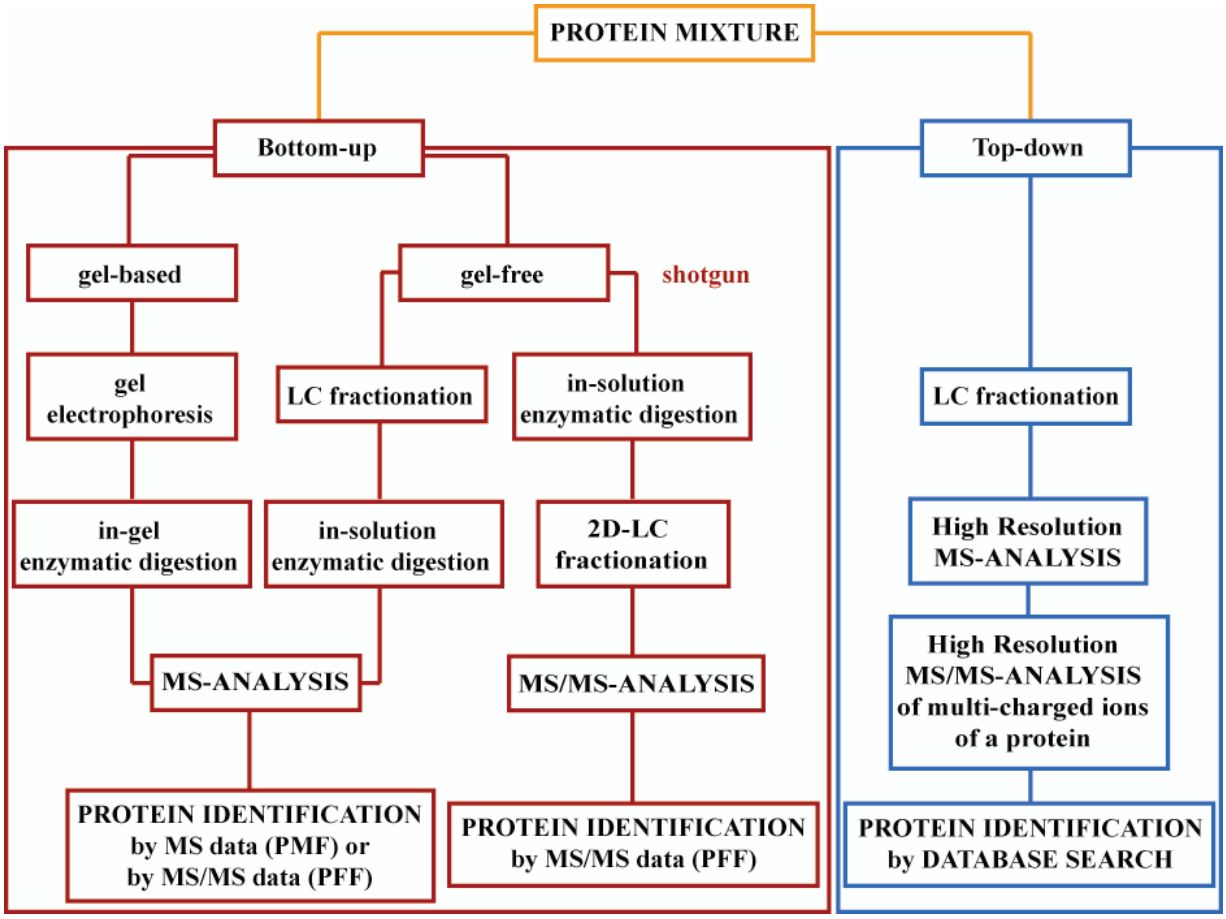


Figure 2.9 Bottom-up and Top-down strategies in the characterization of proteins. The shotgun approach is a kind of Bottom-up technique in which the proteins of a mixture are digested and analysed directly, without prior fractionation.

2.4.1 Shotgun approach

Since the chromatographic separation of entire proteins is frequently problematic due to possible incomplete elution from the column, protein samples are often partially hydrolysed enzymatically prior to chromatography. As a consequence, the sample becomes more complex, however the chromatographic separation and mass spectrometric identification of peptides is well established and straightforward. Single peptides are sequenced using MS/MS technologies and the native proteins are finally identified by database search (**Figure 2.10**).

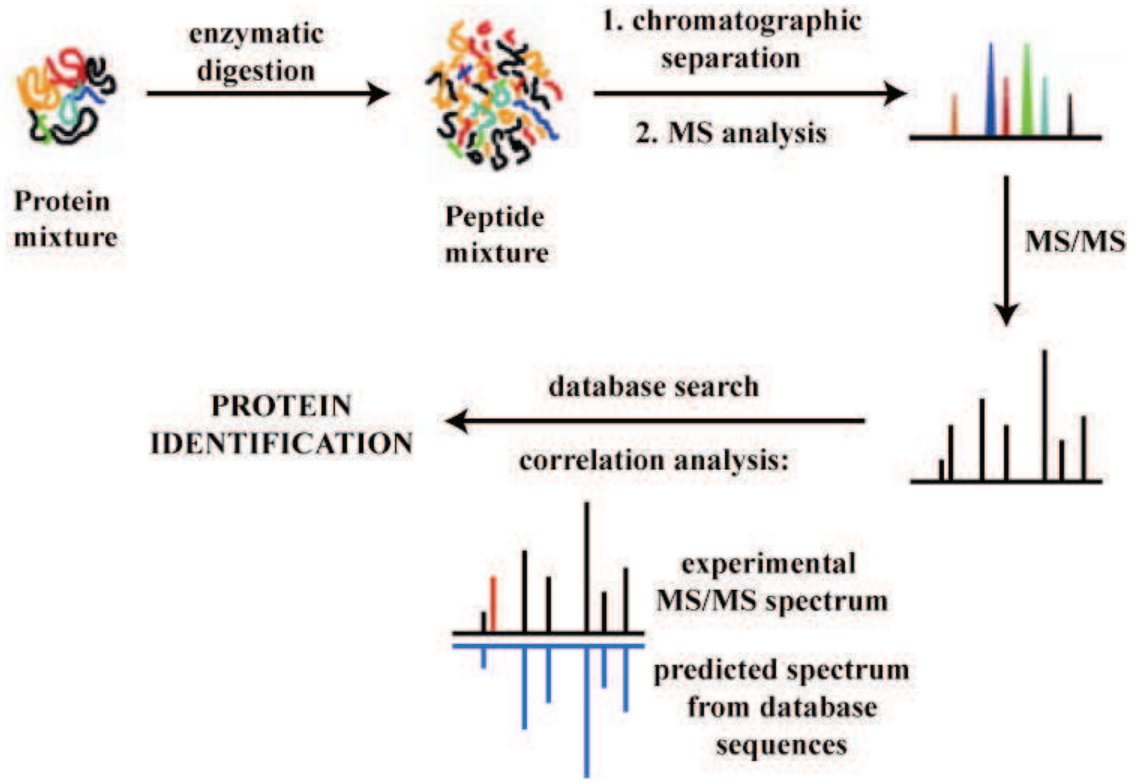


Figure 2.10 Flowchart for a shotgun proteome analysis.

Since complex peptide mixtures are obtained after the enzymatic digestion of protein samples, multidimensional liquid chromatography is usually employed to isolate the single peptides: several separation techniques can be combined, in which each technique utilizes different physicochemical properties of the molecules as a basis for their separation. Commonly used techniques include:

Reverse Phase (RP), strong cation exchange, affinity chromatography and size exclusion.

2.4.2 Top-down approach

Top-down strategy represents the emerging MS-based approach in the study of proteins, providing information on intact protein mass and its amino acid sequence. It involves gas-phase ionization of intact proteins and their direct MS and MS/MS analysis without prior proteolytic digestion. Proteins in complex mixtures are firstly separated into pure single proteins or less complex protein mixtures. MS/MS analysis of the mass-selected multicharged ions of a protein then provides fragment mass values for its structural characterization[9]. High mass accuracy and high resolution mass spectrometers are needed for the successful use of top-down approaches.

2.5 Characterisation of the three-dimensional structure of proteins

Depending on their amino acidic sequence and on their environment, proteins can assume characteristic conformations. Regions in which ordered structures, presenting a repetitive motif, are formed (e.g. α-helix or β-sheet) are said to have a secondary structure, while the tertiary structure of a protein represents its overall spatial conformation.
The three-dimensional structure of a protein is often essential for its biological function, so its characterisation is of crucial importance.
Circular dichroism (CD) is a fast and simple technique, which can be used to detect the presence of secondary structures in a protein or peptide. It involves the differential absorption of the left- and the right-handed component of polarized light by means of chromophores contained the analyte. The profile of a spectrum registered between 190 and 240 nm, region corresponding to the absorption of the peptide bond, depends on the percentages of protein sequence which assume an ordered (helix, sheet or turn) or unordered structure. These percentages can roughly be determined from the spectrum using specific computational software.
Other frequently used methods for the determination of the three-dimensional

structure of proteins are NMR[10], X-ray crystallography[11] and computational prediction techniques[12].

2.6 Food-derived allergens detection

Increased attention has been given in recent years to the detection of allergenic ingredients in food products, both from the food industry and legislative agencies. As a result several analytical methods for the revelation of allergens in foods have been developed.

The sensitivity of the used detection methods must be such that the minimal amount of a food allergen that is capable of triggering an allergic reaction could be revealed.

The determination of the no observed adverse effect level (NOAEL) is highly tricky because the eliciting doses vary widely between individuals; although scientific data on NOAELs are scarce, various studies seem to confirm that analytical methods should be capable of detecting food allergens in the low milligram per kilo range[13].

Several factors can complicate the analytical detection of trace amounts of allergenic ingredients, such as difficulties with their extraction, or the presence of other, often very abundant, components of the food matrix that can mask the allergen; furthermore the recovery of food allergens is dependent on the type of food matrix[14].

Methods designed to reveal the presence of allergenic foods used as ingredients in food products are based either on specific protein or DNA detection. Protein-based methods usually involve immunological techniques, while DNA-based methods rely on the amplification of specific DNA fragments by means of the polymerase chain reaction (PCR). These methods often target a marker indicative for the presence of an offending food in food-products, rather than a specific food allergen. In particular for DNA-based techniques it has to be kept in mind that the presence of DNA of a component in a food product does not guarantee the presence of allergens and vice versa. Thus, these methods can lead to false positive or negative results.

Table 2.2 presents the most common immunological and molecular-biological methods used for food allergens detection.

Table 2.2[15] Commonly used methods for the detection of food allergens.

Methods	Type of method	Specificity achieved by	Target analyte	LOD/ LOQ[a]
RAST/EAST	Immunological	Human IgE binding	Allergenic proteins	1 mg/kg
Immunoblotting	Immunological	Binding to human IgE or antibodies raised in animals	Proteins	2.5-5 mg/kg
Rocket immune-electrophoresis	Immunological	Binding to human IgE or antibodies raised in animals	Proteins	2.5-30 mg/kg
ELISA	Immunological	Binding to human IgE or antibodies raised in animals	Proteins	0.1-2 mg/kg
Dipstick	Immunological	Binding to antibodies raised in animals	Proteins	1 mg/kg
Biosensor	Immunological	Binding to antibodies raised in animals	Proteins	0.5-2 mg/kg
PCR	Molecular biological	Oligonucleotide primers	DNA	1 ppm
PCR ELISA	Molecular biological or immunological	Oligonucleotide primers	DNA	10 ppm
Real-time PCR	Molecular biological	Oligonucleotide primers (and probe)	DNA	2-10 ppm

[a] LOD: lower limit of detection; LOQ: limit of quantification

Food processing often helps to reduce its allergenic potential, as it can cause the breakdown of epitopes in allergenic proteins; however also increases of the allergenic capacity have been sometimes observed as a consequence of food processing: hidden epitopes can in fact be exposed due to the disruption of the tertiary structure of allergens and new epitopes can be formed because of reactions leading to protein modifications.
Also the detection of allergenic ingredients in food is influenced by the changes in protein structure mentioned above. Usually the protein extraction efficiency is lowered after food processing, thus negatively influencing their detection.

2.6.1 Production of antibodies in animal hosts

When an organism is exposed to a foreign protein, the B lymphocytes of its immune system start producing specific antibodies against the introduced antigen. The ability of animal immune systems to produce antibodies capable of binding specifically to antigens can be exploited to manufacture probes for detection of molecules of interest in a variety of research and diagnostic applications.
To produce antibodies able to bind to an allergen, the latter is injected into a laboratory or farm animal (e.g. chickens, goats, guinea pigs, hamsters, horses, mice, rats, sheep); adjuvants are used to improve or enhance an immune response to antigens, as the primary goal is to obtain high titter, high affinity antisera. Antigen-specific antibodies expressed in the serum are then recovered.
Antibodies obtained directly from the blood of an immunized animal are called **polyclonal antibodies**: in an organism, there is an entire population of different types of B cells, which are able to produce different kinds of antibodies.
To obtain **monoclonal antibodies**, which are specific for a single epitope of the antigen, antibody-secreting spleen cells from immunized mice have to be fused with immortal myeloma cells to create monoclonal hybridoma cell lines that express the specific antibody in cell culture supernatant.
The produced antibodies can be easily purified and labelled, as they share a relatively uniform and well-characterized protein structure, except in those portions that determine antigen binding.

A detection signal is needed when using antibodies in biological assays; for this purpose a variety of reagents have been developed to allow labelling of antibodies: enzymes, biotin, fluorophores and radioactive isotopes are commonly employed.
For various applications the immobilization of antibodies on a solid support is required (e.g. immunoassays, biosensors, affinity chromatography); this immobilization can either be direct or mediated by A/G proteins or secondary antibodies[16].

2.6.2 Immunoblotting

2.6.2.1 Western blot

The antibody binding ability of proteins separated by means of SDS-PAGE or 2D-PAGE can be pointed out through immunoblotting. After electrophoresis the proteins are transferred from the gel to a membrane (typically nitrocellulose or PVDF) by means of electroblotting.
To avoid non-specific binding of the antibodies used for detection to the parts of the membrane that do not bind analytes, the latter are blocked by treatment with a dilute solution of protein (e.g. 5% Bovine serum albumin or non-fat dry milk).
The detection of proteins that bind to specific antibodies is usually performed through a two-step process:

- The membrane is first incubated with a dilute solution of primary antibodies, which are generated as a consequence of the exposure of a host species or immune cell cultures to the protein of interest. Solutions containing the sera of patients allergic to the studied antigen can also be used.
- The membrane is rinsed to remove unbound primary antibody, then it is exposed to another antibody, directed to a species-specific portion of the primary antibody. After washing, the bound secondary antibodies can be detected on the membrane, as they are linked to fluorescent moieties or to species capable of interacting with chemiluminescent or staining reagents.

Usually several secondary antibodies will bind to one primary antibody, thus enhancing the intensity of the signal.
The steps of a western blot analysis are summarized in **Figure 2.11**.

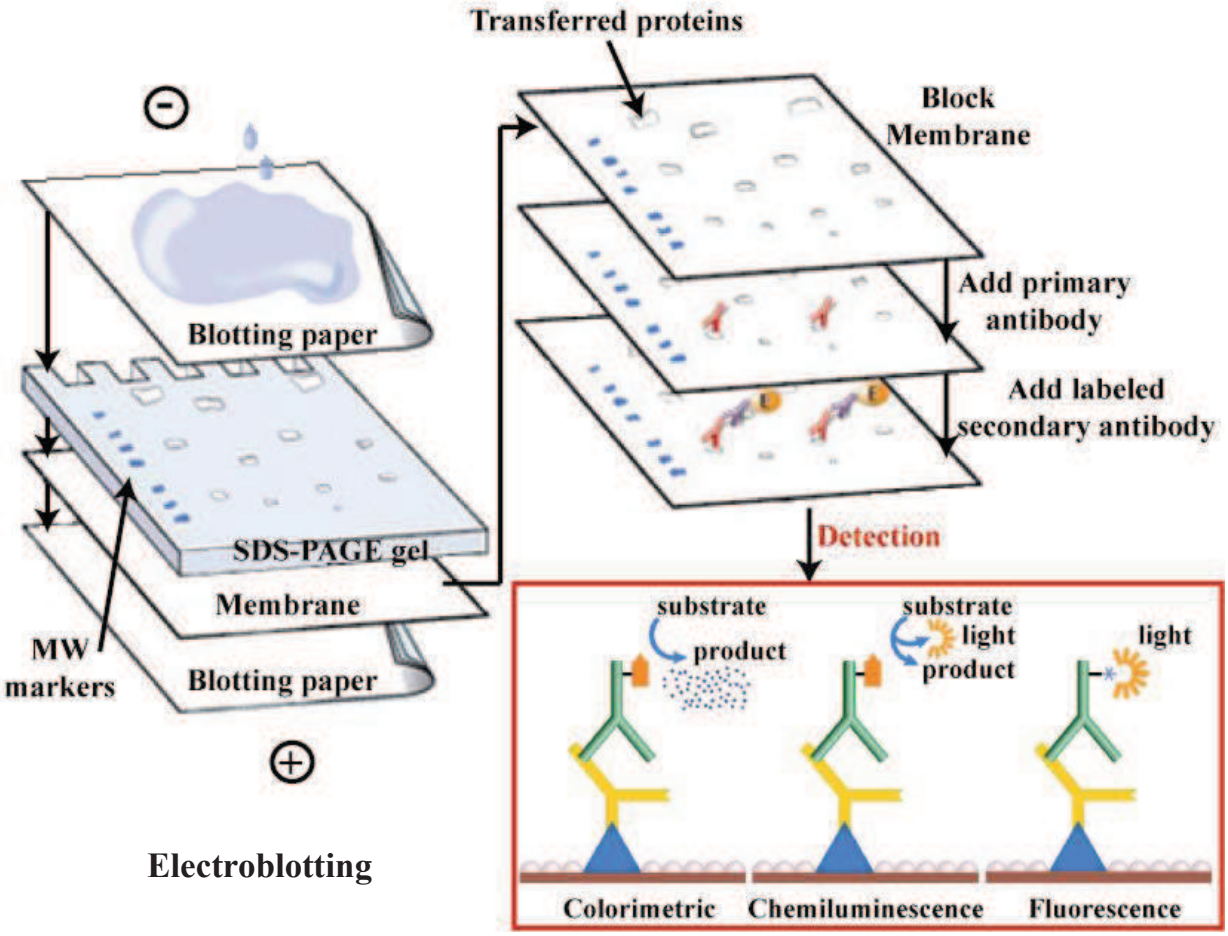


Figure 2.11 Steps of a western blot analysis.

2.6.2.2 Dot blot

To investigate on the antibody binding ability of purified proteins a dot blot analysis can be performed: little volumes of a solution containing the analyte are applied on a membrane to form a dot. After drying the membrane is blocked and the detection is performed following the procedure previously outlined for the western blot.

2.6.3 Competitive binding assays

The binding ability of two different antigens to a specific antibody can be compared through a competitive assay. This can, for example, be useful to verify if a protein produced by means of recombinant DNA techniques has the same properties and acts as the corresponding natural protein. Also the possible cross-reactivity of homologous allergens deriving from different sources can be pointed out.
Another application of this type of assay consist in testing if a small peptide has the potential to inhibit the binding of an allergenic protein to an antibody and could, thus, be used as a possible vaccine.
To perform a competitive binding assay, the antibody is usually pre-incubated with the potential inhibitor and is then used in immunoblotting or immunoCAP assays in which the blot or CAP contains the other studied allergen; the obtained result is compared with the one produced running the same test without inhibition.

2.7 Detection and characterization of food allergens by mass spectrometry

Mass spectrometry has recently been used to successfully identify and detect allergens in various food matrices; the application of proteomic methodologies for the analysis of food allergens has been termed “allergenomics”[17].
Strategies for allergen characterization and monitoring do not substantially differ from the general procedures for protein identification. The IgE-binding properties of type I food allergens, detected through immunoblotting, can ease their identification by MS.
Known food allergens can be detected through the bottom-up approach: LC-MS analysis of the peptide mixture deriving from protein digestion and subsequent database search.
To discover new allergens mass spectrometry is often combined with immunoblotting: the proteins of an extract undergo electrophoretic separation followed by electro-transferring onto nitrocellulose membrane and IgE immunoblotting analysis with the sera of allergic patients. The found potential

allergens can be identified by in-gel enzymatic digestion followed by MS analysis of the produced peptide mixture.
The quantification of known allergens in food matrices through MS is often performed using specific peptides, produced by the enzymatic digestion of allergenic proteins, as analytical targets. A suitable proteotypic peptide should exhibit the following properties: high efficiency to ionize and to fragment; uniqueness for the target protein; absence of amino acid residues susceptible of chemical modifications; absence of post-translational modifications and of sequences that are prone to missed proteolytic cleavage[18].
A big advantage in the use of MS for the detection and quantification of allergens is the possibility of one-step monitoring of more than a single allergen.
Limits of detection and quantification are in the low-ppb range[18].
Moreover, methods based on MS technology are less prone to problems related to cross-reactivity phenomena, which are very frequent in immunoassays, allowing the unequivocal confirmation of the identity of the tested proteins[5].

2.8 Epitope mapping

Knowledge on epitopes is of key importance for understanding the mechanisms of action of an allergen and its interaction with antibodies. This information can aid in the discovery and development of new therapeutics, vaccines, and diagnostics.
Peptide array immunoassays can be used to localize linear epitopes in the amino acidic sequence of an allergenic protein. These analyses involve the use of libraries of overlapping peptides, that together cover the whole studied allergen, whose antibody binding ability is tested through immunoassays, with procedures analogous to those described above.
As the non-specific binding of small peptides to nitrocellulose membranes is often much less efficient than those given by proteins, the tested peptides have to be covalently fixed to a solid support to avoid their loss during washing steps.
Various technologies have been developed to obtain the needed bound peptides: the latter can either be synthesized directly on a membrane (SPOT peptide

synthesis technique[19]) or immobilized on a support (commonly a plastic or glass chip) after their synthesis, through a chemoselective reaction or unspecific covalent binding.
The mapping of conformational epitopes is much more difficult to achieve as it requires investigation on the allergen in its native tridimensional structure and this is generally not present in peptide fragments. **X-ray crystallography** can be used to visualize the binding of the target protein with the antibody, however, many protein-antigen complexes are very difficult to crystallize or can't be crystallized at all.
Information about the spatial interaction between allergen and antibody can also be achieved through **hydrogen–deuterium exchange**: this technique investigates on the solvent accessibility of various parts of the antigen and the antibody, demonstrating reduced values where the interaction occurs.
Single amino acids comprised in an epitope can be identified by **site-directed mutagenesis**: systematic mutations of amino acids are introduced into a protein sequence and then antibody binding is measured and compared with that of the natural allergen. This approach allows the identification of linear and conformational epitopes.

2.9 Obtaining allergenic proteins

In recent years, methods employed for the diagnosis of food allergies are increasingly moving from the use of crude allergen extracts towards component-resolved techniques, in which the specific allergen responsible for sensitization or adverse reaction is searched[20]. This kind of applications requires pure allergens that are also needed to study allergic reactions at a molecular level and as reference materials for the calibration and standardization of quantification methods among different laboratories and operators, as well as for risk assessment in the food industry to detect contamination.
Traditionally, pure allergenic proteins are produced by extraction from biological tissues or by means of recombinant DNA techniques.
To obtain a protein of interest from a biological tissue an extraction of the whole or a part of the proteome is first performed and the target protein is then

purified starting from the extract. The used techniques mainly depend on the nature of the starting biological tissue and usually include defatting, homogenization, centrifugation and desalting.
The recombinant DNA technique involves the introduction of the DNA sequence codifying for the protein of interest into a vector (usually a virus or a plasmid) that is then introduced into a host organism. This host is able to transcript and traduce the inserted gene, producing the target protein. This technique is stereospecific and allows to produce proteins with high yields.

2.10 Recombinant allergenic proteins

Allergenic proteins from the most prevalent allergen sources have been isolated and produced as recombinant proteins, over the last 25 years[21]. These molecules enormously improved the reliability of allergy diagnosis tools, as they allow the production of extracts containing stable and sufficient amounts of allergenic proteins; in contrast, the quality of crude natural extracts and their allergen content is dependent on the source material, which shows intrinsic variations.
Recombinant allergens are also valuable tools in component-resolved allergy diagnostics, in which the specific allergen responsible for sensitization or allergy is searched; unlike natural purified allergens, their recombinant counterparts do not present the risk of being contaminated with other allergens, thus reducing the probability of false positive results.
Furthermore, the development of medicines for immunotherapy, takes advantage from the ability to produce modified allergens by recombinant DNA techniques, as this allows for selective targeting of different facets of the allergic immune response and, therefore, to produce more safe and effective drugs[22].

2.11 References

[1] Beranova-Giorgianni S. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. Trends Anal. Chem. 22, 5, 273-281, **2003**.

[2] Fenn J.B., Mann M., Meng C.K., Wong S.F., Whitehouse C.M. Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64-71, **1989**.

[3] Karas M., Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal. Chem. 60, 2299–2301, **1988**.

[4] Quan L., Liu M. CID, ETD and HCD fragmentation to study protein post-translational modifications. Mod. Chem. Appl. 1, 102, **2013**.

[5] Monaci L., Visconti A. Mass spectrometry-based proteomics methods for analysis of food allergens. Trends Anal. Chem. 28, 5, 581–591, **2009**.

[6]. Gallardo J.M., Ortea I., Carrera M. Proteomics and its applications for food authentication and food-technology research. Trends Anal. Chem. 52, 135–141, **2013**.

[7] Gagnaire V., Piot M., Camier B., Vissers J.P.C., Jan G., Léonil J. Survey of bacterial proteins released in chease: a proteomic approach. Int. J. Food Microbiol. 94, 185-201, **2004**.

[8] Polenta G.A., Calvete J.J., Gonzáles C.B. Isolation and characterization of the main small heat shock proteins induced in tomato pericarp by thermal treatment. FEBS Journal 274, 24, 6447-6455, **2007**.

[9] McLafferty F.W., Breuker K., Jin M., Han X., Infusini G., Jiang H., Kong X., Begley T.P. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. FEBS J. 274, 6256, **2007**.

[10] Billeter M., Wagner G., Wüthrich K. Solution NMR structure determination of proteins revisited. J. Biomol. NMR. 42(3), 155–158, Nov **2008**.

[11] Ilari A., Savino C. Protein structure determination by x-ray crystallography. Methods Mol. Biol. 452, 63-87, **2008**.

[12] Marquez-Chamorro A.E., Asencio-Cortes G., Santiesteban-Toca C.E., Aguilar-Ruiz J.S. Soft computing methods for the prediction of protein tertiary structures: A survey. Appl. Soft. Comput. 35, 398-410, Oct 2015.

[13] Poms R.E., Klein C.L., Anklam E. Methods for allergen analysis in food: a review. Food Addit. Contam. 21, 1, 1-31, Jan **2004**.

[14] Poms R.E., Emons H., Anklam E. In: Koppelman S.J., Hefle S.L. (eds) Detecting allergens in food. Woodhead, Cambridge 348–356, **2006**.

[15] van Hengel A.J. Food allergen detection methods and the challenge to protect food-allergic consumers. Anal. Bioanal. Chem. 389, 111–118, **2007**.

[16] Lee J.E., Seo J.H., Kim C.S., Kwon Y., Ha J.H., Choi S.S., Cha H.J. A comparative study on antibody immobilization strategies onto solid surface. Korean J. Chem. Eng. 30(10), 1934-1938, **2013**.

[17] Akagawa M., Handoyo T., Ishii T., Kumazawa S., Morita N., Suyama K. Proteomic analysis of wheat flour allergens. J. Agric. Food Chem. 55, 6863, **2007**.

[18] Picariello G., Mamone G., Addeo F., Ferranti P. The frontiers of mass spectrometry-based techniques in food allergenomics. J. Chromatogr. A 1218, 7386– 7398, **2011**.

[19] Frank R. The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications. J. Immunol. Methods. 1, 267(1), 13-26, Sep **2002**.

[20] Tuano K.S., Davis C.M. Utility of Component-Resolved Diagnostics in Food Allergy. Curr. Allergy Asthma Rep. 15(6), 534, Jun **2015**.

[21] Marth K., Focke-Tejkl M., Lupinek C., Valenta R., Niederberger V. Allergen Peptides, Recombinant Allergens and Hypoallergens for Allergen-Specific Immunotherapy. Curr. Treat. Options Allergy. 26, 1, 91-106, Feb **2014**.

[22] Valenta R., Hochwallner H., Linhart B., Pahr S. Food allergies: the basics. Gastroenterology. 148(6), 1120-1131, May **2015**.

3 Bottom-up approaches for the detection of LTPs in almond and pistachio

3.1 Introduction

Tree nuts are frequently responsible for IgE-mediated reactions, often associated with acute clinical symptoms and even anaphylactic shock, so the detection and the characterisation of the allergens responsible for these reactions, are of crucial importance.

Almond and pistachio are largely used in sweet bakery products as ingredients and pistachio is as well employed in the production of food of animal origin (e.g. mortadella). Their presence as hidden traces in food, originated by cross-contamination, is particularly critical. Like other nuts, almond and pistachio are currently ingredients that have to be mandatorily declared and highlighted in label, accomplishing to EU Regulation 1169/2011 and following amendments.

Beside allergens that are major seed-storage proteins, such as Legumins, Vicilins and 2S Albumins, also species that are only minor constituents of the protein fraction are increasingly being recognized as relevant allergens in tree nuts. This latter group includes lipid transfer proteins[1].

In several fruit species belonging to the Rosaceae family, nsLTPs have been associated with a variety of adverse reactions with allergic feature. Pru p 3, the major allergen of peach in the Mediterranean area, is one of the most studied and best characterized LTPs and it is considered as the prototypic LTP allergen[2, 3].

Like peach (*Prunus persica*) also almond (*Prunus dulcis*) belongs to the genus Prunus, included in the family Rosaceae, but only few studies on LTP induced allergy in almond are available.

In a case study, an almond LTP has been associated with eyelid edema and contact dermatitis following exposure to almond tree leaves and almond fruit epicarp/mesocarp: immunodetection performed on almond epicarp/mesocarp and tree leaf extracts with an anti-Pru p 3 monoclonal antibody revealed an uptake band of approximately 9 kDa. When immunodetection was performed using the same extracts with the patient's serum and anti-IgE as a secondary

antibody, a band of approximately 9 kDa was observed, indicating the existence of lipid transfer proteins with allergenic potential in almonds[4].
In another work the genes encoding for three almond nsLTPs (Pru du 3.01-3.03) were sequenced[5]; the identified isoallergens have similar molecular weights (about 9 kDa), but different sizes of 117, 123 and 116 amino acids respectively.
However, LTPs from almonds have never been isolated and characterized at protein level yet.
Pistachio (*Pistacia vera*) belongs to the Anacardiaceae family and an LTP may be present in it, as suggested by a study aiming to investigate the IgE cross-reactivity between Rosaceae and non-Rosaceae LTPs, which found that IgE antibodies to Rosaceae LTPs reacted to a broad range of non-related plant-foods, including pistachio[6].
Recently, the sequence coding for an LTP was found in an EST database for pistachio and the corresponding amino acidic sequence was deduced[7].
The extraction and the purification of an almond LTP, followed by its complete chemical characterization using liquid chromatography and mass spectrometry techniques, are here described. The full amino acidic sequence of the protein was identified through a bottom up approach and an in silico evaluation of its potential allergenicity was performed.
The same experimental procedure, applied to pistachio, unfortunately showed to be unsuccessful.

3.2 Results and discussion

3.2.1 LTP extraction and purification from almonds

Commercial almond samples were obtained from the market.
Initial SDS-PAGE analysis of the protein extracts (**Figure 3.1 A**), obtained as described in the experimental part, showed a particularly relevant abundance of high molecular weight species, such as Amandin, the major storage protein in almond, which is composed of two polypeptides with estimated molecular weights of 42-46 and 20-22 kDa linked via disulphide bonds[8] (**Figure 3.1 A**, molecular weight range: 21500 – 45000 Da).

In order to better visualize and isolate the target protein, since LTPs usually have a molecular weight between 7 and 10 kDa, the analysis was narrowed down to those proteins with masses in the appropriate range, following a recently developed approach[9]. A double ultrafiltration step was carried out on the protein extract: the first one was performed by using centrifugal filter devices with membranes having a nominal molecular weight limit of 30 kDa; the filtered solution, containing all the molecules with weights lower than 30 kDa, was recovered and then filtered through devices containing membranes with a nominal molecular weight limit of 3 kDa. The retentate fractions from this last 3 KDa ultrafiltration step were finally recovered, as they contained proteins with molecular weight higher than 3 kDa. In this way, theoretically, all the molecules with masses included in the range 3-30 kDa were isolated from the total protein extract of almond samples, getting rid of low molecular weight contaminants (like polyphenols and salts) and of high molecular weight proteins, which can all interfere with the analysis.

As it is shown in **Figure 3.1 B**, the ultrafiltration process allowed to isolate proteins having molecular weight in the range of interest and to increase their concentration. This procedure was also useful to desalt the solution. In particular, the SDS-PAGE analysis revealed the presence of 3 main protein bands in the range 6500 - 14400 Da (**Figure 3.1 B**: numbered bands).

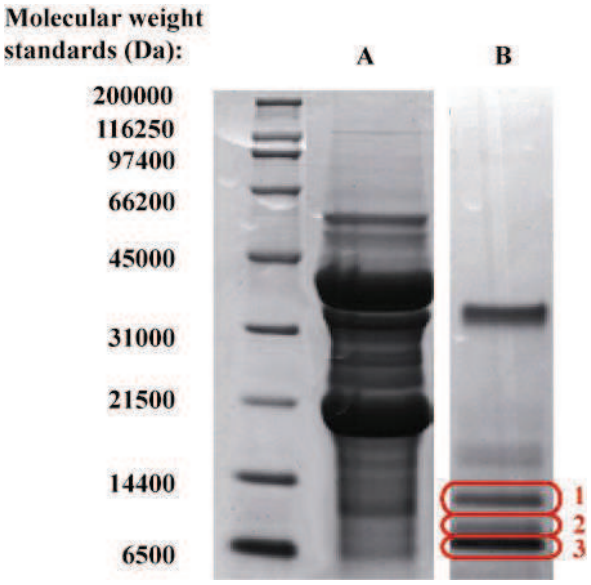


Figure 3.1 SDS-PAGE of the almond protein extract prior (**A**) and after (**B**) the ultrafiltration process.

These bands were subjected to in-gel tryptic digestion and the produced peptide mixtures were directly analysed on an HPLC system coupled with an LTQ-Orbitrap mass spectrometer. Excalibur software was used to search for matches in the Uniprot database for green plants: a significant correspondence was found between the protein band with lowest MW (numbered as 3 in **Figure 3.1 B**) and A7Y7K3, a fragment of 66 AA of an almond LTP, which has been previously characterized only by genomic methods[10]. For the other two bands, the software gave a lower score of attribution for the same fragment and for this reason, they were not considered in this work. The significant matching coverage by the peptides produced from tryptic digestion of the band number 3 (**Figure 3.1 B**) is depicted in **Figure 3.2.**

Figure 3.2 Coverage of the protein fragment A7Y7K3 by peptides (underlined parts) generated after tryptic digestion of the band number 3 (**Figure 3.1 B**). C indicates the carboxymetylation sites.

In order to perform a deep characterization and to obtain the full sequence, the putative LTP protein was then purified through semi-preparative HPLC using a C18 column, starting from the retentate of the ultrafiltered solution. Fractions of 8 ml each were collected during the HPLC gradient and analysed by means of both SDS-PAGE and LC-MS.
The electrophoretic profile of the fraction eluting between the 28th and the 30th minute contained only the band of interest (putatively considered an almond LTP). The LC-UV profile at 240 nm of the purified protein is depicted in **Figure 3.3**.

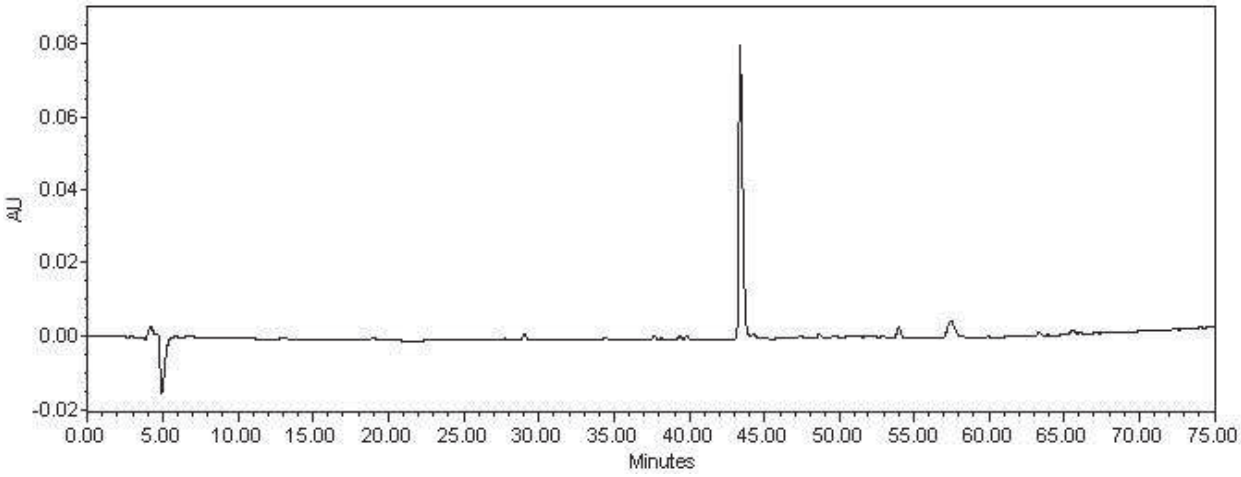


Figure 3.3 HPLC profile (UV detection at 240 nm) of the purified protein.

3.2.2 Almond LTP characterization by mass spectrometric techniques

The monoisotopic MW of the isolated protein, in its oxidized form, is 9579.0, as determined through MS analysis (**Figure 3.4**, ESI-LTQ Orbitrap spetrum).

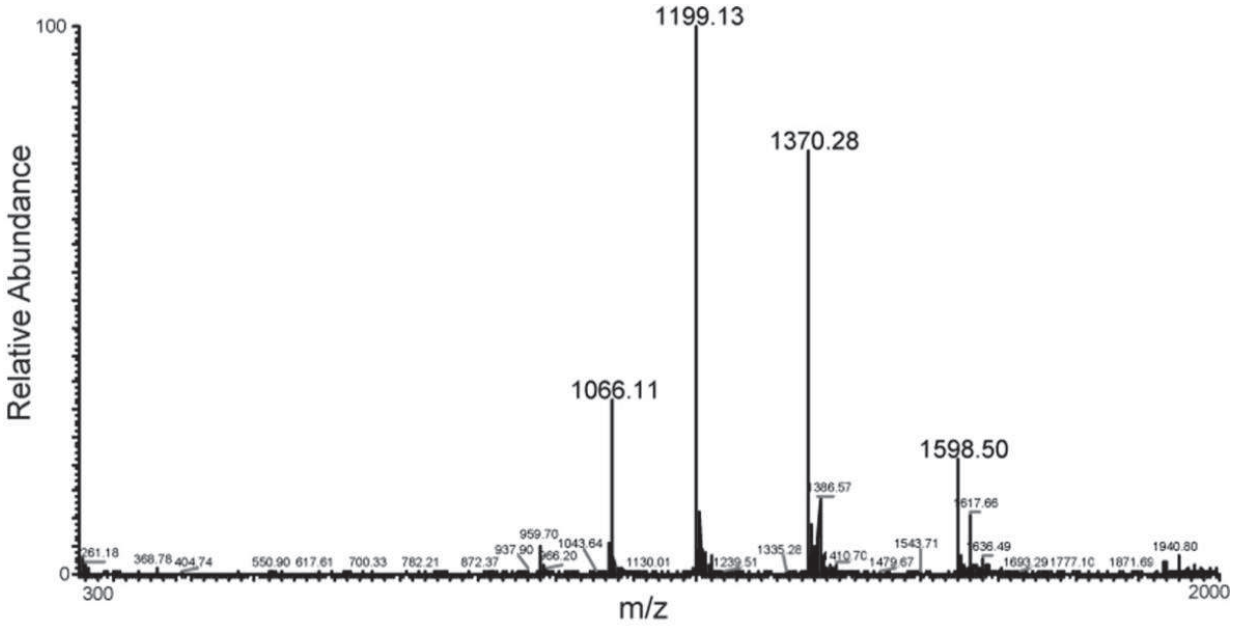


Figure 3.4 Mass spectrum of the 9579.0 Da purified protein (found ions: 1598.50 m/z $[M+6H]^{6+}$, 1370.28 m/z $[M+7H]^{7+}$, 1199.13 m/z $[M+8H]^{8+}$, 1066.11 m/z $[M+9H]^{9+}$).

To completely define its primary sequence, the purified protein was reduced with Dithiotreitol (DTT), alkylated with Iodoacetamide (IAA) and finally digested with trypsin and chymotrypsin. The peptides produced from the enzymatic digestion processes were characterized by means of HPLC-ESI-Orbitrap-MS and the obtained data were analysed using Peaks software (http://www.bioinfor.com/) to compare the experimental information with data relative to known proteins from green plants.
The peptides produced by tryptic digestion gave an excellent alignment with the C-terminal part of M5WKL7[11], an LTP deduced from the genome of peach (*Prunus persica*). For the same protein no match was initially found from the analysis of the chymotryptic digest, however a very good coverage was achieved after the elimination of the cleavage specificity of the enzyme from the software parameters (**Figure 3.5**).

Figure 3.5 Coverage of the protein M5WKL7 by peptides (underlined parts) generated after tryptic and chymotryptic digestion. C indicates carboxymethylation sites.

The protein M5WKL7 completely includes the A7Y7K3 fragment, previously recognized after in-gel digestion.
Table 3.1 shows the list of identified peptides generated by tryptic and chymotryptic digestion of the extracted protein.

Table 3.1 Identification of the peptides generated from the tryptic and chymotryptic digestion of the purified LTP.

Peptide	Position	-10lgP[a]	Calculated MW (Da)[b]	Found MW (Da)	Diagnostic ions (m/z)	Enzyme[c]
LPYPIGPNVDC*NR	102-114	95.36	1513.73	1513.74	757.87 (MH_2^{2+}) 1141.53 (y_{10})	T
SLIGLAGTTPDR	58-69	88.64	1199.65	1199.66	600.83 (MH_2^{2+}) 646.37 (y_6)	T
VSQYVGPC*ISYLK	31-43	85.32	1512.76	1512.78	757.39 (MH_2^{2+}) 937.32 (y_8)	T
TGGAVPVPC*C*NGIR	44-57	74.53	1456.69	1456.70	729.35 (MH_2^{2+}) 1072.49 (y_9)	T
IGLAGTTPDRQGVC*	60-73	73.04	1443.71	1443.72	722.86 (MH_2^{2+}) 831.42 (y_7)	C
YPIGPNVDC*NR	104-114	72.74	1303.60	1303.62	652.81 (MH_2^{2+}) 931.35 (y_8)	T
VRLPYPIGPNVDC*NRIH	100-116	68.92	2019.05	2019.06	674.02 (MH_3^{3+}) 591.45 (y_{10}^{2+})	C

GELVSGLPR	88-96	68.69	926.52	926.54	464.27 (MH_2^{2+}) 529.26 (y_5)	T
LKTGGAVPVPC*C*N	42-54	68.2	1371.66	1371.68	686.84 (MH_2^{2+}) 823.43 (b_9)	C
SITGIKGELVSGLPR	82-96	67.28	1525.88	1525.92	509.64 (MH_3^{3+}) 664.04 (y_{13}^{2+})	T
IGLAGTTPDR	60-69	63.85	999.53	999.56	500.78 (MH_2^{2+}) 646.40 (y_6)	T
C*LVATAK	75-81	59.66	761.41	761.42	381.71 (MH_2^{2+}) 372.90 (b_3/ y_4-NH_3)	T
IGLAGTTPDRQGV	60-72	58.92	1283.68	1283.7	642.85 (MH_2^{2+}) 671.37 (y_6)	C
VATAKSITGIKGELVSGLPRAC*	77-98	57.83	2227.24	2227.26	743.42 (MH_3^{3+}) 760.50 (y_7)	C
LKTGGAVPVPC*	42-52	55.06	1097.59	1097.60	549.80 (MH_2^{2+}) 823.41 (b_9)	C

VATAKSITGIKGELVSGLPR	77-96	54.05	1996.17	1996.20	666.40 (MH_3^{3+}) 927.51 (y_9)	C
VATAKSITGIKGELVSGLPRAC*N	77-99	53.34	2341.28	2341.29	781.43 (MH_3^{3+}) 874.45 (y_8)	C
PIGPNVDC*NRIH	105-116	52.65	1390.68	1390.70	696.35 (MH_2^{2+}) 591.23 (y_{10}^{2+})	C
VIPC*SRVSQYVGPC*ISY	25-41	51.57	1983.95	1984.00	993.00 (MH_3^{3+}) 886.82 (y_{15}^{2+})	C
SITGIKGEL	82-90	48.48	916.52	916.54	459.27 (MH_2^{2+}) 717.34 (y_7)	C
AC*NVRLPYPIGPNVDC*NRIH	97-116	47.83	2364.16	2364.18	789.06 (MH_3^{3+}) 1105.86 (b_{19}^{2+})	C
LPYPIGPNVDC*NRIH	102-116	44.47	1763.88	1763.91	588.97 (MH_3^{3+}) 696.30 (y_{12}^{2+})	C

KTGGAVPVPC*	43-52	43.62	984.51	984.52	493.26 (MH_2^{2+}) 710.35 (b_8)	C
VSQYVGPC*ISY	31-41	41.60	1271.58	1271.6	636.8 (MH_2^{2+}) 696.34 (y_6)	T
AC*NVR	97-101	37.37	618.29	618.30	310.15 (MH_2^{2+}) 388.28 (y_3)	T
SITGIK	82-87	35.79	617.37	617.38	309.69 (MH_2^{2+}) 418.35 (y_4)	T
VSQYVGPC*ISYLKTGGAVPVPC*C*NGIR	31-57	35.48	2951.44	2951.46	984.82 (MH_3^{3+}) 1072.40 (y_9)	T
VGPC*ISY	35-41	33.16	794.36	794.38	398.19 (MH_2^{2+}) 614.20 (b_6)	C
LPYPIGPNVDC*	102-112	32.01	1243.59	1243.6	622.8 (MH_2^{2+}) 951.34 (b_9)	T
PIGPNVDC*NR	105-114	29.67	1140.53	1140.54	571.27 (MH_2^{2+}) 438.07 (y_7^{2+})	T

PYPIGPNVDC*NR	103-114	29.48	1400.65	1400.66	701.33 (MH_2^{2+}) 571.45 (y_{10}^{2+})	T
ELVSGLPR	89-96	29.27	869.50	869.52	435.76 (MH_2^{2+}) 581.40 ($b_6 - H_2O$)	T
AGTTPDR	63-69	25.47	716.34	716.36	359.18 (MH_2^{2+}) 387.24 (y_3)	C
VSGLPR	91-96	20.15	627.37	627.38	314.69 (MH_2^{2+}) 442.23 (y_4)	C
ISYLK	39-43	19.53	622.37	622.38	312.19 (MH_2^{2+}) 510.33 (y_4)	T
VATAKSITGIKGEL	77-90	15.17	1386.81	1386.82	694.41 (MH_2^{2+}) 454.47 (b_5 -NH_3)	C

[a] Peptide feature significance filter: The significance of a peptide is denoted by its -10LogP score. The rule of thumb is to set the cutoff value at 20 which is equivalent to a P-value of 0.01.

[b] Monoisotopic MW

[c] T: trypsin; C: chymotrypsin

C* indicates a carboxamidomethylcysteine

Merging the obtained data, 91 of the 92 C-terminal amino acids of M5WKL7 were covered; the protein constituted by this 92 aa has a calculated monoisotopic MW of 9587.0, which matches perfectly with the experimental value for the purified protein, considering the presence of 4 disulphide bridges in it. This indicates that also the only one amino acid not covered by the peptides obtained from the enzymatic digestion, is actually coincident with the one present in M5WKL7, thus the sequence constituted by the 92 C-terminal amino acids of this protein corresponds to the one of the almond LTP here isolated.

Moreover, the presence of the 4 disulphide bridges was confirmed after the reduction treatment with Tributylphosphine (TBP), as shown by a mass increment of 8 Da (**Figure 3.6**).

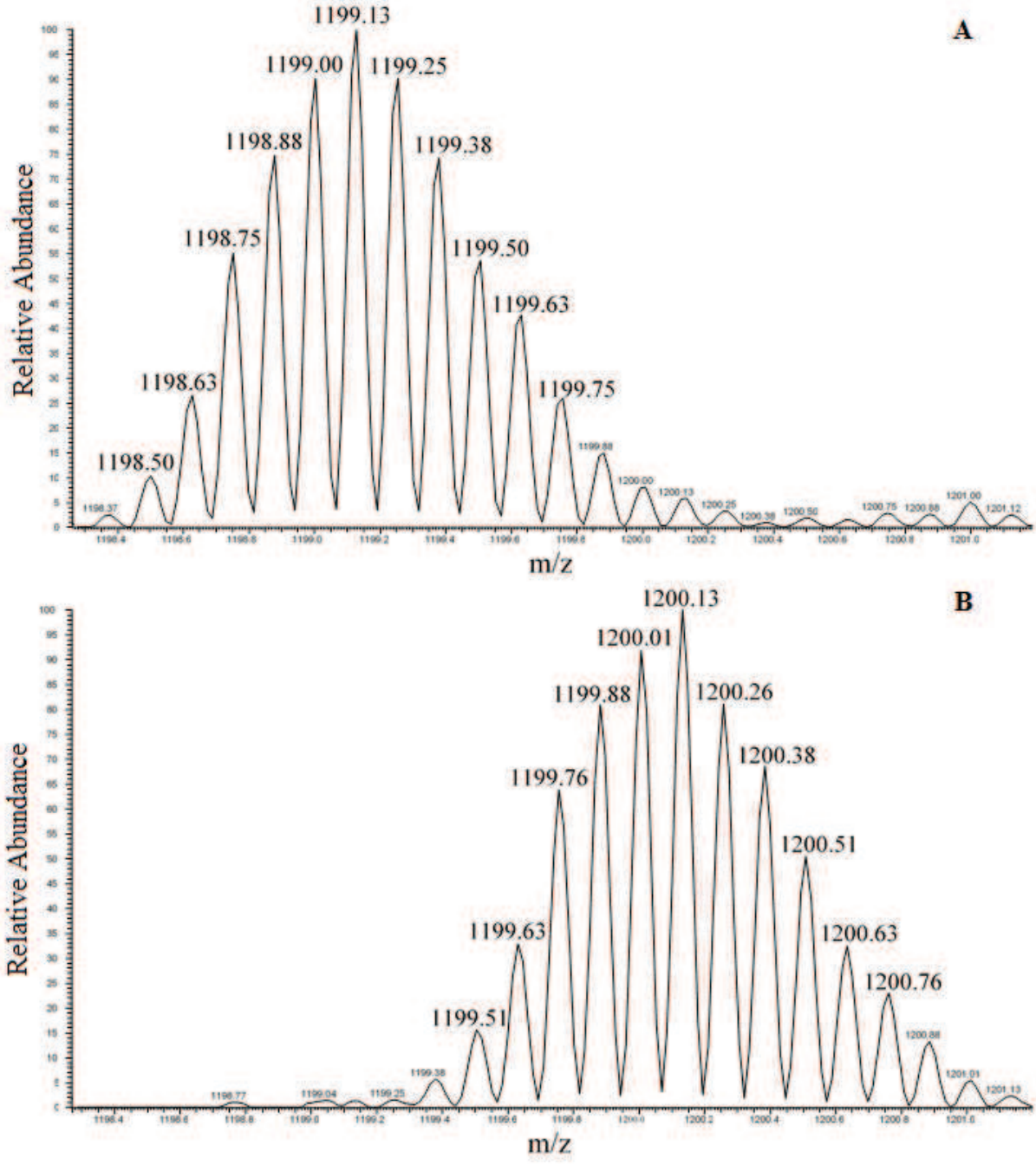


Figure 3.6 Isotopic pattern of the purified LTP octaprotonated ion (**A**: oxidized form; **B**: reduced form).

3.2.3 In silico analysis for the prediction of allergenicity of the almond LTP

According to the CDD tool of the NCBI portal[12], the identified amino acidic sequence can be classified as a member of the nsLTP1 subfamily. Currently

there are 12 entries corresponding to a lipid transfer protein from *Prunus dulcis* in the UniProt database: eleven of them display an almost perfect match with either the first, the second or the third isoform of Pru du 3 previously identified[5], while the new characterized protein is the only one that does not match. Based on our observations it appears how this protein could represent a fourth, previously uncharacterized, LTP isoform of almond.

In order to evaluate the potential allergenicity, we performed an in silico analysis. The FAO/WHO's *Codex Alimentarius* and an Expert Consultation Group have established guidelines to assess potential allergenicity of proteins with bioinformatics in a step-by-step procedure. According to these guidelines[13, 14] a protein should be considered potentially allergenic when there is either:

I) more than 35 % similarity over a window of 80 amino acids in the primary structure of the query protein (without the leader sequence, if any) with an entry known as allergen;

II) a stretch of identity of 6 to 8 contiguous amino acids with the sequence of a known allergen.

There are several available on line software that can be used to predict the allergenicity of a given protein; among them, the Allermatch™ web tool complies with the FAO/WHO criteria given above[15]. The sequence analysis by AllerMatch™ indicated that the identified protein shows a 52.5 % of similarity against Cit s 3 allergen from Citrus sinensis (Q6EV47), and a lower degree of similarity (but anyway higher than the 35 % cutoff) with other LTP allergens from different fruits and vegetables, thus satisfying the criteria of *Codex Alimentarius* for being a potential allergen. Also by using other web tools for the prediction of allergenicity, like AlgPred, AllergenFP and EVALLER™, the outcome was always a probable allergen.

The use of computational techniques allowed to determine that the identified protein is presumably allergenic in a very short time. Moreover, these tools offer the advantage of being much less expensive than routinely used immunoassays like ELISA, which require specific antibodies to be performed.

3.2.4 Identification of the low MW proteins in pistachio extracts

Unroasted pistachio seeds were used for the experiments.
The SDS-PAGE analyses of the total protein extract and of the 3-30 kDa fractionated extract, obtained as previously described for almonds, do not contain sharp protein bands in the range 6500 - 14400 Da (**Figure 3.7**).

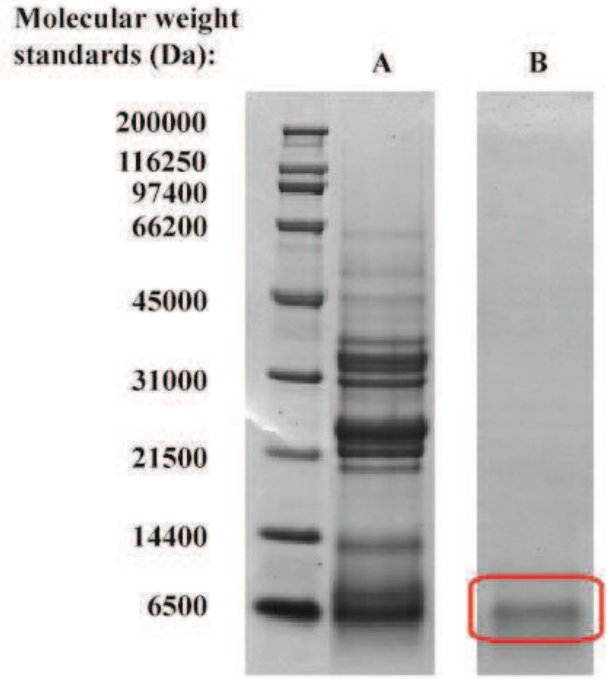


Figure 3.7 SDS-PAGE of the pistachio protein extract prior (**A**) and after (**B**) the ultrafiltration process.

The whole region evidenced in **Figure 3.7**, was excised from the gel and subjected to tryptic digestion. The produced peptide mixture was analysed on an LTQ-Orbitrap and Excalibur software was used to search for matches in the Uniprot database for green plants and to compare the experimental data with the sequence deduced from an EST database, by Garino et al., for the pistachio LTP[7] (**Figure 3.8**).

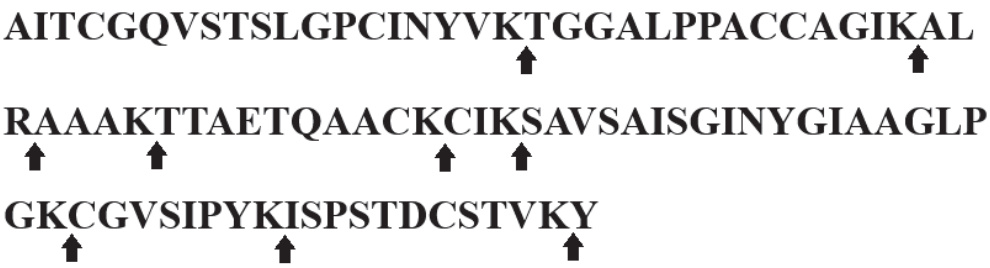


Figure 3.8 Sequence deduced from an EST database for the pistachio LTP; the arrows evidence the expected cleavage sites for tryptic digestion.

The main results of the protein identification are reported in **Table 3.2**. No match was found for the theoretical LTP sequence.

Table 3.2 Matches between the experimental data and the Uniprot database for green plants. Only results for which a score higher than 50 was obtained, were considered reliable and reported.

Protein	Accession n.	MW (Da)	Coverage	Score
Pis v 1 allergen 2S albumin (*Pistacia vera*)	B7P072	17278	36.91%	262.98
Pis v 5.0101 allergen 11S globulin precursor (*Pistacia vera*)	B7SLJ1	53325	23.26%	98.22
11S globulin (*Pistacia vera*)	B2KN55	53231	21.61%	93.21

As shown in **Table 3.2**, the identified proteins, which belong to the families of albumins and globulins and are the major known allergens for pistachio, have molecular weights that are significantly higher than the values that could be expected from the SDS-PAGE analysis. Thus, the found matches are relative to fragments of these proteins, which are probably formed due to the action of natural endo-proteases inside the pistachio seeds.

Since LTPs from different plant sources usually show relevant grades of homology, the presence, in the analysed mixture, of peptides deriving from the digestion of a protein belonging to this family, should result in the partial

identification of at least one LTP present in the database for proteins from green plants. As this was not the case and also no match with the LTP sequence deduced from pistachio EST database was found, we were not able to confirm the presence of LTPs in the analysed nuts.

3.3 Conclusions

The extraction, the purification and the full characterization, at the protein level, of a lipid transfer protein in almond, detailing its complete amino acidic sequence, has been described, with a procedure that combined extraction and targeted ultrafiltration techniques.

The results presented demonstrate the enormous potential of advanced MS techniques for obtaining high quality structural and functional data of allergenic proteins in a short time.

LTQ-Orbitrap-MS allowed to define the unambiguous sequence of a new isoform of almond LTP by the high-resolution measurement of the monoisotopic masses of the purified intact protein and its fragments, obtained by bottom-up approach. Moreover, the potential allergenicity of this previously uncharacterized protein has been confirmed via an *in silico* approach.

As an important issue of food safety, the assessment of the presence of the protein and its amount in different almonds varieties from different geographic origin should be investigated, in order to evaluate the natural biodiversity and the environmental pressure on the LTP expression in seeds.

The experimental procedure used for the isolation of the novel identified LTP from almond was proven not successful to confirm the presence of an LTP in pistachios, indicating that this kind of protein is probably absent or present only in traces in these nuts. However, since LTPs are only minor constituents of the protein fraction of tree nuts, their loss during the extraction process cannot be completely excluded.

3.4 Experimental part

3.4.1 Chemicals

Acetic acid (CH_3COOH), Acetonitrile (ACN), Ammonium hydrogen carbonate (NH_4HCO_3), Calcium chloride ($CaCl_2$), Dithiotreitol (DTT), Formic acid (FA), Hexane, Iodoacetamide (IAA), Methanol (CH_3OH), Potassium chloride (KCl), Potassium dihydrogen phosphate (KH_2PO_4), Protease inhibitor "Sigmafast", Sodium chloride (NaCl), Sodium hydrogen phosphate (Na_2HPO_4), Tributylphosphine (TBP), Trifluoroacetic acid (TFA), Trypsin and α-Chymotrypsin were purchased from Sigma Aldrich (Sigma, St. Louis, MO, USA).
SDS-PAGE Molecular Weight Standard Broad Range, XT MES Running buffer 20X, XT Reducing Agent 20X, XT Sample buffer 4X and Coomassie Brilliant Blue R-250 were purchased from Bio-Rad (Hercules, CA, USA).
Deionised water was obtained using a Milli-Q system from Millipore (Bedford, MA, USA).

3.4.2 Procedure

3.4.2.1 Defatted nut flour production

Almonds and pistachios were ground in liquid nitrogen and defatted by stirring in five volumes (w/v) of hexane at room temperature for 1 h. After drying, the obtained flour was reground and the hexane extraction repeated; from 29.9 g of almonds, 12.3 g of defatted flour were produced, while 8.4 g of flour were obtained starting from 29.8 g of pistachios.

3.4.2.2 Protein extraction

The protein fractions were extracted suspending the defatted flours in a phosphate-buffered saline solution (PBS, NaCl 137 mM, KCl 2.7 mM, Na_2HPO_4 10 mM, KH_2PO_4 1.5 mM; protease inhibitor "Sigmafast" 1 tablet/100

ml; pH 7.4) at 1:50 w/v ratio. The mixtures were stirred at room temperature for 1 h, then the solutions were recovered by sequential filtration through membrane filters (5 µm and 0.45 µm from Millipore). The extracted proteins were visualized through SDS-PAGE.

3.4.2.3 SDS-PAGE

The samples were prepared as follows:
20 µl of each protein extract were dried under nitrogen flux and the obtained residues were re-dissolved in a mixture composed of 16.3 µl of water, 8.3 µl XT sample buffer and 0.4 µl XT reduction buffer.
To obtain the marker, 1 µl of the molecular weight standard was mixed with 5 µl XT sample buffer, 9 µl of water and 0.24 µl XT reduction buffer.
The samples and the marker were incubated at 95°C for 5'.
The obtained solutions were loaded on a Criterion XT Precast gel, 12% Bis-Tris (Bio-Rad); the voltage applied to the Criterion Cell (Bio-Rad) was 150 V. The run lasted 60'.
Gel staining: the staining solution was made up of 0.1% w/v of Coomassie brilliant blue R-250 dissolved in 10% CH_3COOH, 40% CH_3OH and H_2O to 1 l. The gel was covered with the staining solution, where it was allowed to soak for 3 hours. Any dye that was not bound to proteins diffused out of the gel during the de-staining steps, when it was rinsed with a solution of 10% CH_3COOH, 40% CH_3OH, 50% H_2O, changed every 30', since achievement of the desired contrast.

3.4.2.4 Protein fractionation in the range 3-30 kDa

Protein extracts were fractionated in the 3-30 KDa range using Amicon Ultra devices (Millipore), according to the following procedure:
Pre-rinsing: the ultra-filtration membranes in Amicon Ultra devices were pre-rinsed 5 times: each wash was performed with CH_3OH:H_2O (1:1) solution, centrifuging at 5000 rpm and at room temperature for 15'. The devices were stocked in CH_3OH:H_2O (5:95) solution at 4°C until use.

Fractionation below 30 kDa: the total protein extracts from almond and pistachio were loaded onto 30 kDa filter devices (4 ml each), then centrifuged at 7000 rpm, 4°C for 1 h. The filtrates were recovered and used in the following step.

Fractionation above 3 kDa: the solutions containing proteins with MW below 30 kDa were loaded onto 3 kDa filter devices and centrifuged at 7000 rpm, 4°C until the depletion of the whole solution. Two washes with 500 µl of 0.1% FA in H_2O of the retentates were performed with the same centrifugation settings. The retentates were finally recovered using 0.1% FA in H_2O (1 ml for each filter). An SDS-PAGE analysis was performed as described above, using 100 µl of the recovered fractions to prepare the samples, allowing the identification of the bands in the right molecular weight range, which were then subjected to in-gel tryptic digestion.

3.4.2.5 In-gel tryptic digestion

Protein bands were excised from the stained polyacrylamide gel and cut into 1 x 1 mm pieces that were placed into 2 mL tubes.

Destaining: the gel pieces were incubated for 15' in 1 ml of H_2O, then the liquid was discarded and 0.5 ml of ACN 50% in H_2O were added to the tube; after 15' the solution was refreshed and the incubation was repeated for other 2 times.

0.5 ml of ACN were added to the gel pieces and, after the latter became white, the liquid was substituted with 0.5 ml of NH_4HCO_3 0.1 M in H_2O; 5' later 0.5 ml of ACN were added to the tubes and the samples were incubated for further 15'.

The solution was discarded and the gel was dried under N_2 flux.

Reduction and alkylation of the free thiols: 0.5 ml of an aqueous solution containing DTT 50 mM and NH_4HCO_3 0.1 M, were added to the de-stained gel and the samples were incubated for 45' at 56°C. After cooling the liquid was substituted with 0.5 ml of an aqueous solution containing IAA 0.2 M and NH_4HCO_3 0.1 M and the tubes were kept in the dark for 30'.

The gel pieces were incubated for 5' in 1 ml of NH_4HCO_3 0.1 M in H_2O, then the liquid was discarded and 1 ml of ACN 50% in H_2O was added to the tube and the latter was left to stand for 15'.
0.5 ml of ACN were added to the gel pieces and, after the latter became white, the liquid was substituted with 0.5 ml of NH_4HCO_3 0.1 M in H_2O; 5' later 0.5 ml of ACN were added to the tubes and the samples were incubated for further 15'.
The solution was discarded and the gel was dried under N_2 flux.
In-gel digestion: the gel pieces were incubated on ice for 45' in 1 ml of an aqueous solution containing 25 µg/mL trypsin, NH_4HCO_3 25 mM and $CaCl_2$ 2,5 mM.
The liquid was substituted with 1 ml of an aqueous solution containing NH_4HCO_3 25 mM and $CaCl_2$ 2,5 mM and the samples were left overnight at 37°C.
The solution was recovered and 200 µl of NH_4HCO_3 25 mM in H_2O were added to the gel; after 15' with occasional sonication, 200 µl of ACN were added to the tubes and the samples were incubated for further 15'.
The liquid was recovered and the gel was incubated with 200 µl of FA 5% in H_2O for 15', then 200 µl of ACN were added to the tubes and the samples were incubated for further 15'.
The washes produced from the previous steps were combined with the solutions obtained from the digestion and DTT was added to it, to a final concentration of 1 mM. The whole solutions were dried under N_2 flux.

3.4.2.6 Analysis of the peptide mixture formed by the tryptic digestion of the gel bands

The dried samples obtained from in-gel digestion were dissolved in 0.1% FA and analysed with a Dionex Ultimate 3000 micro HPLC (Dionex, Sunnyvale, CA, USA) coupled with an LTQ-Orbitrap mass spectrometer (Thermofisher, San José, CA, USA) equipped with a conventional ESI source.
For the chromatographic separation, a Jupiter 4U Proteo (150 x 0.3 mm, 90 Å) column (Phenomenex, Torrance, CA, USA) was used, and the column oven temperature was set to 25°C. HPLC column flow was 5 µL /min, eluent A

(0.1% FA in H_2O) was kept at 95% for 4', then eluent B (0.1% FA in ACN) was ramped to 50% at 60'.
The ESI source parameters were configured as follows: spray voltage 3.5 kV, capillary voltage 35 V and tube lens 75 V.
MS1 scans were performed in an m/z window from 200 to 1800. For MS/MS in the LTQ a maximum of four precursor ions (most intense) were selected for activation and subsequent MS/MS analysis. CID was performed at 35% of the normalized collision energy (NCE). The obtained data were analysed using Excalibur software.

3.4.2.7 Almond LTP purification

The fractionated protein extract was purified on a semipreparative Jupiter C18 column (250 x 10 mm, 300 Å) (Phenomenex), using a Waters 1525 Binary HPLC Pump system; HPLC column flow was 4 ml/min, eluent A (0.1 TFA in H_2O) was kept at 100% for 10', then eluent B (0.1 TFA in ACN) was ramped to 60% in 30'. The LTP-containing fraction was collected between 28' and 30' and dried under N_2 flux.

3.4.2.8 LC-UV analysis of the purified almond protein

The dried LTP-containing fraction was redissolved in 0.1% FA and analysed on a Waters Alliance 2695 HPLC system (Waters,Milford, MA, USA) equipped with a Jupiter C18 column (250 x 2 mm, 300 Å) (Phenomenex); the column oven temperature was set to 35°C. HPLC column flow was 0.2 ml/min, eluent A (0.1 TFA in H_2O) was kept at 100% for 12', then eluent B (0.1 TFA in ACN) was ramped to 50% in 65'.
The UV absorption profile was registered at 240 nm using a 2487 Dual λ Absorbance Detector (Waters).

3.4.2.9 Exact mass determination of the purified almond protein

The dried protein was resuspended in 0.1% FA in H_2O and analysed with a Dionex Ultimate 3000 micro HPLC coupled with an LTQ-Orbitrap mass

spectrometer equipped with a conventional ESI source. The chromatographic separation was performed using a Luna 5U C18 (50x4.6mm 100 Å) column (Phenomenex); HPLC column flow was 500 µL /min, eluent A (0.1% FA in H_2O) was kept at 85% for 5', then eluent B (0.1% FA in ACN) ramped to 90% in 30'.
The spectra were registered at a nominal resolution of 30000 (at m/z=400) and a mass range from 600 to 2000 m/z.

3.4.2.10 Tryptic and chymotryptic digestion of the purified almond protein

The dried protein was resuspended in 100 µl of NH_4HCO_3 50 mM in water; the disulphide bridges were reduced by adding 5 µl of an aqueous solution containing DTT 200 mM and NH_4HCO_3 0.1 M, and boiling the sample for 10'. The formed free thiols were alkylated through the addition of 4 µl of an aqueous solution containing IAA 1 M and NH_4HCO_3 0.1 M; the sample was kept in the dark for 1 h, then the alkylation was stopped by adding 20 µl of an aqueous solution containing DTT 200 mM and NH_4HCO_3 0.1 M.
After 1 h the sample was split in 2 equal parts to be digested respectively with trypsin and α-chymotrypsin. For this purpose 20 µl of an aqueous solution 0.1 M in NH_4HCO_3, containing 1 mg/ml enzyme, were added to each sample and the digestion was allowed to go on at 37°C overnight. Samples were desalted using a C18 sep pak cartridge (Waters), then the produced peptide mixture was analysed on a LTQ Orbitrap using the same parameters described above.

3.4.2.11 Determination of the number of disulfide bridges of the purified almond protein

The dried protein was resuspended in 0.1% FA in H_2O/ACN (1:1) containing TBP 5 mM and the obtained solution was left at 37°C for 30' in order to reduce the disulphide bridges. The mass spectrum of the reduced LTP was recorded with the same parameters described above.

3.4.2.12 In silico evaluation of potential allergenicity of the purified almond protein

Information on all the known sequences of lipid transfer protein from *Prunus dulcis* were retrieved using the UniProt database (www.uniprot.org). BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the CCD web tool[12] were employed to unveil the presence of conserved domains within our protein query. ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2) was employed to perform multiple alignment between retrieved protein sequences. For prediction of the potential allergenicity we used four web-based tools: Allermatch™ (www.allermatch.org), AlgPred (www.imtech.res.in/raghava/algpred/ index.html), AllergenFP v.1.0 (http://ddg-pharmfac.net/AllergenFP/ index.html) and EVALLER™ (www.slv.se/en-gb/Group1/Food-Safety/e-Testing-of-protein-allergenicity).

3.5 References

[1] Roux K.H., Teuber S.S., Sathe S.K. Tree nut allergens. Int. Arch. Allergy Immunol. 131(4), 234-244, Aug **2003**.

[2] Pastorello E.A., Farioli L., Pravettoni V., Ortolani C., Ispano M., Monza M. The major allergen of peach (Prunus Persica) is a lipid transfer protein. J. Allergy Clin. Immunol. 103, 520-526, **1999**.

[3] Fernandez-Rivas M., Gonzalez-Mancebo E., Rodriguez-Perez R., Benito C., Sanchez-Monge R., Salcedo G., Alonso D., Rosado A., Tejedor M.A., Vila C., Casas M.L. Clinically relevant peach allergy is related to peach lipid transfer protein, Pru p3, in the Spanish population. J. Allergy Clin. Immunol. 112, 789-795, **2003**.

[4] Garrido-Fernandez S., Garcia B.E., Sanz M.L., Ariz S., Tabar A.I. The role of lipid transfer proteins in the almond tree and almond fruit as contact and airborne allergens. J. Investig. Allergol. Clin. Immunol. 19(1), 61–63, **2009**.

[5] Chen L., Zhang S., Illa E., Song L., Wu S., Howad W., Arus P., Weg E., Chen K., Gao Z. Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. BMC Genomics. 9, 543-558, **2008**.

[6] Asero R., Mistrello G., Roncarolo D., de Vries S.C., Gautier M.F., Ciurana C.L., Verbeek E., Mohammadi T., Knul-Brettlova V., Akkerdaas J.H., Bulder I., Aalberse R.C., van Ree R. Lipid transfer protein: a pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. Int. Arch. Allergy Immunol. 122(1), 20-32, May **2000**.

[7] Garino C., Coïsson J.D., Arlorio M. In silico allergenicity prediction of several lipid transfer proteins. Comput. Biol. Chem. 28, 60, 32-42, Nov **2015**.

[8] Sathe S.K., Wolf W.J., Roux K.H., Teuber S.S., Venkatachalam M., Sze-Tao K.W. Biochemical characterization of amandin, the major storage protein in almond (Prunus dulcis L.) J. Agric. Food Chem. 50(15), 4333-4341, **2002**.

[9] Bencivenni M., Faccini A., Zecchi R., Boscaro F., Moneti G., Dossena A., Sforza S. Electrospray MS and MALDI imaging show that non-specific lipid-transfer proteins (LTPs) in tomato are present as several isoforms and are concentrated in seeds. J. Mass Spectrom. 12, 1264-1271, **2014**.

[10] Santos A.M., Oliver M.J., Sanchez A.M., Payton P.R., Oliveira M.M. Adventitious shoot regeneration in Prunus dulcis: suppression subtractive hybridization and gene expression profiling toisolate stage specific genes." Submitted to the EMBL/GenBank/DDBJ databases.6. May **2007**.

[11] Verde I., Abbott A.G., Salamini F., Schmutz J., Sosinski B., Morgante M., Rokhsar D.S. The genome of Prunis persica. The International Peach Genome Initiative Submitted to the EMBL/GenBank/DDBJ databases, Feb **2013**.

[12] Marchler-Bauer A., Lu S., Anderson J.B., Chitsaz F., Derbyshire M.K., DeWeese-Scott C., Fong J.H., Geer L.Y., Geer R.C., Gonzales N.R., Gwadz M., Hurwitz D.I., Jackson J.D., Ke Z., Lanczycki C.J., Lu F., Marchler G.H.,

Mullokandov M., Omelchenko M.V., Robertson C.L., Song J.S., Thanki N., Yamashita R.A., Zhang D., Zhang N., Zheng C., Bryant S.H. CDD: a Conserved Domain Database for the functional annotation of proteins., Nucleic Acids Res., 39:database issue D225-D229, **2011**.

[13] FAO and WHO. Evaluation of allergenicity of genetically modified foods: report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology, 22 – 25 January 2001. FAO, Rome. [http://www.who.int/foodsafety/publications/biotech/en/ec_jan2001.pdf], **2001**.

[14] FAO and WHO. Codex principles and guidelines on foods derived from biotechnology. FAO, Rome. [ftp://ftp.fao.org/codex/standard/en/CodexTextsBiotechFoods.pdf], **2003**.

[15] Fiers M.W.E.J., Kleter G.A., Nijland H., Peijnenburg A.A.C.M., Nap J.P., van Ham R.C.H.J. Allermatch (TM), a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC BIOINFORMATICS. 5, 133-139, **2004**.

4 Recombinant protein production: synthesis of the allergen Jug r 3

4.1 Introduction

Walnut is the edible seed of the trees of the genus *Juglans*, among which *Junglas regia*, the English walnut, is the most well known. Its regular consumption seems to improve body's lipid profile, contributing to the decrease of cholesterol levels, thus reducing the risk of coronary heart diseases. Other positive beneficial health effects that can be attributed to the intake of walnuts and other nuts are associated with their active intervention on inflammatory processes, oxidative stress, vascular reactivity and glycaemic control[1].

The recognized health benefits of walnut, combined with its pleasant taste, have led to its inclusion in several cookery dishes and pastry products, both as seed or as oil[2]. As a consequence of the increasing request for walnuts, their global production has grown of more than 300% in the last 20 years[3]. However, the ingestion of walnut can also represent a health risk due to the possibility of inducing hypersensitivity in allergic individuals: among tree nuts, walnut is in fact the most common cause of allergic reactions and, together with other tree nuts, it represents the food most frequently involved in anaphylaxis and anaphylaxis deaths, after peanut[4].

The protein Jug r 3, isolated for the first time in 2004, belongs to the nsLTP family, as determined after N-terminal sequencing, and has been classified as a food allergen in walnut seeds[4, 5].

In a study performed by Pastorello et al.[4], 36 out of 46 Italian patients with a convincing history of mild or life-threatening reactions to walnut, evidenced IgE reactivity to Jug r 3. Most of these patients were allergic also to other fruits containing LTPs, like Prunoideae (peach, apricot, cherry, plum), apple, grape, maize and hazelnut; thus, allergy to walnut can be considered a clinical manifestation of the LTP syndrome.

The sequence of Jug r 3 is constituted by 119 amino acids, with a signal peptide of 26 residues and it is encoded by the nucleotide sequence Juglans regia nonspecific lipid transfer protein mRNA, composed of 360 base pairs[6]. The

sequence of the gene was determined by Teuber and co-workers and submitted to the GenBank database with the accession number EU780670, in 2008[7] (**Figure 4.1**).

Figure 4.1 Nucleotide sequence (5'→3') and deduced amino acid sequence of Jug r 3. The signal peptide is underlined and * is the stop codon.

The recombinant production of the mature form of Jug r 3 (without signal peptide) was approached in the present work. The use of recombinant allergens from nuts is advantageous for diagnostic purposes since their natural counterparts might be underrepresented in natural extracts, due to loss during the defatting process; this is particularly true for allergens that, like LTPs, are only minor constituents of the protein fraction.

The experimental part of this work was performed at the Department of Experimental Immunology, Academic Medical Center, Amsterdam. The research group of Prof. Ronald van Ree is gratefully acknowledged.

4.2 Results and discussion

4.2.1 Production of recombinant proteins

The use of recombinant proteins as tools for cellular and molecular biology has become a commercial reality in several fields that comprise, among many others, production of pharmaceuticals, development of bio-insecticides,

diagnostic kits, enzymes with numerous applications and bioremediation processes[8].

Recombinant proteins are obtained from the expression within living cells of recombinant DNA; the latter is formed by the artificial joining of genetic material deriving from multiple sources, which is possible because DNA molecules from all organisms share the same chemical structure.

Using recombinant DNA technology together with synthetic DNA production, literally any DNA sequence may be created and introduced into any of a very wide range of living organisms, to obtain its expression.

A cloning vector is needed to insert the desired DNA sequence into the host and to allow its replication and expression; for this purpose plasmids or viruses can be used.

Once the vector containing the DNA fragment of interest has been inserted into the host organism, the foreign DNA might either be only replicated or also transcribed and translated, so that the recombinant protein is produced.

4.2.2 Choice of the host system

Several different host organisms, which can either be prokaryotic (bacteria such as *Escherichia coli*) or eukaryotic (yeasts, plant cells, cultured animal cells), are currently used for the production of recombinant proteins.

When choosing the host, the chemical properties of the recombinant protein to be produced have to be evaluated: if the desired protein is glycosylated or otherwise extensively post-translationally modified, an eukaryotic expression system should be used; bacteria in fact lack the endoplasmic reticulum and the Golgi apparatus, the sites where post-translational modifications occur.

As bacteria are, however, much more easily grown and manipulated in laboratory, the development of eukaryotic-like post-translational modification in bacterial hosts is a current area of research[9].

Actually, *E. coli* remains the dominant host for producing recombinant proteins, with 30% of the recombinant biopharmaceuticals licensed up to 2011 by the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMEA) being obtained using this host cell[10].

Many molecular tools and protocols, such as a vast catalogue of expression

plasmids, a great number of engineered strains and many cultivation strategies, have been developed for the high-level production of heterologous proteins in *E. coli*[11].

Since LTPs do not usually contain post-translational modifications, an *E. coli* strain was chosen in the present work for the production of the target protein.

4.2.3 Preparation of the DNA to be cloned

The DNA sequence codifying for the protein of interest might either be extracted from a biological tissue, be obtained from extracted mRNA using reverse transcriptase, or be synthesized artificially.

In the case of extraction from natural sources, the gene relative to the protein to be produced is amplified by **Polymerase Chain reaction** (PCR).

The PCR process allows to obtain thousands to millions of copies of the DNA sequence of interest.

Short DNA segments containing sequences complementary to the beginning and the end of the gene to be amplified, called forward and backward **primer** respectively, are used to initiate the selective copying of the region comprised between these two fragments.

A heat-stable DNA polymerase, such as the Taq polymerase, is employed to assemble new DNA strands.

The PCR consists of cycles (generally 28 to 35) of repeated heating and cooling of the reaction system for DNA melting and its enzymatic replication: first the template DNA is denatured by heating, so that single strands are obtained; the sample is then cooled to allow the primers to anneal, that is, to bind the appropriate complementary strand. The temperature for this step varies depending on the size of the primers, their GC content, and their homology to the target DNA. Finally, the primers are extended on both strands by DNA-polymerase, in the presence of Mg^{2+}; the temperature of this step is optimized for the used polymerase.

The segment of interest accumulates exponentially and strands with extension beyond the target sequence become negligible in the final mixture, since these accumulate in a linear manner.

The amplification of a gene of interest by PCR is illustrated in **Figure 4.2**.

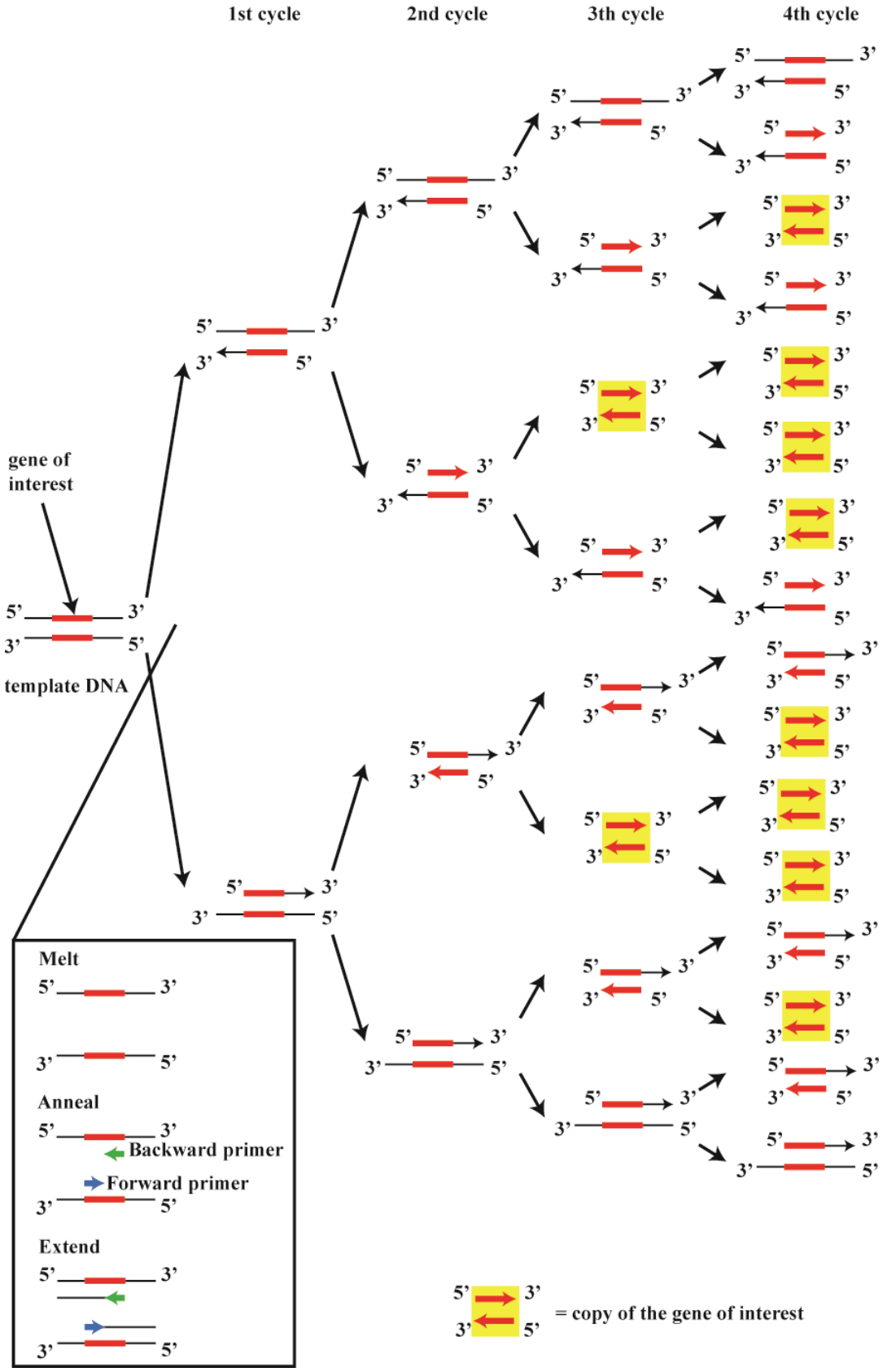


Figure 4.2 The first 4 cycles of the PCR process; in the n^{th} cycle, (2^n-2n) copies of the gene of interest will be produced.

PCR products can be visualized using agarose gel electrophoresis: the phosphate in the sugar-phosphate backbone confers a negative charge to the DNA fragments, which migrate towards the anode, through the pores present between agarose molecules; the latter act like sieves, separating the fragments by size. Staining is achieved using a nucleic acid-specific fluorescent compound, such as ethidium bromide (EtBr): this intercalates between the bases causing DNA to fluoresce orange when the dye is illuminated by UV light.
When the starting material is mRNA, reverse transcription PCR is used to obtain multiple copies of the gene of interest: first, the DNA complement of the mRNA is produced through the use of reverse transcriptase and subsequently, the newly synthesized cDNA is amplified using traditional PCR. The advantage of starting from mRNA is that this allows to selectively obtain the DNA coding regions of expressed genes only.
If the amino acidic sequence of the protein to be produced is known, but the nucleotide sequence codifying for it is not, to amplify the gene by means of PCR, degenerated primers have to be used. The latter are mixtures of primers including all the possible combinations of the codons codifying for the amino acids forming the fragment encoded by the primers.
As single amino acids can be codified by up to 6 different codons (**Table 4.1**) the number of possible primers, which is given by the product of the degeneracies of all the amino acids composing the fragment, rises quickly.

Table 4.1 Number of different codons coding for each amino acid.

Amino acid	**Number of possible codons (degeneracy)**
Met, Trp	1
Cys, Asp, Gln, Phe, Glu, His, Lys, Asn, Tyr	2
Ile	3
Ala, Gly, Pro, Thr, Val	4
Leu, Arg, Ser	6

The use of degenerate primers can greatly reduce the specificity of the PCR amplification, thus the artificial synthesis of the needed gene could be a good choice in this case. This also allows to use codons that are more commonly employed by the host organism, increasing the chances of success of the recombinant protein production.
In the present work, cDNA, produced starting from mRNA extracted from walnut, was used as starting material for the amplification of the gene coding for Jug r 3. Since the sequence of the latter is known, non-degenerated primers were employed.

4.2.4 Recombinant DNA production

Once synthesized or isolated from natural sources and amplified, the DNA sequence codifying for the protein of interest is bound to a **vector** to allow its insertion into the host organism. Vectors contain necessary genetic signals for replication, elements for convenience in inserting and expressing the foreign DNA and parts that allow to identify cells that contain recombinant DNA. Also genes for antibiotic resistance are usually present in vectors; this allows to select the cells that took up the vector by addition of antibiotics to the culture medium.
The choice of the vector depends on the host organism, the size of the DNA to be cloned and how the foreign DNA is to be expressed[12].
For the production of recombinant proteins, the most commonly used vectors are plasmids, which are extrachromosomal, double-stranded, generally circular, DNA sequences that are capable of automatically replicating in a host cell.
The DNA sequence codifying for the protein of interest can be combined with a plasmid through restriction enzyme digestion/ligation: first, the vector and the DNA fragment to be bound into it, are both digested using the same two restriction enzymes, which are able to cut double stranded DNA at specific 4 to 8 base pair, usually palindromic, recognition sequences. This process creates complementary ends in the insert and in the linearized vector and the two obtained fragments can be bound together employing a DNA ligase, which forms new phosphodiester bonds (**Figure 4.3**).

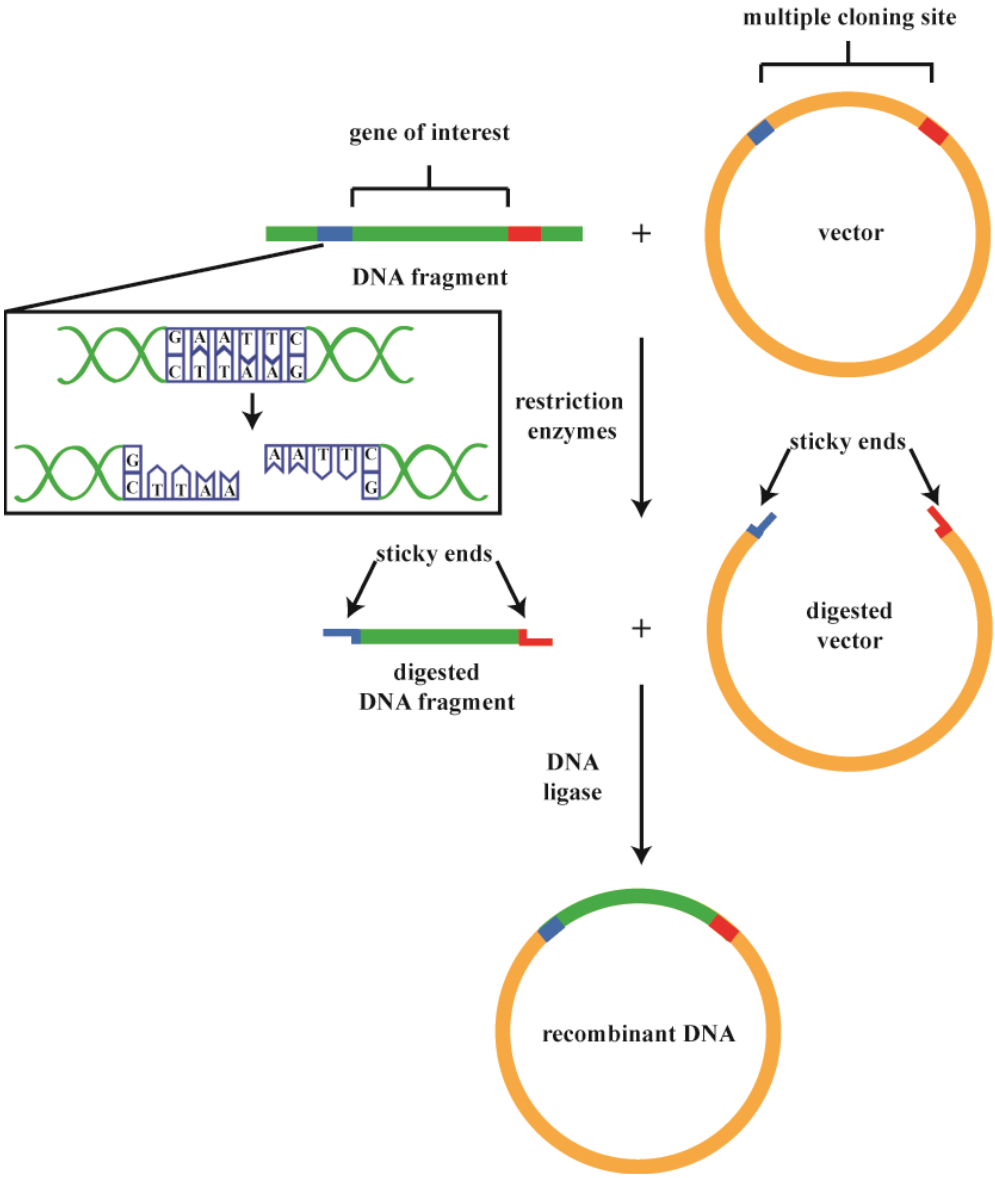


Figure 4.3 Insertion of the gene of interest into a vector through restriction enzyme digestion/ligation; the reported example of cleavage pattern is specific for the restriction enzyme *EcoR*I.

The multiple cloning site in the vector contains several different possible cleavage sites, to be cut selectively employing diverse restriction enzymes; this allows choosing the most convenient ones for the insertion of the fragment of interest. While amplifying the gene to be expressed by PCR, specific restriction sites can be included at the extremities of the product, inserting the corresponding sequences in the primers.

Commonly used vectors comprise a promoter site, or operon, which regulates

the expression of the recombinant DNA, so that this can be controlled and starts only in certain conditions, as a consequence of variations of physical properties of the system (pH or temperature) or after the addition of a chemical compound.
Often the sequence coding for the protein of interest is cloned in-frame with **tags** (peptides or proteins) such that they are expressed as contiguous to the protein, in a single unit. Tags can be used for various purposes: affinity tags ease the purification of the recombinant protein, as they can selectively be bound by specific ligands, usually immobilized on a resin, such that the protein containing the tag is isolated from all other (host) proteins that do not bind (e.g. a sequence of six histidine residues is able to coordinate divalent cations and proteins tagged by it can be purified by means of immobilized metal affinity chromatography (IMAC))[13]. Other tags might help to obtain the proper folding of the target protein and its solubilisation, to avoid its incorporation in inclusion bodies.
Also the protection of the target protein from degradation caused by proteases can be achieved through the use of specific fusion tags, that promote the translocation of the protein to different cellular locations.
Generally, there is an encoded protease cleavage site just before the sequence codifying for the protein of interest; this allows the enzymatic removal of the tags from the purified protein.

For the production of Jug r 3, the pE-SUMO3 vector was chosen; this plasmid contains a gene that confers resistance to Ampicillin and it leads to the expression of the target protein as fused with the SUMO tag, which was reported to improve expression, folding and solubility[14]. Moreover, a 6 His stretch is bound to the N-terminal of this tag, allowing the facile purification of the fusion protein by means of IMAC.
The sequence of the primers used to amplify the gene coding for Jug r 3 by means of PCR, is reported in **Figure 4.4**. In addition to parts corresponding to the beginning (forward primer) and the complement to the end (backward primer) of the nucleotide sequence for Jug r 3, these primers also contain recognition sites for the restriction enzymes BsaI and BamHI respectively, so that the digestion of the PCR product with these enzymes created sticky ends that allowed its binding into the pE-SUMO3 vector, previously digested with

the same enzymes.

Forward primer:

CC**GGTCTC**A AGGT*GTCATAACATGTGGGCAGGTGG*

BsaI

Backward primer:

CGC**G** **GATCC***TCATTTCACAGCTTTGCAG*

BamHI

Figure 4.4 Sequence (5'→3') of the primers used to clone the gene coding for Jug r 3 from walnut cDNA. Recognition sites for the used restriction enzymes are reported in bold; the cleavage sites are indicated by the arrows and the overhangs generated by the digestion are underlined. Parts corresponding to the beginning and the complement to the end respectively of the sequence codifying for Jug r 3, are reported in italics.

The success of the amplification of the gene coding for Jug r 3 was checked analysing the PCR product on agarose gel. As shown in **Figure 4.5**, a single band corresponding to a fragment of about 300 base pairs, compatible with the expected sequence of 304 bp, resulted from the analysis.

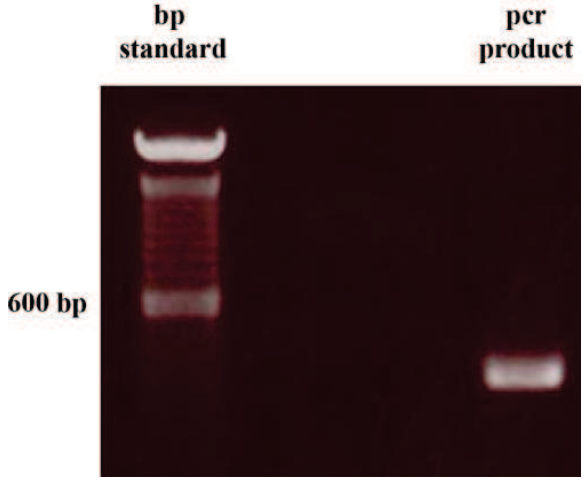


Figure 4.5 Agarose gel of the product obtained from the PCR amplification of the nucleotide sequence coding for Jug r 3.

As stated before, the amplified gene was digested with the restriction enzymes BsaI and BamHI, in a one-step process. The pE-SUMO3 vector was also subjected to digestion with the same enzymes but, due to the proximity of the two restriction sites in the vector, the process was carried out in two sequential steps.
The ligation of the amplified sequence into the linearized vector was performed using a T4 DNA ligase, according to the instructions of the manufacturer.

4.2.5 Insertion of the recombinant DNA into the host

Various methods can be used to get the previously produced recombinant DNA into host cells. Some microorganisms are able to take up DNA from their local environment; in this case, the insertion process is termed **transformation** and the host cells are said to be competent. Transformation usually requires preparation of the cells through a special growth regime and chemical treatment processes.
Already competent *E. coli* cells are commercially available and were purchased for the production of Jug r 3 described in the present work.
Other possible techniques are electroporation, which uses high voltage electrical pulses to translocate DNA across the cell membrane and, if present, the cell wall, and transduction, which involves the use of virus-like particles to introduce the DNA into the cell through a process resembling viral infection.

4.2.6 Recombinant DNA amplification and expression

To obtain amplification of the recombinant DNA, the same is firstly transformed into specific host cells, generally constituted by *E. coli* mutants, that are able to produce several clones of the plasmid DNA. These cells are plated on a selective agar medium containing the antibiotic against which the introduced vector confers resistance; cells that took up the plasmid are able to proliferate on this plate, generating colonies, which might either be constituted by cells containing the recombinant vector or the vector without insert. To distinguish non-recombinant from recombinant transformants, blue white screening might be performed[15, 16], or colony PCR might be carried out: single

colonies are transferred to liquid growth medium and the plasmids present in its cells are analysed through PCR, to check if they contain the insert.
Positive colonies are further grown in liquid medium and, finally, the cloned plasmids are recovered and purified chromatographically using silica gel or anion exchange resins.
Once the amplified recombinant DNA has been purified, it is introduced into the host cells used for expression.

In the present work, the recombinant DNA produced by the insertion of the amplified gene into the pE-SUMO3 vector was first transformed into Subcloning Efficiency™ DH5α™ competent *E. coli* cells to produce multiple copies of the plasmid of interest. Twelve different colonies generated from this transformation were grown and checked by means of colony PCR for content of the recombinant plasmid DNA.
The primers used for this analysis (T7 and T7T) are specific for external priming sites present in the vector. The PCR products were visualized on an Agarose gel (**Figure 4.6**): the colonies **10** and **12** contained the vector with insert (expected PCR product of 847 base pairs), while all the other colonies contained the empty vector (expected PCR product of 562 base pairs).

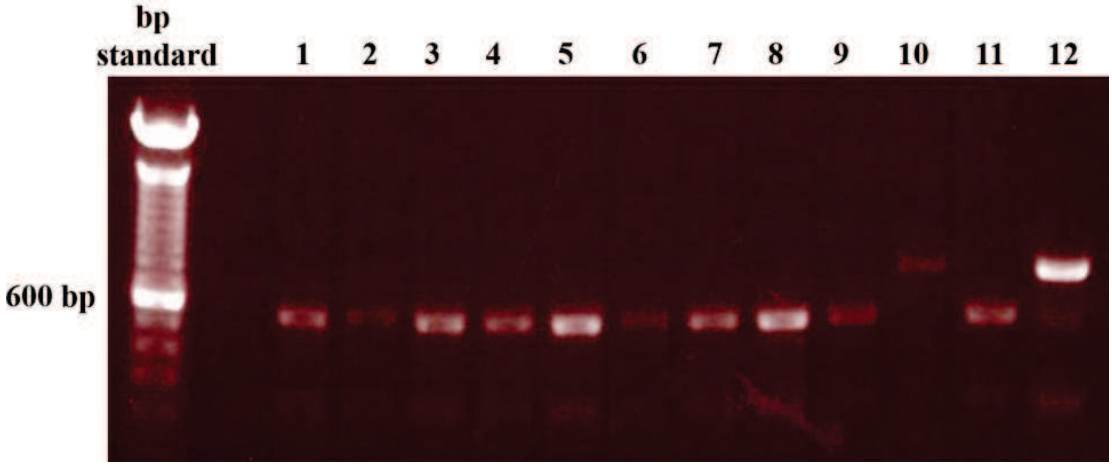


Figure 4.6 Agarose gel of the products obtained from colony PCR of the cells transformed with the ligation product. Each numbered column corresponds to a colony.

The 2 positive colonies where further grown, then the plasmid DNA contained in their cells was purified and transformed into Rosetta 2 (DE3) pLysS *E. coli* cells for expression.

As the best conditions for cell growth often differ from those for recombinant protein production, vectors containing promoters for inducible gene expression are usually employed in the recombinant DNA production, and induction is performed after a particular cell density has been reached.
Most commonly used vectors for recombinant protein production in *E. coli* contain the T7 promoter, that needs T7 RNA polymerase to be activated; the T7 RNA polymerase gene, required for the production of the latter, is usually under the control of a *Lac* promoter, which can be activated by the addition of Isopropyl β-D-1-thiogalactopyranoside (**IPTG**) to the culture medium. IPTG is a molecular mimic of allolactose, able to bind the *Lac* repressor *LacI*, which blocks transcription in the absence of inducers. The use of IPTG is advantageous because, unlike allolactose, it is not hydrolysable by β-galactosidase, and its concentration therefore remains constant in an experiment.
In absence of the inducer, the *Lac* repressor minimizes the basal transcription and this is particularly important when the expression target introduces cellular stress situations and thereby selects for plasmid loss[17].
The pET system, in which IPTG induced activation occurs, is illustrated in **Figure 4.7**.

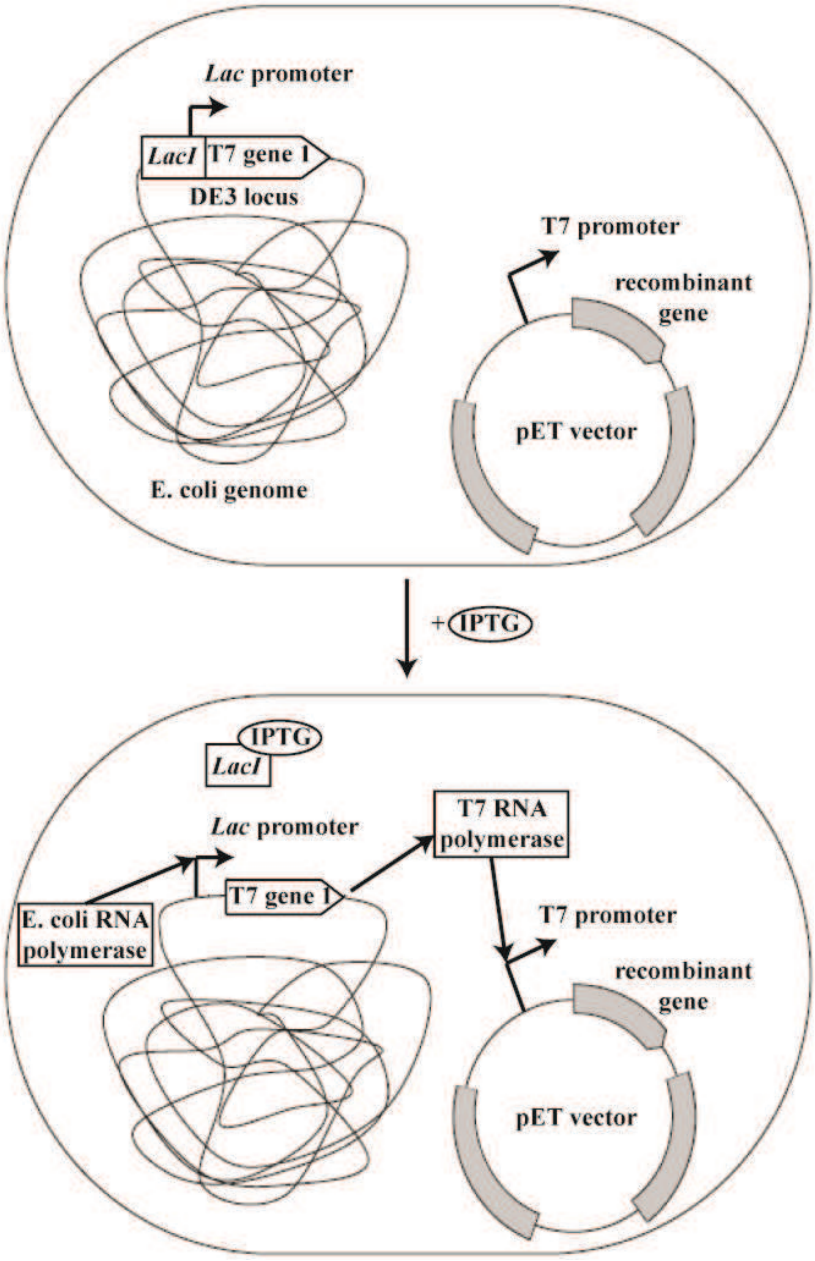


Figure 4.7 IPTG induced activation of the T7 promoter in the pET system.

To obtain a tighter control on the recombinant gene expression, a *Lac* operon can be introduced also in the vector, overlapping the T7 promoter, so that the *Lac* repressor controls the expression of both, the T7 RNA polymerase and the recombinant gene.

After expression, the host cells containing the produced recombinant protein are harvested from the culture medium by centrifugation or filtration. Methods used for subsequent extraction of the protein from the cells highly depend on the

used host and include, among others, cell lysis by osmotic shock, enzymatic digestion, freeze/thaw and ultrasonication.

For the production of Jug r 3 described in this work, the expression of the recombinant DNA was achieved through the addition of IPTG to the medium, as the cell-growth was in mid-log phase (OD600=0.6); aliquots of cells were harvested before induction and after 90' and 180'. To assess the expression of the recombinant fusion protein, the protein profile of the collected samples was visualized through SDS-PAGE (**Figure 4.8**).

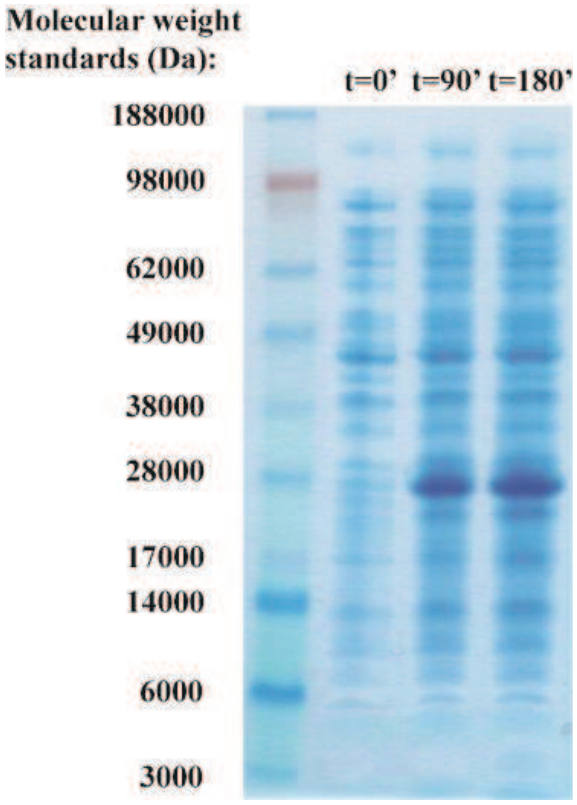


Figure 4.8 SDS-PAGE of the samples collected before (t=0') and after (t=90', 180') addition of IPTG to the culture medium.

In the SDS-PAGE analysis of the samples collected after induction, a band corresponding to the recombinant fusion protein can clearly be distinguished at a molecular weight of about 28 kDa.
The expression was repeated in a bigger experiment to obtain consistent quantities of target protein; following induction, bacteria were further grown at

30 ºC to prevent possible formation of inclusion bodies, due to too fast protein production. After 5 hours, the cells were harvested by centrifugation and suspended in a lysis buffer; the obtained slurry was sonicated to facilitate the dissolution of the proteins.

The target protein can be separated from the host proteins by classical purification techniques used in proteomics; chromatographic purification might be simplified by the use of affinity tags, as previously mentioned.
The recombinant walnut protein was separated from the host proteins through immobilized metal ion affinity chromatography, exploiting the presence of a six His stretch in the SUMO3 tag, which was able to interact with the stationary phase of the column, while all the proteins without tag were quickly eluted. The purified recombinant protein was recovered increasing the imidazole content of the mobile phase.
After the purification, a buffer exchange process was performed on the fractions containing the recombinant protein, so that the latter could be directly purified after the tag cleavage.

Any fusion tags are finally removed from the recombinant protein by chemical or, more commonly, enzymatic cleavage.
The SUMO-fusion tag, introduced in the strategy chosen for the production of Jug r 3, was cleaved employing the SUMO protease 2; unlike other proteases, for which the recognition sites are short linear sequences, thus carrying the risk of undesired cleavages within the protein of interest, the SUMO protease 2 recognizes the tertiary structure of the tag and cleaves the junction with the recombinant protein, irrespective of the N-terminal sequence of the latter (except proline). Therefore, the cleavage yields the native protein without extraneous residues at the N-terminus, if the target protein is fused directly to the C-terminus of the SUMO tag, as it was the case in the present work.
The progress of the digestion with the used protease was checked through SDS-PAGE and Immunoblotting with antibodies raised in a rabbit immunized with Mal d 3[18] (**Figure 4.9**).

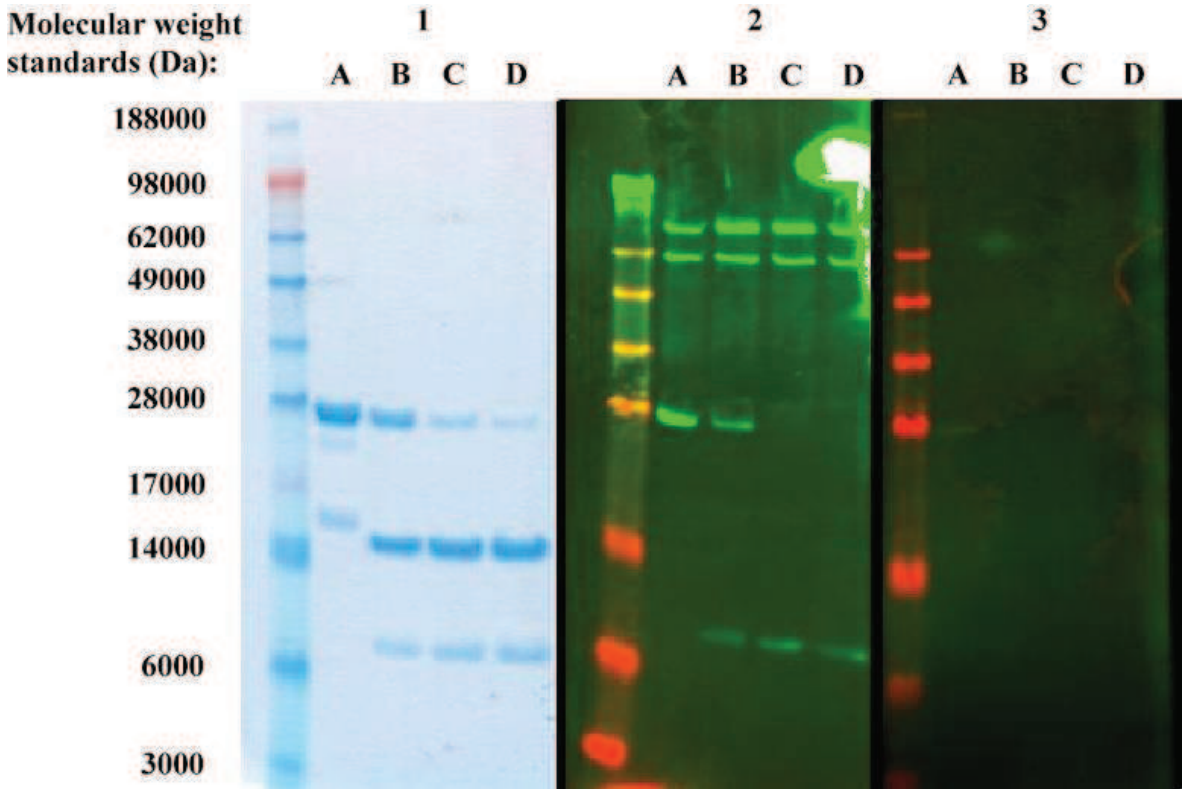


Figure 4.9 SUMO3 Cleavage: SDS-PAGE (**1**) and Immunoblot results produced by incubation with LTP specific polyclonal IgG rabbit antibodies: rabbit-anti LTP IgG 125487 (**2**); label control (**3**). **A**: undigested sample; **B**: sample after 1 h of digestion; **C**: sample after 2 h of digestion; **D**: sample after 4 h of digestion.

As shown by the disappearance of the band relative to the fusion protein (about 28 kDa) in the SDS-PAGE and Immunoblot results, the digestion was almost complete after 4 hours (**Figure 4.9 1D, 2D**). The band, in the SDS-PAGE analysis, at a molecular weight of about 14 kDa corresponds to the cleaved SUMO3 tag, while the band above the 6 kDa standard is relative to the recombinant LTP, which is also recognized by the rabbit-anti LTP IgG antibodies (**Figure 4.9 2**). No non-specific binding of the used secondary antibodies to the blotted proteins was detected (**Figure 4.9 3**).
The digested protein was purified through immobilized metal ion affinity chromatography: the cleaved SUMO3 tag and the SUMO protease 2 both contain His 6 tags and were therefore bound by the stationary phase, while the recombinant LTP was not retained.

4.2.7 Immunological tests

The purified protein was used to prepare an Immunoblot, to test its reactivity with antibodies raised in two rabbits immunized with Mal d 3[18] and IgE antibodies contained in sera of LTP allergic patients (**Table 4.2**) (**Figure 4.10**).

Table 4.2 Sera of LTP allergic patients used for the Immunoblot analyses.

Serum	Origin	Remarks
D	Italy	Peach allergy (IgE: 41.1 kUA/l); positive ImmunoCAP for Pru p 3
E	Italy	Peach allergy (IgE: 160.0 kUA/l); positive ImmunoCAP for Pru p 3 (IgE: 140.0 kUA/l)
F	Spain	Only peach allergy (IgE: 20.1 kUA/l); positive ImmunoCAP for Pru p 3
G	Spain	Peach allergy (IgE: 25.2 kUA/l); positive ImmunoCAP for Pru p 3 (IgE: 23.0 kUA/l); reported OAS with walnut
H	Netherlands	Peach allergy (IgE: 31.0 kUA/l); positive ImmunoCAP for Pru p 3 and LTPs from various other fruits

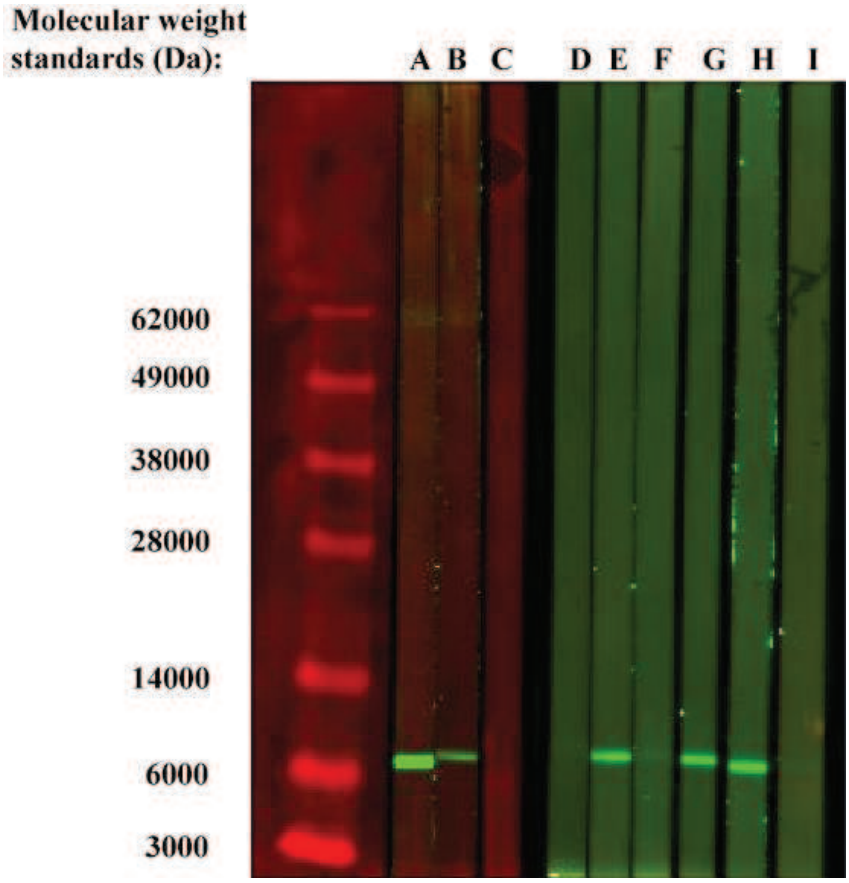


Figure 4.10 Immunoblot results produced by incubation with LTP specific polyclonal IgG rabbit antibodies; **A**: rabbit-anti LTP IgG 125487; **B**: rabbit-anti LTP IgG 126/41; **C**: label control. Immunoblot results produced by incubation with sera of LTP allergic patients; detection with anti-Human IgE. **D-H:** sera (**Table 4.2**); **I:** label control.

As shown in **Figure 4.10**, LTP specific polyclonal IgG antibodies raised in rabbits recognize the produced protein. Concerning the tests performed with human sera, positive results are found for the sera with high IgE titles against Pru p 3 and for patients with allergy to multiple fruits.

4.2.8 Challenges in recombinant protein production

Even when hundreds of proteins are produced successfully by recombinant techniques and these processes are apparently straightforward, many challenges might be encountered.

Expression of the recombinant gene can be lost due to structural changes in the latter or its disappearance from host cells: as the replication and the expression of the recombinant DNA imposes an additional metabolic load on the host, the growth rate of the cells results decreased and eventually faster-growing recombinant DNA free cells may overtake the culture. Generally, selective pressure is utilized to overcome this problem, using vectors containing genes that provide resistance to particular antibiotics; as previously mentioned, the vector employed in the production of the walnut LTP conferred resistance to ampicillin.

In some cases the produced recombinant protein results to be toxic for the host organism; as a consequence, slow grow rates or even host cell death can be observed. To try to overcome this problem the expression can be tuned, regulating the induction process, so that the concentration of the toxic target protein is maintained just below the host strain's tolerance. As no growth decrease was observed during the expression of our recombinant protein, the latter wasn't apparently toxic for the used host.

Stressful situations for the host cell, resulting in **protein misfolding** and consequent aggregation into inclusion bodies, might result from a strong and rapid protein production. To prevent or, at least, minimize these phenomena, various measures can be taken:

- Change of the cultivar conditions: the composition of the cell growth medium and the fermentation variables such as temperature, pH, induction time, and inducer concentration, might be varied to optimize the production of soluble recombinant proteins[19].
  Protein aggregation is usually reduced if the cultivation temperature is lowered, as this slows down the synthesis and the folding of the proteins, thus decreasing the hydrophobic interactions that are involved in self-aggregation. As stated before, in the production of Jug r 3 described in this work, the temperature was lowered during expression to prevent possible aggregation.
- Use of engineered host strains: mutant strains of commonly used host, like *E. coli*, have been produced through the introduction of DNA

mutations that affect protein synthesis, degradation, secretion, or folding.

- Co-production of molecular chaperones and folding modulators: chaperons and folding catalysts prevent protein aggregation and assist their proper folding (e.g. directing the formation of disulphide bonds); the co-production of these factors together with the target protein was reported to be a suitable strategy for the improvement of soluble protein production[20].
- Use fusion tags that enhance solubility, like the SUMO tag employed in our synthesis.

Finally, amino acid misincorporation and/or truncation of the recombinant protein might occur due to codon bias: when the frequency of occurrence of synonymous codons in the foreign coding DNA is significantly different from that of the host, depletion of low-abundance tRNAs occurs during the protein synthesis. Codon optimization of the foreign coding sequence or increasing the availability of underrepresented tRNAs by host modification are two common techniques used for solving codon usage bias[21]. The latter strategy was used to prevent those problems in the production of Jug r 3: the employed Rosetta 2 (DE3) pLysS *E. coli* cells, provide for "universal" translation by supplying tRNAs that are rarely used by *E. coli* cells.

4.2.9 Plasmid sequence and exact mass of the purified protein

To check the correctness of the sequence of the produced walnut LTP, the purified plasmid used for expression was sent to Baseclear (Leiden, the Netherlands) for sequencing and the obtained recombinant protein was analysed through mass spectrometry to determine its exact mass.
The sequencing of the purified plasmid evidenced that the gene coding for Jug r 3 was correctly inserted into the vector, but it also highlighted the presence of a wrong base in the seventeenth position of the sequence relative to the target protein (A replaced by G).

As the mutated base was covered by the forward primer, it is possible that the error was already present in the latter, or that the mutation was introduced as a consequence of an imperfect binding of the primer to the gene, during PCR.
Unfortunately, the error in the plasmid has an impact on the produced protein, since it leads to the introduction of an Arginine residue in the sixth position of Jug r 3, in the place of the Glutamine residue.
The presence of the mutation in the produced protein could easily be confirmed through its analysis by means of mass spectrometry (**Figure 4.11**): the exact mass was determined to be 9199.6 Da, which perfectly matches with the calculated mass of the mutated sequence of Jug r 3, in its oxidized form (i.e. with four disulphide bridges, as characteristic for LTPs).

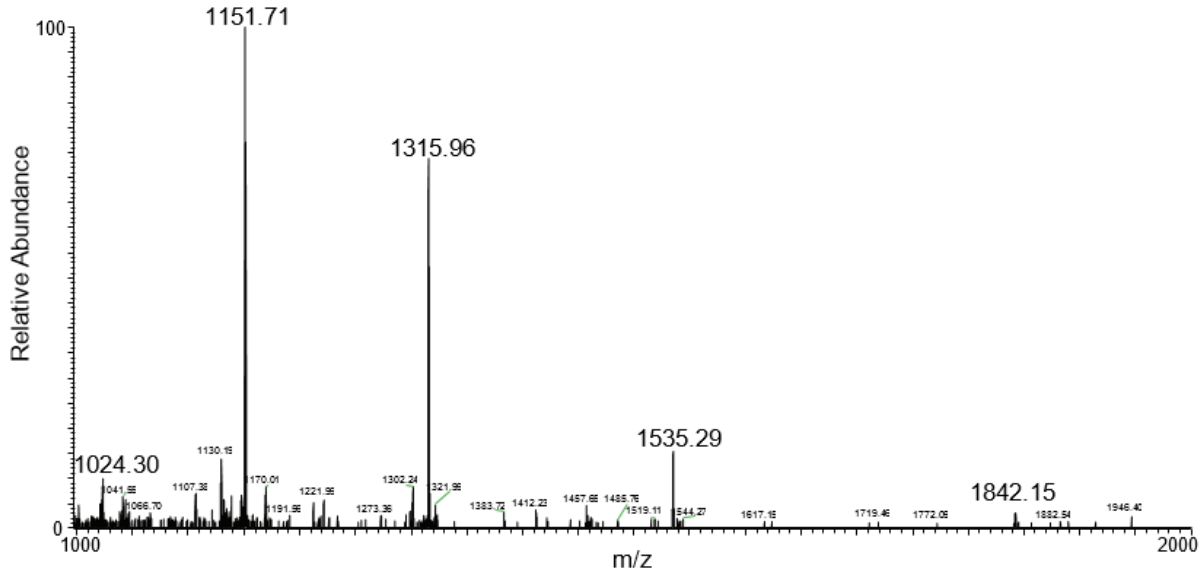


Figure 4.11 ESI-LTQ Orbitrap spetrum of the purified recombinant protein, having an exact mass of 9199.6 Da (found ions: 1842.15 m/z $[M+5H]^{5+}$, 1535.29 m/z $[M+6H]^{6+}$, 1315.96 m/z $[M+7H]^{7+}$, 1151.71 m/z $[M+8H]^{8+}$, 1024.30 m/z $[M+9H]^{9+}$).

The same kind of analysis proved to be a quick manner to confirm the success of the recombinant synthesis of a pistachio LTP, performed by Garino et al.
The sequence of the produced protein, comprehensive of the fusion tag used to ease its purification, is reported in **Figure 4.12**.

MGHHHHHHHHHHSSGHIEGRH**AITCGQVSTSLGPCINYVK TGGALPPACCAGIKALRAAAKTTAETQAACKCIKSAVSAI SGINYGIAAGLPGKCGVSIPYKISPSTDCSTVKY**

Figure 4.12 Sequence of the recombinant pistachio protein produced by Garino et al.; the N-terminal fusion tag is underlined.

The Orbitrap mass spectrum of the purified protein is depicted in **Figure 4.13**.

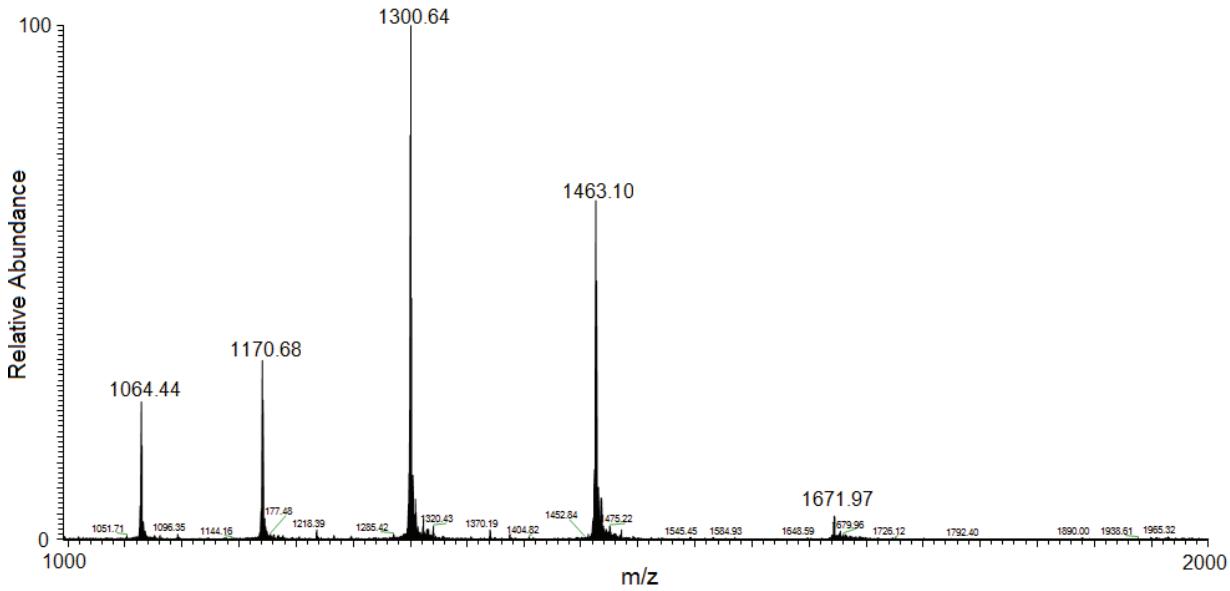


Figure 4.13 ESI-LTQ Orbitrap spetrum of the purified recombinant pistachio LTP, having an exact mass of 11689.7 Da (found ions: 1671.97 m/z $[M+7H]^{7+}$, 1463.10 m/z $[M+8H]^{8+}$, 1300.64 m/z $[M+9H]^{9+}$, 1170.68 m/z $[M+10H]^{10+}$, 1064.44 m/z $[M+11H]^{11+}$).

The exact mass determined through mass spectrometry, 11689.7 Da, matches with the theoretical mass of the expected sequence (**Figure 4.12**), considering the presence of 8 disulphide bridges in it.

4.3 Conclusions

The production of the walnut LTP, Jug r 3, through recombinant DNA techniques and its immunological characterization, are here described.
As LTPs do not contain post-translational modifications the process can be carried out using prokaryotic cells and *E. coli* was chosen as an easy to handle and economically convenient host.
The use of the pE-SUMO3 vector allowed the facile purification of the target, expressed as a fusion-tag protein, by IMAC chromatography; thereafter the tag was cleaved enzymatically.
Advanced MS techniques were shown to be a quick and powerful tool to confirm the success of recombinant protein production: the analysis of a recombinant pistachio LTP, produced by Garino et al., proved that the protein had the right sequence and formed four disulphide bridges, as characteristic for LTPs. Conversely, a mutation was detected in the sequence of the walnut protein.
The sequencing of the used recombinant plasmid evidenced that the mutation was already present in the latter and was therefore not introduced during expression.
Despite the presence of the error, the produced protein was shown to be immunologically active.

4.4 Experimental part

4.4.1 Chemicals

100 bp DNA Ladder (Invitrogen™), dNTP Mix 10 mM ea (Invitrogen™), Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Invitrogen™), Magnesium chloride ($MgCl_2$) 50 mM (Invitrogen™), NuPage® Antioxidant (Invitrogen™), NuPage® MES SDS Running Buffer (20X) (Invitrogen™), NuPage® Sample Reducing Agent (10X) (Invitrogen™), NuPage® LDS Sample Buffer (4X) (Invitrogen™), NuPage® Transfer buffer (Novex), PCR Rxn buffer (10X) (Invitrogen™), Pierce™ BCA Protein Assay Kit, SeeBlue® Plus2 Prestained Standard (Novex), Simplyblue™ SafeStain (Novex), Taq DNA polymerase

(Invitrogen™) and UltraPure™ Agarose (Invitrogen™) were purchased from Thermo Fisher Scientific (Waltham, MA, USA).

Acetic acid, Ampicillin, Dithiotreitol (DTT), Ethylenediaminetetraacetic acid (EDTA), Methanol (CH_3OH), Imidazole and Tris(hydroxymethyl) aminomethane hydrochloride (TRIS-HCl) were purchased from Sigma Aldrich (Sigma, St. Louis, MO, USA).

Blotting-Grade Blocker and Ethidium Bromide Solution 10 mg/ml were purchased from Bio-Rad (Hercules, CA, USA).

Di-sodium hydrogen phosphate dihydrate ($Na_2HPO_4 \cdot 2H_2O$), Glycerol, Sodium chloride (NaCl), Sodium dihydrogen phosphate monohydrate ($NaH_2PO_4 \cdot H_2O$) and Tween 20 were purchased from Merck-Millipore (Merck KGaA, Darmstadt, Germany).

Buffered sodium chloride solution, pH=7.4 (PBS) was purchased from Fresenius Kabi (Graz, Austria).

Difco™ Terrific Broth was purchased from Becton (Dickinson and Company, Sparks, MD, USA).

Gel Loading Dye, Orange (6X), NEBuffer 3, NEBuffer 4, Purified BSA 10 mg/ml, Restriction enzymes (BsaI, BamHI, 10 U/ml) and T4 DNA Ligase Reaction Buffer were purchased from New England Biolabs (Ipswich, MA, USA).

pE-SUMO3 Vector, T7, Amp and SUMO protease 2 were purchased from LifeSensors (Malvern, PA, USA).

Secondary IgG antibodies labelled with IRDye® 800 CW (IRDye® 800CW Goat anti-Rabbit IgG (H + L), 0.5 mg; IRDye® 800CW Donkey anti-Mouse IgG (H + L), 0.5 mg) were purchased from LI-COR Biotechnology (Lincoln, NE, USA), while IRDye® 800 CW Conjugated monoclonal Mouse anti-Human IgE were obtained from Rockland (Gilbertsville, PA, USA).

T4 DNA Ligase was purchased from Roche (Roche Diagnostics, Indianapolis, IN, USA).

T7 and T7T primers were obtained from BaseClear (BaseClear Lab Services, Leiden, The Netherlands).

4.4.2 Procedure

4.4.2.1 Cloning

Walnut RNA was extracted and cDNA was prepared from it, in a previous work and stored at -80 ºC; this cDNA was used for our following experiments.
A primer pair designed on the basis of the published sequence of the gene codifying for Jug r 3, was purchased from Invitrogen™ (Thermo Fisher Scientific) and used to amplify the latter gene by means of PCR: 2.5 µl PCR Rxn buffer, 2.5 µl dNTP Mix, 0.75 µl $MgCl_2$ 50 mM, 0.5 µl forward primer 10 µM, 0.5 µl backward primer 10 µM, 1 µl walnut cDNA, 0.25 µl Taq DNA polymerase and 17 µl H_2O were mixed in a PCR tube that was placed in a C1000™ Thermal Cycler (Bio-Rad).
The PCR conditions were: 3' at 95 ºC; 35 cycles of: 30" at 95 ºC, 30" at 60 ºC, 1' at 72 ºC; 5' at 72 ºC.
Agarose gel: an 1% agarose gel was prepared according to the following procedure: 0.5 g of agarose were dissolved in 50 ml TAE buffer (40mM Tris, 20mM acetic acid, and 1mM EDTA) in a microwave; the obtained solution was cooled, 2 µl of Ethidium Bromide Solution were added to it and finally it was poured into a suitable mould. 30' later, the produced gel was used to analyse the PCR product (9 µl sample + 1 µl Gel Loading Dye); 100 bp DNA Ladder was used as marker. After a 10' run at 140 V, the gel was visualized using a GeneFlash Gel Documentation System (Syngene-Synoptics, Cambridge, United Kingdom).
Digestion: the PCR product was purified using a QIAquick® Gel Extraction Kit (Qiagen, Venlo, the Netherlands), according to the instructions of the manufacturer, then it was digested with the restriction enzymes BsaI and BamHI: 76 µl purified PCR product, 10 µl Purified BSA, 10 µl NEBuffer 3, 2 µl BsaI and 2 µl BamHI were mixed in a tube and incubated at 37 ºC. After 1 h a new purification with the QIAquick® Gel Extraction Kit was performed and the concentration of the DNA in the eluate was determined using a Nanodrop system (Model 2000c - Thermo Fisher Scientific).

The pE-SUMO3 vector was digested with the BsaI enzyme according to the following procedure: 17.6 µl pE-SUMO3 vector, 5 µl Purified BSA, 5 µl NEBuffer 4, 21.4 µl H_2O and 1 µl BsaI were mixed and incubated at 37 ºC for 1 h. The product was then purified using the QIAquick® Gel Extraction kit (Qiagen) and the whole eluate (40 µl) was mixed with 5 µl Purified BSA, 5 µl NEBuffer 3 and 1 µl BamHI. After 1 h of incubation at 37 º C, a new purification with the QIAquick® Gel Extraction Kit was performed and the concentration of the DNA in the eluate was determined using a Nanodrop system.

Ligation: 10 fmol of the insert and 30 fmol of the digested vector were mixed together with 10 µl T4 DNA Ligase Reaction Buffer and 1 µl T4 DNA Ligase; after 5'at room temperature the tube was placed on ice for 1 h.

Transformation for amplification: 5 µl (corresponding to about 5 ng of DNA) of the ligation product were added to a tube containing 50 µl of Subcloning Efficiency™ DH5α™ competent *E. coli* cells (Invitrogen™ -Thermo Fisher Scientific); after gentle mixing and 30' of incubation on ice, the tube was placed for 20" in a 42 ºC water bath and again on ice for 2'. 950 µl of TB medium (46.7 g Difco™ Terrific Broth + 4 ml glycerol in 1 l of aqueous solution) were added to the cells and the suspension was incubated for 1 h at 37 ºC with 225 rpm shaking. 100 µl of the transformation product were spread on a pre-warmed LB plate containing 0.05 mg/ml ampicillin and the plate was incubated overnight at 37 ºC.

12 clones were picked from the plate and transferred into aliquots of 200 µl of TB medium containing 0.05 mg/ml ampicillin, and the obtained suspensions were incubated for 4 h at 37 ºC, with 200 rpm shaking.

Colony PCR: 10 µl of each suspension were added to 90 µl H_2O and 5 µl of the obtained mixtures were placed in PCR tubes together with 10 µl of a solution made of: 1.5 µl PCR Rxn buffer, 1.5 µl dNTP Mix, 0.6 µl $MgCl_2$ 50 mM, 0.75 µl T7 primer 10 µM, 0.75 µl T7T primer 10 µM, 0.13 µl Taq DNA polymerase and 4.8 µl H_2O.

The PCR conditions were: 3' at 95 ºC; 35 cycles of: 30" at 95 ºC, 30" at 50 ºC, 1' at 72 ºC; 5' at 72 ºC.

The obtained products were visualized on an agarose gel following the procedure outlined above.

5 µl of the suspensions containing the positive clones were inoculated in 5 ml TB medium containing 0.05 mg/ml ampicillin, and incubated overnight at 37 ºC with 200 rpm shaking.
The tubes containing the cultures were centrifuged at 4600 rpm for 15' at 4 ºC; the obtained pellets were treated with the QIAprep® Spin Miniprep Kit, according to the instructions of the manufacturer, to purify the included plasmids.
The DNA concentration in the produced eluates was determined using a Nanodrop system and a part of them was sent for sequencing to Baseclear (Leiden, the Netherlands).

4.4.2.2 Expression

The purified plasmid DNA was transformed into Rosetta 2 (DE3) pLysS *E. coli* cells (Novagen-Merck), according to the following procedure: 0.1 µl of the eluate obtained from the previous step, were added to 20 µl of cells and the tube was incubated on ice for 5'; after that it was placed for 30" in a 42 ºC water bath and again on ice for 2'. 80 µl of TB medium were added to the cells and the suspension was incubated for 1 h at 37 ºC with 250 rpm shaking. The transformation product was spread on a pre-warmed LB plate containing 0.1 mg/ml ampicillin and the plate was incubated overnight at 37 ºC.
Colonies were picked from the plate and transferred into 5 ml TB medium containing 0.05 mg/ml ampicillin, and the obtained suspensions were incubated at 37 ºC with 200 rpm shaking.
Expression of the recombinant DNA was induced by the addition of an aqueous solution of IPTG, to a final concentration of 1 mM, as the OD600 value, measured through a Novaspec II spectrophotometer (LKB Biochrom LTD, Cambridge, England), reached 0.6.
To assess protein expression, samples (250 µl) were collected at different times (t=0', 90' and 180') and centrifuged for 20' at 13200 rpm; the obtained pellets were sheared, using a syringe, in 100 µl of an aqueous solution containing 25 µl of NuPage® LDS Sample Buffer and 10 µl of NuPage® Sample Reducing Agent, and analysed through SDS-PAGE.

SDS-PAGE: the samples were heated at 70 ºC for 10' and run at 200 V for 40' on a NuPage® 4-12% Bis-Tris gel (Novex-Thermo Fisher Scientific), employing a XCell SureLock™ Mini-Cell electrophoresis system (Novex-Thermo Fisher Scientific); SeeBlue® Plus2 Prestained Standard was used as MW marker. The gel was stained for 1 h employing Simplyblue™ SafeStain and destaining to the desired contrast was achieved washing with H_2O.

Expression scale-up: clones were inoculated in 100 ml TB medium containing 0.05 mg/ml ampicillin, and the obtained suspensions were incubated at 37 ºC with 200 rpm shaking until the OD600 value reached 1. Cells were recovered centrifuging at 4600 rpm for 20', and transferred in 1 l TB medium containing 0.05 mg/ml ampicillin. The system was incubated at 37 ºC with 200 rpm shaking until the OD600 value reached 0.6, then an aqueous solution of IPTG was added to a final concentration of 1 mM. After 5 h of incubation at 30 ºC with 250 rpm shaking, the suspension was centrifuged (4600 rpm, 20') and the supernatant was discarded.

The pellet was dissolved in 50 ml of an aqueous solution containing TRIS-HCl 50 mM and EDTA 2 mM, pH=8. After 30' on ice, the slurry was sonified for 90" (70% amplitude, 30'' on, 30" off) using a Branson digital sonifier (Model 250 - Branson-Emerson, Ferguson, MO, USA) and then centrifuged for 30' at 9500 rcf at 10 ºC. The obtained supernatant, containing the recombinant protein, was used for the subsequent steps.

4.4.2.3 Purification

The recombinant fusion protein was purified using an FPLC system (AKTA purifier, Amersham Biosciences, Uppsala, Sweden) and the following buffers:

- A: sodium phosphate 20 mM, NaCl 0.5 M, imidazole 20 mM; pH=7.4
- B: sodium phosphate 20 mM, NaCl 0.5 M, imidazole 0.5 M; pH=7.4

A HiTrap Desalting column (GE Healthcare, Little Chalfont, United Kingdom) was used to achieve buffer exchange of the proteins from the sample to eluent A. A HisTrap HP column (GE Healthcare) was then employed to separate the recombinant protein from the host proteins; flow rate was 2.5 ml, eluent A was kept at 100% for 10 column volumes and then eluent B was ramped to 100% in 10 column volumes. Fractions of 4 ml were collected during the isocratic

elution, while fractions of 2 ml were collected during the gradient phase. The absorbance of the eluate was monitored at 280 nm and fractions corresponding to peaks in the chromatogram were analysed through SDS-PAGE, as previously described (samples were prepared mixing 13 µl of eluate with 5 µl of NuPage® LDS Sample Buffer and 2 µl of NuPage® Sample Reducing Agent).
Fractions containing the recombinant fusion protein were pooled and concentrated by ultrafiltration, using an Amicon stirred cell (Model 8050 - Merck) equipped with an Ultracel® 10 kDa ultrafiltration disk (Millipore, Bedford, MA, USA).
The HiTrap Desalting column was used again to achieve buffer exchange of the proteins from the sample to eluent A, then the concentration of the eluate was determined using the Pierce™ BCA Protein Assay Kit following the instruction of the producer.

4.4.2.4 SUMO3 cleavage

DTT to a final concentration of 1 mM and SUMO protease 2 (1 Unit per 10 µg of protein) were added to the sample and the digestion was allowed to go on at 37 ºC with slight shaking; the progress of the reaction was checked after 1 h, 2 h and 4 h by SDS-PAGE, using a NuPage® 12% Bis-Tris gel (Invitrogen™ - Thermo Fisher Scientific), according to the previously described procedure (samples were prepared mixing 13 µl of reaction mixture with 5 µl of NuPage® LDS Sample Buffer and 2 µl of NuPage® Sample Reducing Agent). A sample prepared with the undigested protein was run on the same gel for comparison.
Immunoblotting: a part of the gel was not stained, but used to obtain an immunoblot, employing a XCell II™ Blot Module (Novex-Thermo Fisher Scientific); transfer buffer was prepared mixing 50 ml NuPage® Transfer buffer, 100 ml CH_3OH and 350 ml H_2O. Proteins were transferred on a Nitrocellulose membrane (0.2 µm pores) (Novex-Thermo Fisher Scientific), which was then incubated for 1 h in a blocking buffer made of PBS containing 5% (w/v) Blotting-Grade Blocker and finally washed for 3 times with PBS containing 0.1% (v/v) Tween 20.

The blot was cut in 2 equal parts, one of which was incubated for 7 h at 4 ºC in 5 ml of PBS containing 0.1% (v/v) Tween 20, 0.5% (w/v) Blotting-Grade Blocker and 2 µl of solution containing IgG antibodies raised in a rabbit immunized with Mal d 3[18].
The blot was washed for 3 times with PBS containing 0.1% (v/v) Tween 20 and then it was incubated overnight at 4 ºC in 5 ml of PBS containing 0.1% (v/v) Tween 20, 0.5% (w/v) Blotting-Grade Blocker and 2 µl of solution containing goat-anti-rabbit IgG secondary antibodies labelled with IRDye® 800 CW.
The part of the blot that had not been treated with the primary rabbit antibodies was incubated with the secondary antibodies, as described above, to detect possible nonspecific binding of the latter with the proteins.
The blots were washed for 3 times with PBS containing 0.1% (v/v) Tween 20 and then they were visualized using an infrared fluorescence detection Odyssey Imager and software (LI-COR Biotechnology).
Purification of the digested protein: the recombinant LTP was separated from the SUMO3 tag and the SUMO protease 2 through FPLC purification using a HisTrap HP column as described above.
The collected fractions were analysed by means of SDS-PAGE and those containing the recombinant protein were pooled and concentrated by ultrafiltration, using Amicon filter devices with nominal MW cut-off of 3 kDa (Millipore, Bedford, MA, USA). The concentration of the recovered retentate was determined using the Pierce™ BCA Protein Assay Kit following the instructions of the producer.

4.4.2.5 Immunological tests

A volume of retentate corresponding to 25 µg of protein was mixed with 100 µl of NuPage® LDS Sample Buffer, 40 µl of NuPage® Sample Reducing Agent and PBS to a final volume of 400 µl; the so produced sample was used for SDS-PAGE on a NuPage® 4-12% Bis-Tris 2D gel (Novex™ -Thermo Fisher Scientific) and after the electrophoretic run the proteins were transferred on a Nitrocellulose membrane and the latter was suddenly blocked, as previously described.
Strips having a width of about 3 mm were cut from the blot and the binding

ability to the blotted protein of polyclonal IgG antibodies raised in two rabbits immunized with Mal d 3[18] and IgE antibodies contained in sera of patients allergic to LTPs, was tested.
Each strip was incubated overnight at 4 ºC in 3 ml of PBS containing 0.1% (v/v) Tween 20, 0.5% (w/v) Blotting-Grade Blocker and 1 µl of solution containing IgG antibodies or 150 µl of human serum.
The strips were washed for 3 times with PBS containing 0.1% (v/v) Tween 20 and then they were incubated for 5 h at 4 ºC in 3 ml of PBS containing 0.1% (v/v) Tween 20, 0.5% (w/v) Blotting-Grade Blocker and 1 µl of solution containing respectively goat-anti-rabbit IgG or mouse-anti-human IgE antibodies labelled with IRDye® 800 CW.
Two strips that had not been treated with the primary antibodies were incubated with the two secondary antibodies respectively, as described above, to detect possible nonspecific binding of the latter to the purified protein.
The strips were washed for 3 times with PBS containing 0.1% (v/v) Tween 20 and then they were visualized using an infrared fluorescence detection Odyssey Imager and software (LI-COR Biotechnology).

4.4.2.6 Exact mass determination

The purified protein was analysed with a Dionex Ultimate 3000 micro HPLC (Dionex, Sunnyvale, CA, USA) coupled with an LTQ-Orbitrap mass spectrometer (Thermofisher, San José, CA, USA) equipped with a conventional ESI source.
The chromatographic separation was performed using a Luna 5U C18 (50x4.6mm 100 Å) column (Phenomenex, Torrance, CA, USA); HPLC column flow was 500 µL /min, eluent A (0.1% FA in H_2O) was kept at 85% for 5', then eluent B (0.1% FA in ACN) was ramped to 90% in 30'.
The ESI source parameters were configured as follows: spray voltage 3.5 kV, capillary voltage 30 V and tube lens 200 V.
The spectra were registered at a nominal resolution of 30000 (at m/z=400) and a mass range from 600 to 2000 m/z.
The recombinant pistachio LTP, produced by Garino et al., was analysed using the same experimental settings.

4.5 References

[1] Costa J., Carrapatoso I., Oliveira M.B., Mafra I. Walnut allergens: molecular characterization, detection and clinical relevance. Clin. Exp. Allergy. 44(3), 319-341, Mar **2014**.

[2] Alasalvar C., Shahidi F. Tree nuts: composition, phytochemicals, and health effects: an overview. In: Alasalvar C, Shahidi F, eds. Tree nuts: composition, phytochemicals, and health effects. Boca Raton: CRC Press, 1–6, **2008**.

[3] FAOSTAT, The Statistics Division of the Food and Agriculture Organization of the United Nations. Available at: http://faostat3.fao.org/

[4] Pastorello E.A., Farioli L., Pravettoni V., Robino A.M., Scibilia J., Fortunato D., Conti A., Borgonovo L., Bengtsson A., Ortolani C. Lipid transfer protein and vicilin are important walnut allergens in patients not allergic to pollen. J. Allergy Clin. Immunol. 114(4), 908-914, Oct **2004**.

[5] ALLERGEN, official site for the systematic allergen nomenclature. Available at: http://www.allergen.org/.

[6] NCBI, National Center for Biotechnology Information, Bethesda, USA. Available at: http://www.ncbi.nlm.nih.gov/.

[7] Teuber S.S., Dandekar A.M., Uratsu S., Peterson W.R. Juglans regia nonspecific lipid transfer protein mRNA, complete cds. Available at: http://www.ncbi.nlm.nih.gov/. May **2008**.

[8] Palomares L.A., Estrada-Mondaca S., Ramírez O.T. Production of Recombinant Proteins. Methods in Molecular Biology, vol. 267: Recombinant Gene Expression: Reviews and Protocols, Second Edition Edited by: P. Balbás and A. Lorence © Humana Press Inc., Totowa, NJ. **2004**.

[9] Nothaft H., Szymanski C.M. Bacterial protein N-glycosylation: new perspectives and applications. J. Biol. Chem. 288, 6912–6920, **2013**.

[10] Berlec A., Strukelj B. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. J. Ind. Microbiol. Biotechnol. 40(3-4), 257-274, Apr **2013**.

[11] Rosano G.L., Ceccarell E.A. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 5, 172, **2014**.

[12] Russell D.W., Sambrook J. Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory. **2001**.

[13] Jayaraj R., Smooker P.M. So you Need a Protein - A Guide to the Production of Recombinant Proteins. Open Vet. J. 3, 28-34, **2009**.

[14] Butt T.R., Edavettal S.C., Hall J.P., Mattern M.R. SUMO fusion technology for difficult-to-express proteins. Protein Expr. Purif. 43(1), 1-9, Sep **2005**.

[15] Langley K.E., Villarejo M.R., Fowler A.V., Zamenhof P.J., Zabin I. Molecular basis of beta-galacosidase α-complentation. Proc. Natl. Acad. Sci. USA. 72(4), 1254-1257, Apr **1975**.

[16] Zamenhof P., Villarejo M. Construction and properties of *E. coli* strains exhibiting α- complementation of beta-galactosidase fragments in-vivo. J. Bacteriol. 110(1), 171-178, Apr **1972**.

[17] Sørensen H.P., Mortensen K.K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J. Biotechnol. 115(2), 113-128, Jan **2005**.

[18] Zuidmeer L., van Leeuwen W.A., Budde I.K., Cornelissen J., Bulder I., Rafalska I., Besolí N.T., Akkerdaas J.H., Asero R., Fernandez Rivas M.,

Gonzalez Mancebo E., van Ree R. Lipid transfer proteins from fruit: cloning, expression and quantification. Int. Arch. Allergy Immunol. 137(4), 273-281, Aug **2005**.

[19] Jana S., Deb J.K. Strategies for efficient production of heterologous proteins in *Escherichia coli*. Appl. Microbiol. Biotechnol. 67(3), 289-298, May **2005**.

[20]. Kolaj O., Spada S., Robin S., Wall J.G. Use of folding modulators to improve heterologous protein production in *Escherichia coli*. Microb. Cell Fact. 27, 8, 9, Jan **2009**.

[21] Sorensen H. P., Mortensen K.K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J. Biotechnol. 115, 113–128, **2005**.

5 Total chemical synthesis of proteins: design of a synthetic strategy to obtain the allergen Pru p 3

5.1 Chemical synthesis of proteins

Chemical synthesis of proteins enables a level of control over protein composition beyond that attainable by ribosome-dependent protein expression. It also facilitates the study of the structure-property relationship of protein functions at atomic resolution and moreover, it holds promise for creating proteins with usage in biomedicine that would otherwise be difficult-to-obtain[1, 2].

Chemical synthesis also allows to easily incorporate unnatural amino acids, post-translational modifications or labelling agents, opening new opportunities for the understanding of protein molecules and controlling their biological mechanism of action.

Moreover, the chemical production of peptides and proteins can avoid contamination and other issues that can arise from their production from natural or recombinant sources[3].

In the last decades several techniques allowing the total chemical synthesis of proteins have been developed. Proteins and enzymes of more than 200 amino acids have been successfully synthesized chemically and their biological properties were proven to be identical to those of their natural occurring homologues[1].

However, these techniques have not been applied yet to the production of allergenic proteins.

In the present work strategies for the total chemical synthesis of Pru p 3, the prototypic LTP, were developed.

5.2 Solid Phase Peptide Synthesis (SPPS)

Peptide synthesis was traditionally performed in solution; this required long production times and purifications of the reaction product were necessary after each amino acidic coupling.

Solid Phase Peptide Synthesis (SPPS) was introduced for the first time by Bruce Merrifield in 1963: this technique involves the use of an inert, insoluble solid support on which the growing peptide chain is attached[4]. Since the forming peptide is bound to a solid, large excesses of the entering amino acids and of the coupling reagents can be used, driving the reaction to completion. Once the bond with the new amino acid has been formed, all the species in excess and the by-products can be eliminated simply by means of washing and filtration.

The most used solid supports for SPPS are polymer resins like polystyrene, polyacrylamide and polyethylene glycol. Crosslinking with about 1% divinylbenzene confers mechanical stability, while maintaining good swelling properties.

Various factors influence the choice of the resin to be used:

- Swelling properties in different reaction solvents: SPPS relies on proper swelling in polar solvents as polar aprotic solvents facilitate coupling and washing steps; good swelling means good accessibility of coupling sites and thus, a smooth reaction.
- Beads size: diffusion of the reagents might be limited when using resins with large beads size. A narrow particle size distribution is crucial for achieving uniform reaction conditions.
- Extent of loading of the linker or of the first amino acid: higher loadings allow obtaining higher peptide yields, but for the production of long or difficult peptide sequences, because of steric hindrance, the use of resins with low loadings leads to better results.
- Nature of the linker used to bind the first amino acid to the support: particular linkers can be used to obtain C-terminal functionalised peptides.

As the loading of the first amino acid on the resin is often difficult and the used reaction conditions can lead to racemization, most of the resins are purchased already preloaded.

The synthesis starts from the C-terminal of the peptide to be produced and the amino acid coupling process is repeated until the sequence is completed, whereupon the desired peptide is cleaved from the resin.

The first amino acid is bound to the linker on the resin through its carboxylic group; the entering amino acid forms an amide bond with the aminic group of the first amino acid, using its carboxylic group. To avoid auto-condensation of the entering amino acid in solution, its aminic group is protected and the protecting group is removed only once the coupling reaction is completed and the excess of the reagents has been washed away. The reaction with the next amino acid can then take place and so on.

Amino acids bearing groups in their side chains which would be reactive in the coupling conditions have additional protections; these are stable during the whole formation of the peptide (amino acidic coupling and α–amino deprotection) and are usually removed simultaneously to the cleavage of the peptide from the resin (**Figure 5.1**).

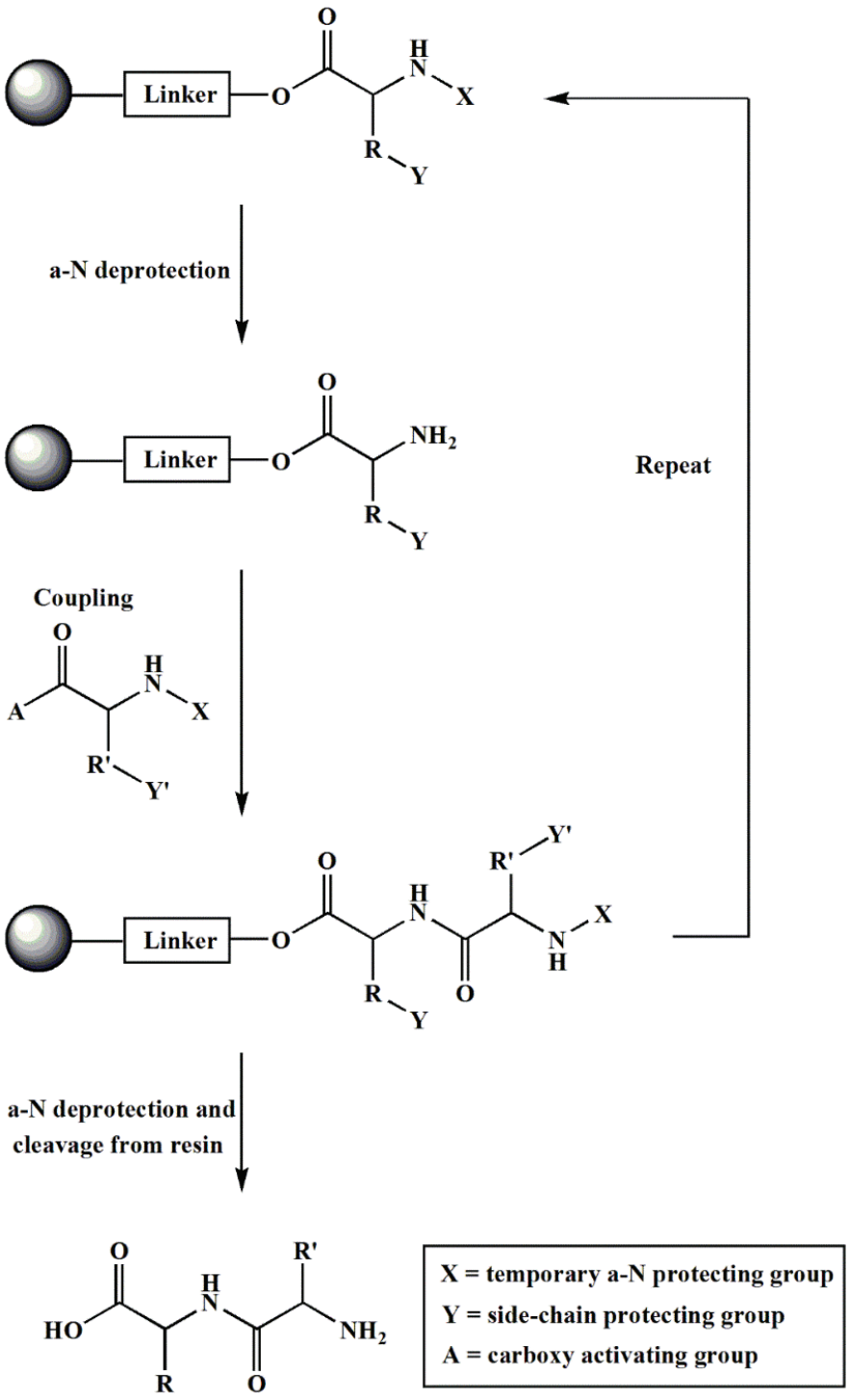


Figure 5.1 Solid Phase Peptide Synthesis (SPPS).

Amino acids bearing special side chain protecting groups, stable to the cleavage conditions used to detach the peptide from the resin, can be used in the synthesis if the peptide to be produced has to be used further in reactions that could generate undesired modifications on the amino acid at issue.

Based on the group used to protect the α–amino N of the entering amino acid and the protections of reactive functions in the side chains, two main strategies are used for SPPS: the Boc/Bzl and the Fmoc/tBu protocol.

5.2.1 Fmoc-protocol for SPPS

The α–N is protected by the N- Fluorenylmethyloxycarbonyl (Fmoc) group (**Figure 5.2**); this protection is removed in a basic environment, by treatment with piperidine/DMF (1:4) and the cleavage of the peptide from the support can be obtained in acidic conditions, for example in trifluoroacetic acid (TFA).

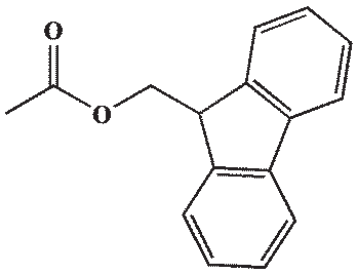


Figure 5.2 Structure of the Fmoc-protecting group.

The pairing Fmoc/tBu is truly orthogonal; TFA-labile and simultaneously base-stable groups as tBu and Boc (in combination with a TFA-labile anchor) are the perfect choice for side-chain protection. Orthogonal protection schemes permit milder overall reaction conditions as well as the synthesis of partly protected or side-chain modified peptides.
Thanks to its mild reactive conditions, peptide synthesis using the Fmoc-protocol can be automated and has become the most popular way to produce peptides in research and industry laboratories all over the world.
The removal of the Fmoc-group occurs through a base induced β-elimination (**Figure 5.3**). As a result, dibenzofulvene and carbon dioxide are split off; the former molecule is scavenged by piperidine, avoiding its irreversible attachment to the deprotected amino group.

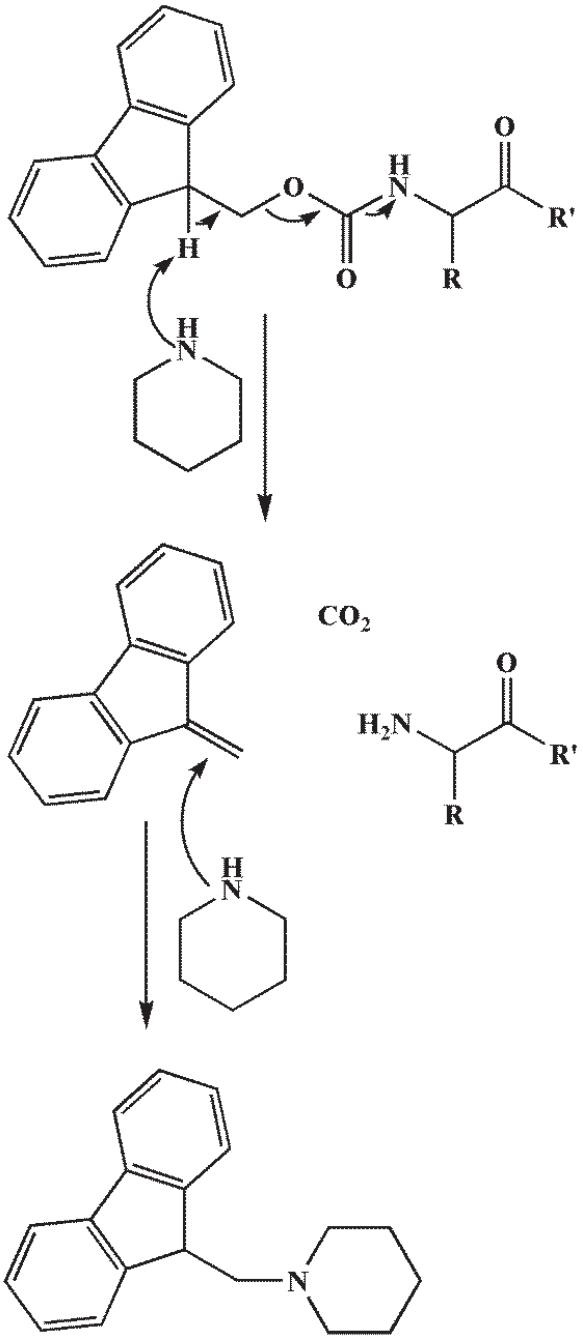


Figure 5.3 Piperidine induced Fmoc-deprotection.

Standard side-chain protecting groups are completely stable to Fmoc-deprotection and are removed during the cleavage of the peptide from the resin.

5.2.2 Amino acidic coupling: HBTU/DIPEA mediated activation

The carboxylic group of the entering amino acid needs to be made more reactive to form an amide bond with the growing peptide chain, by transformation into an active ester; to obtain this activation, coupling reagents are added to the system.

In the case of HBTU/DIPEA mediated couplings, an active ester is formed in situ by the reaction of the carboxylic group of the entering amino acid with HBTU (O-(Benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate) (**Figure 5.4**).

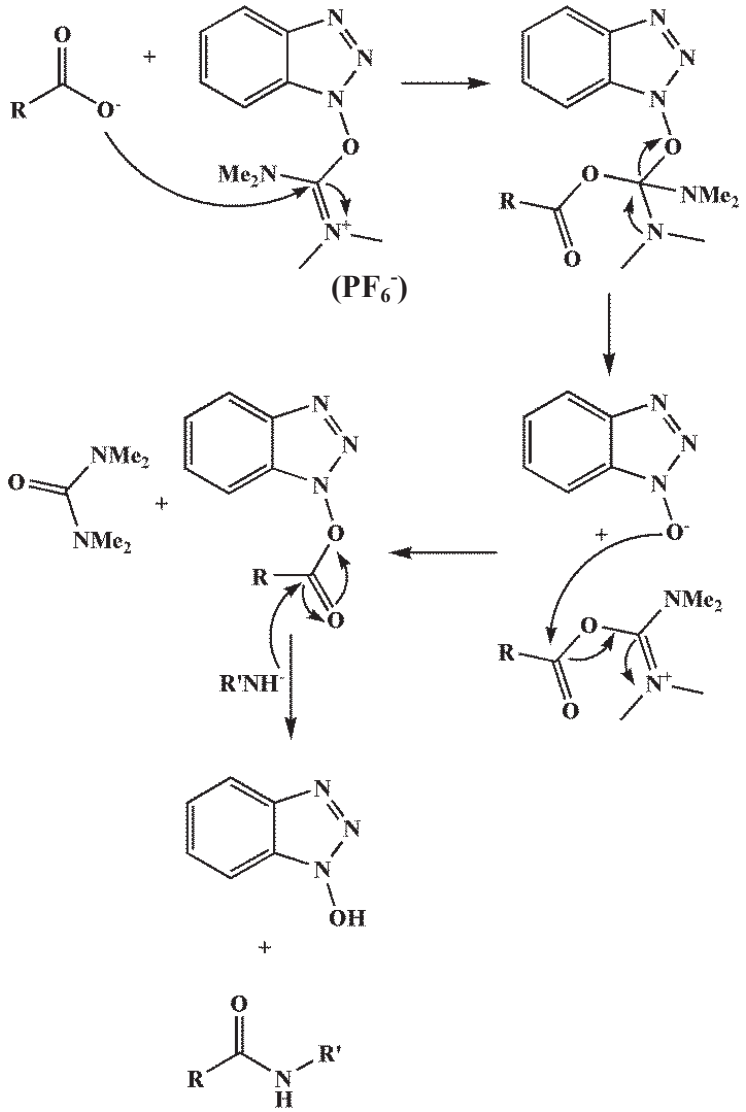


Figure 5.4 In situ activation of a carboxylic group by means of HBTU and following amide bond formation.

An excess of di-isopropylethylamine (DIPEA) is added to the system to keep it basic and to bind the released protons.
Usually 4 equivalents of amino acid and HBTU and 8 equivalents of DIPEA, to the initial loading of the resin, are employed.

5.2.3 Cleavage of the peptide from the resin

Concentrated TFA is the most commonly used reagent to perform the final cleavage of the peptide from the resin after Fmoc-SPPS. The standard side-chain protecting groups are removed concomitantly and during these processes highly reactive carbocations are produced; to avoid undesired attachment of the latter to sensitive amino acids like Cys, Met, Ser, Thr, Trp and Tyr, the released carbocations have to be trapped. For this purpose various scavengers can be added to the cleavage cocktail: H_2O, thiols like dithiotreitol (DTT) and 1,2-ethanedithiol (EDT) and sylane derivates like triisopropylsilane (TIS) and triethylsilane (TES) are commonly employed.
Usually the cleavage reaction goes to completion within 2-3 hours at room temperature.
The crude peptide product is obtained after precipitation in cold diethyl ether.
In order to have the molecule pure for further applications, usually a HPLC purification step is performed.

5.3 Chemoselective ligation reactions

Solid phase peptide synthesis gives good results for the production of linear peptides of a maximum length of about 30 amino acids; the chemical synthesis of larger peptides or even of proteins was made possible by the development of chemoselective ligation reactions, which allow to link synthetic peptides to obtain larger macromolecules.
To produce a protein by total chemical synthesis, a retrosynthetic pathway is designed, in which the target protein is divided in peptides to be bound together by ligation reactions. The use of short starting peptide fragments, that can be produced in high purity, should be preferred[5].

The hallmark of most ligation strategies known today is a capture step that binds two peptides in a chemoselective way followed by an intramolecular rearrangement[6].

5.3.1 Native Chemical Ligation (NCL)

In 1994, Kent's research group demonstrated that a chemoselective ligation occurs between C-terminal peptide-thioesters and N-terminal Cys peptides, so this reaction can be used to bind unprotected peptide fragments in solution[7] (**Figure 5.5**).

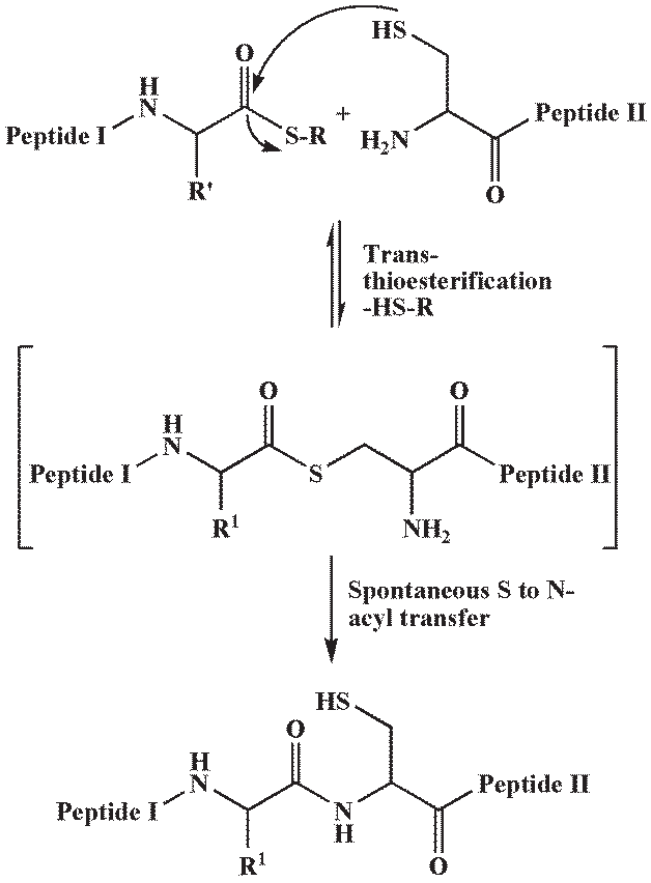


Figure 5.5 Native Chemical Ligation.

The chemoselective capture step is mediated by a reversible thiol–thioester exchange between an electrophilic (aryl) thioester at the C-terminus of the N-peptide and the nucleophilic thiol of a Cys residue located at the N-terminus of the C-peptide. In the following rearrangement, the Cys-thioester undergoes a

rapid intramolecular S→N transfer, via a favourable five-membered transition state, to form a native peptide bond between the C and the N-peptide. Importantly, additional internal Cys residues in the peptide do not interfere with the overall reaction pathway, since the irreversible intramolecular S→N shift can only occur at the unique N-terminal Cys residue. Any internal cysteines that participate in the formation of thioesters rapidly exchange backwards[6].
A major advantage of NCL is represented by the mild reaction conditions, the ligation can in fact be carried out in buffered aqueous solutions of neutral pH value and at room temperature.
The nature of the thioester has a huge impact on the NCL reaction[8]. Since alkyl-thioesters are generally less reactive than aryl-thioesters, the former are more attractive for synthesis and handling. However, NCL requires a rapid thiol–thioester exchange, which proceeds much more efficiently with aryl-thioesters. Thus, peptide thioesters are commonly synthesized as their alkyl-derivatives and converted into the corresponding aryl-thioesters in situ, by the addition of an excess of aryl-thiols, such as the water-soluble 4-mercaptophenylacetic acid (MPAA), to the reaction medium[9].

5.3.2 Ligations at cysteine-free junctions

The need for a Cys residue at the ligation site represents a particular limitation for NCL, as cysteines are not particularly abundant in natural occurring peptides and proteins. To overcome this restriction, several approaches have been developed to allow a chemoselective connection of amide bonds at cysteine-free peptide junctions.
A possible strategy consists in the use of β- and γ-mercapto amino acid derivatives at the N-terminus of the C-peptide, residue that can participate in the previously discussed capture step and subsequent nucleophilic attack on an appropriate electrophilic functionality. The corresponding amino acid can then be unmasked through desulfurization, once the two peptides are linked.
Native cysteines that might be present in the target polypeptide chain have to be protected properly to avoid their desulfurization.

5.3.2.1 Ligation at alanine

Following the above outlined strategy, ligations in correspondence of Ala residues can be easily achieved considering that the desulfurization of a Cys yields an Ala (**Figure 5.6**).

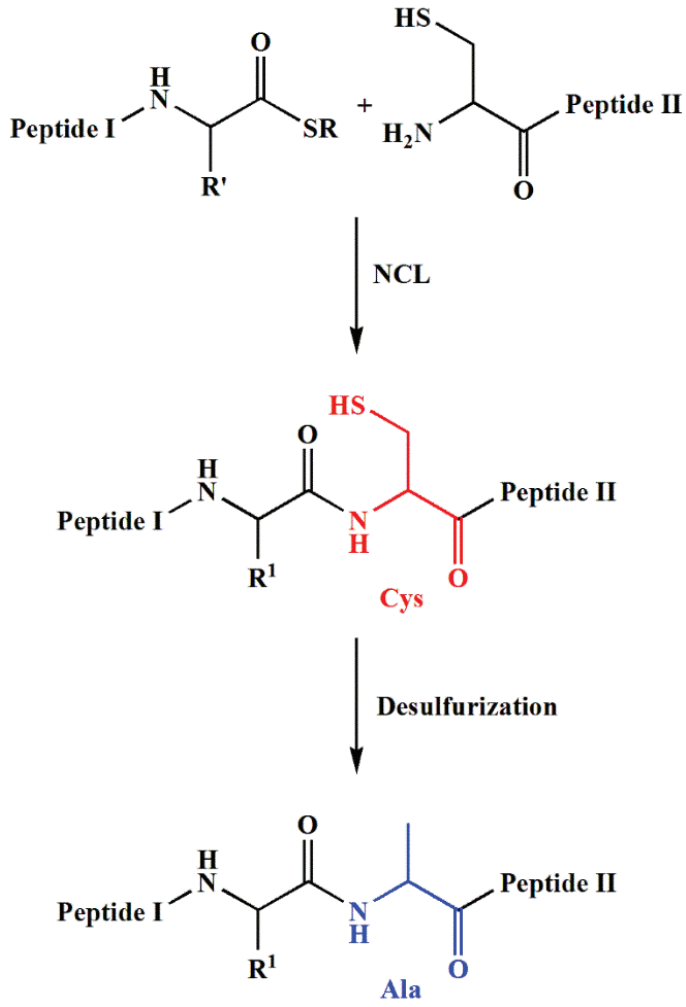


Figure 5.6 Native Chemical Ligation-Desulfurization.

This approach was first developed by Perlstein[10] and since then, various desulfurization reactions have been performed on NCL products to introduce an alanine in correspondence of the ligation site.
The first trials of desulfurization involved the use of metal catalysts, like Raney Ni or Pd, to induce the desulfurization, however aggregation and side reactions were observed in certain cases.

This disadvantage has been addressed recently by Wan and Danishefsky, who reported a metal-free radical-based desulfurization technique to convert Cys residues, in the presence of residues such as Met, Acm-protected Cys, and Thz, in high yields[11].

The developed experimental plan for the total chemical synthesis of the allergen Pru p 3 involves the use of NCL reactions coupled with metal-free desulfurization processes, to bind five peptides, constituted by a maximum of 24 amino acids, which have been defined evaluating the positions of alanines within the protein to be obtained (**Figure 5.7**).

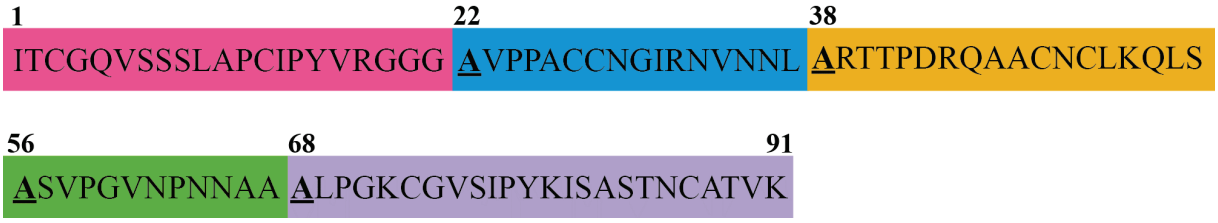


Figure 5.7 Theoretical splitting of Pru p 3.

Ligations at alanines were preferred to simple NCL processes, since these allow to split the target protein homogeneously and the starting peptides can thus more easily be produced.
Moreover, native cysteines play a crucial role in the folding of the target protein and the chosen strategy allows keeping them protected during the whole assembly and control the formation of the disulphide bridges at the end of the synthesis.

5.4 Quantification of synthetic peptides

Generally, for ligations equal molar amounts of the two reacting fragments are used, thus these need to be quantified.
Quantification of peptides by weight is generally not really accurate, as they can absorb humidity and easily crystallize together with impurities.
Peptides containing tryptophan or tyrosine residues can be quantified exploiting the ultraviolet absorption of the latter: at neutral pH, their molar extinction

coefficients at 280 nm are 5690 (Trp) and 1280 (Tyr) $M^{-1}cm^{-1}$.
Peptides bearing unprotected Cys residues can be quantified using Ellman's reagent[12] (DTNB, 5,5′-Dithiobis(2-nitrobenzoic acid)): this compound reacts stoichiometrically with free thiols releasing one equivalent of a yellow p-nitrothiophenol anion (TNB^{2-}, **Figure 5.8**), which has a molar extinction coefficient of 13,600 $M^{-1}cm^{-1}$ at 412 nm in aqueous medium.

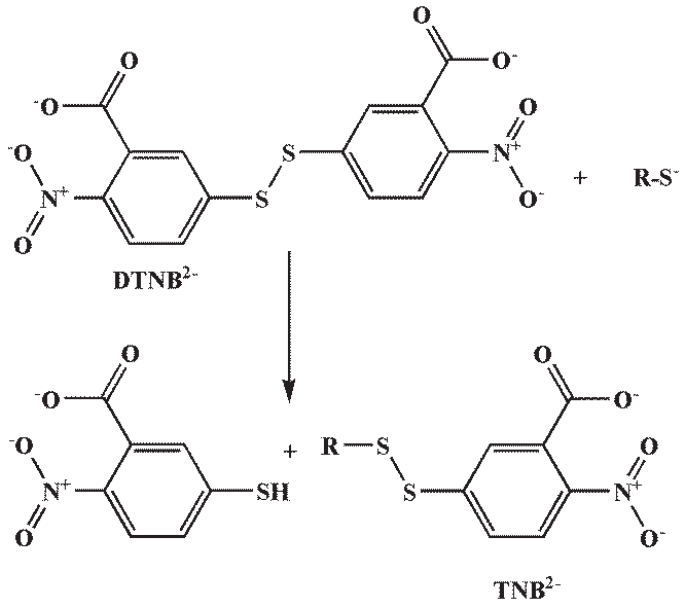


Figure 5.8 Reaction between a thiol and Ellman's reagent.

Once the light absorbance at 412 nm of a solution containing the peptide and an excess of DTNB has been measured spectrophotometrically, the concentration of the peptide can be calculated through Lambert-Beer's law.

5.5 Synthesis of C-terminal peptide-thioesters

The peptide thioesters, needed for NCL reactions, can be successfully synthesized by using Boc/Bzl chemistry[13], but their preparation by Fmoc-SPPS is much more cumbersome and often gives low yields. Nevertheless, various strategies have been developed in the recent years to make that kind of synthesis feasible.
The main problem when using Fmoc-SPPS for the production of peptide-thioesters is the lability of the thioester moiety to piperidine, which is employed

to remove the Fmoc-group after each amino acidic coupling. One possible strategy that can be used is the minimization of the aminolysis of the linker, achieved replacing piperidine with less nucleophilic bases for the Fmoc-deprotection[14].

Another option consists in masking the thioester moiety as a less reactive group, which can be transformed in thioester once the Fmoc-SPPS took place.

The design of "safety-catch" linkers includes a chemical modification reaction, which increases the reactivity of the peptide-acyl group. The desired peptide thioester is obtained upon nucleophilic cleavage of the activated peptidyl linkage[15]. A frequently used method of that kind is the one employing Kenner's sulphonamide linker[16]: the latter gets activated by an appropriate transformation after Fmoc-SPPS and becomes vulnerable to nucleophilic attack; thiolysis provides fully protected peptide-thioesters, which are then treated with TFA for global deprotection.

5.5.1 Nbz-peptides as C-terminal thioester equivalents

Blanco-Canosa and Dawson recently introduced a new safety-catch principle[17]: the peptide is assembled on a 3,4-diaminobenzoyl linker (Dbz) which is transformed in N-acylbenzimidazolone (Nbz) after the Fmoc-SPPS, prior to the detachment of the peptide from the resin. The peptide-Nbz produced after standard cleavage (TFA), undergoes rapid thiolysis in neutral aqueous buffers, so it can be easily transformed in a C-terminal peptide-thioester.

The lability of the Nbz group to the conditions used during NCL reactions also allows to use Nbz-peptides directly as reagents for ligations, without prior conversion to thioester-peptides (**Figure 5.9**).

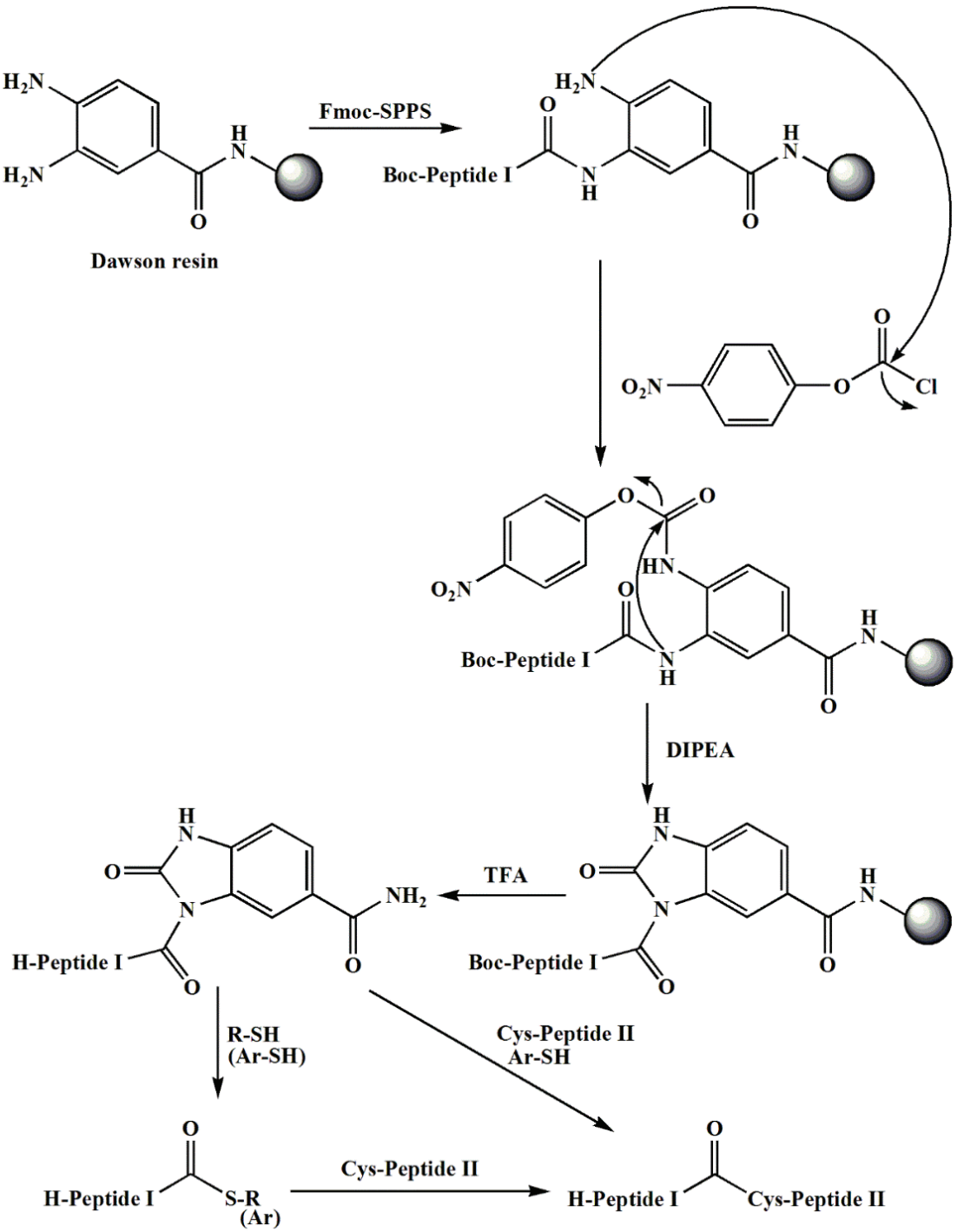


Figure 5.9 Synthesis of Nbz-peptides and their use in NCL.

As the reactive (p-nitrophenyl chloroformate) used for the closure of the Nbz moiety interacts with free amino groups, the N-terminal amino acid of the peptide chain is introduced as Boc-protected on its α-N during Fmoc-SPPS, to avoid side-reactions on it. The Boc-group is then removed, together with all the

side-chain protecting groups, during the cleavage of the peptide from the resin. The usefulness of this approach, which was chosen for the strategy we developed for the production of Pru p 3, has been demonstrated in the total chemical synthesis of various proteins, like HIV-1 Tat protein[18] and the second type 1 repeat of thrombospondin-1[19].

5.6 Linear vs convergent synthesis

Once the starting fragments have been defined, two different strategies can be used to assembly the whole protein:

- Linear synthesis: the peptide fragments are bound sequentially, starting from the C-terminal of the protein (**Figure 5.10 A**).
- Convergent synthesis: bigger building blocks are formed at the same time, binding together single peptide fragments; this kind of process is repeated until the whole protein is obtained (**Figure 5.10 B**).

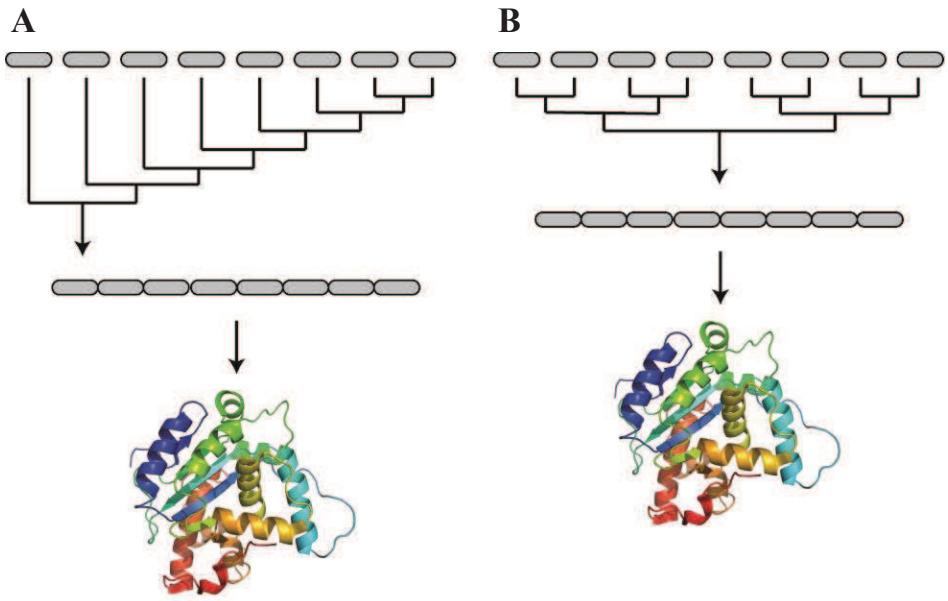


Figure 5.10 A Linear (**A**) and convergent (**B**) synthesis of a protein.

In linear syntheses, the overall yield quickly drops with each reaction step, while higher yields can be obtained through a convergent synthesis, as each peptide fragment undergoes a smaller number of steps before being incorporated in the final product. Thus, convergent approaches are preferable when planning the synthesis of a protein.

Using NCL reactions, proteins can easily be assembled in a linear fashion through C-to-N sequential ligations: the C-terminal peptide of the target protein is bound to the thioester-peptide of the second fragment which is protected on its N-terminal to avoid auto-condensation or cyclization; once the ligation has occurred, the N-terminal Cys of the obtained product is unmasked so that the third peptide can be bound and so on. The N-terminal Cys of the entering peptide is typically masked as L-thiazolidine-4-carboxylic acid (Thz); conversion of the joined Thz-peptide product to a Cys-peptide can be achieved by simply adding methoxyamine hydrochloride to the crude ligation mixture and lowering the pH to 4 (**Figure 5.11**).

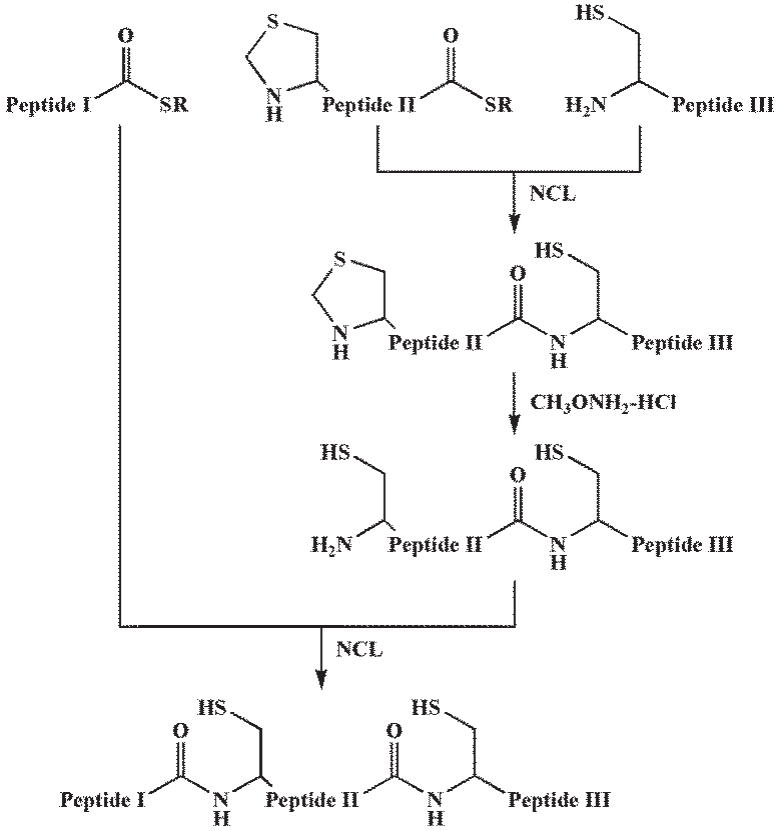


Figure 5.11 C to N sequential synthesis of a protein using Thz as a protecting group for Cys.

Convergent protein syntheses by NCL are more challenging because the C-terminal thioester moieties cannot be masked as easily as an N-terminal Cys and the N-to-C ligations are so less straightforward.
To overcome this problem in 2006 a kinetically controlled convergent synthesis

method (KCL) was developed, based on the differential reactivities of different types of thioesters[20]: Bang and his group exploited the knowledge that a Cys-peptide would react with a peptide-thioarylester much faster than it would with a peptide-thioalkylester. It was conjectured that the large reactivity difference would allow a thioalkylester functional group of a peptide-thioalkylester to remain unreacted in the same solution with a thioarylester, under competitive reaction conditions and in the absence of exogenous aryl-thiol. The practicality of this approach was illustrated through a fully convergent synthesis of Crambin from six unprotected peptide fragments[20]. Since then, KCL has been successfully employed for the synthesis of various proteins, however sometimes the reaction does not proceed smoothly because the dual reactivity of the middle segment is sensitive to the steric hindrance of the ligation sites[21].

5.7 Peptide-hydrazides as thioester equivalents

Recently, it has been proposed that an hydrazide moiety could be used as an efficient protecting group for C-terminal thioesters[22]: the C-terminal hydrazide is stable in standard ligation conditions and a Cys-peptide-hydrazide can chemoselectively undergo a NCL reaction at its N-terminal Cys; once this took place, the hydrazide moiety can be transformed in thioester. This conversion occurs in two steps that are carried out in a one-pot fashion: first, the peptide hydrazide is oxidized to peptide-azide, treating with $NaNO_2$ at low temperature (-10^0 C) and low pH (3), then a thiol compound is added and the pH is adjusted to 7 to initiate the conversion to C-terminal thioester (**Figure 5.12**).

Figure 5.12 Peptide hydrazide to peptide thioester conversion.

The produced peptide thioester can be directly used for NCL without isolation, just adding the desired peptide bearing a Cys at its N-terminal to the reaction medium. The whole process is named **peptide-hydrazide ligation**.
Importantly, the brief treatment with $NaNO_2$ does not cause any oxidation for

Met and Trp and unprotected Ser, Thr, Tyr, His, Lys, Asp, Glu, and even Cys were found to be compatible with the above process[22].

Two possible strategies were developed for the assembly of Pru p 3 in a convergent fashion, exploiting the possibility to selectively achieve C to N ligations masking the cysteine as Thz and N to C ligations using peptide-hydrazides as thioester equivalents (**Figure 5.13**). The Acm group was chosen as side chain protection for native cysteines, as it is stable to the reactive conditions of the processes foreseen for the protein assembly.

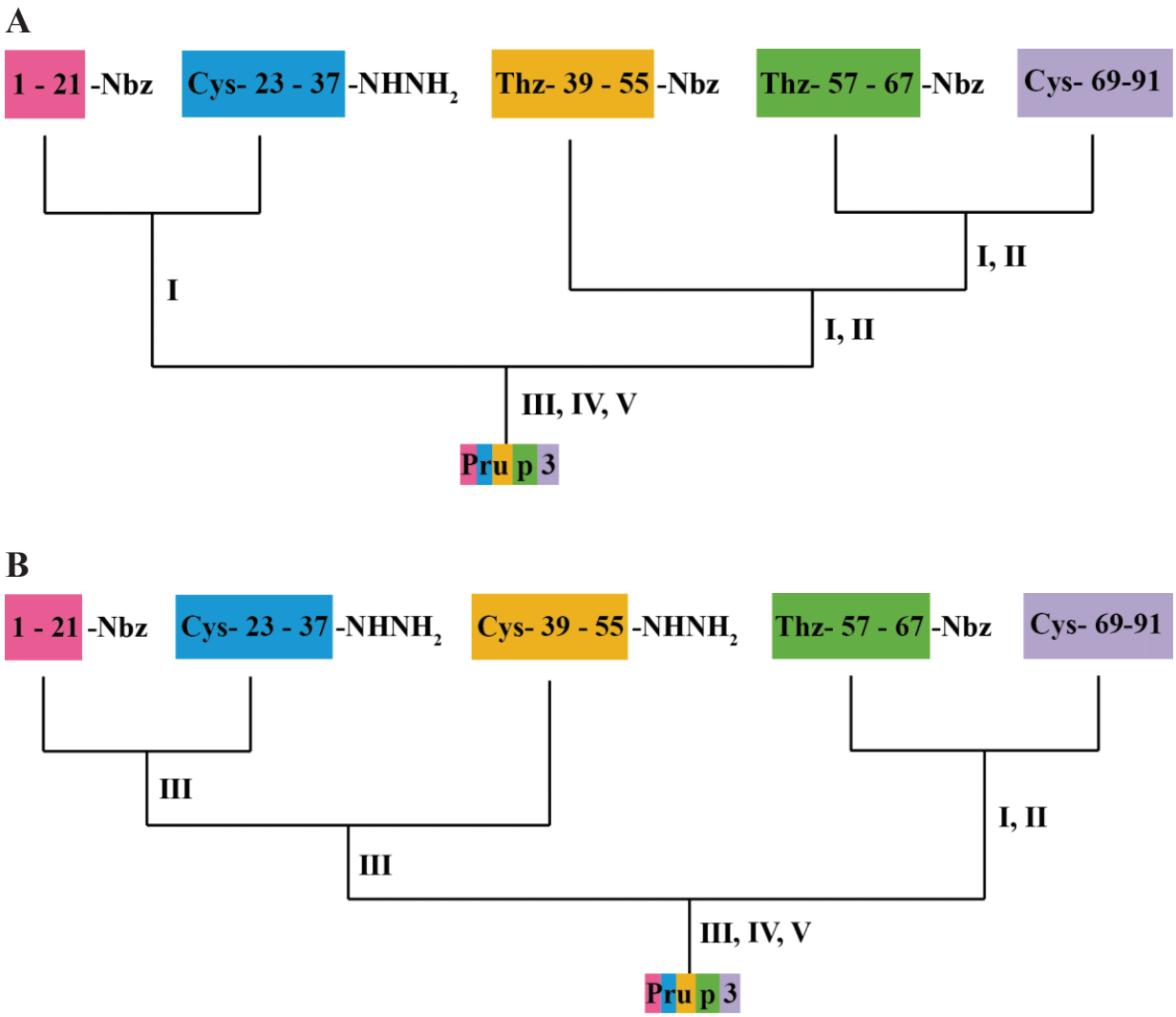


Figure 5.13 Strategies developed for the convergent assembly of Pru p 3. I: NCL; II: Thiazolidine deprotection; III: peptide-hydrazine ligation; IV: Cys desulfurization; V: Acm deprotection.

The needed peptide-hydrazides can conveniently be obtained starting from Nbz-peptides, by nucleophilic substitution of the Nbz group. Since we decided to use Nbz-peptides also as thioester equivalents for C to N ligations, the C-terminal functionalised peptide fragments to be bound can all be produced using the same solid support (Dawson resin).

5.8 References

[1] Kent S.B.H. Total chemical synthesis of proteins. Chem. Soc. Rev. 38, 338–351, **2009**.

[2] Nilsson B.L., Soellner M.B., Raines R.T. Chemical synthesis of proteins. Annu. Rev. Biophys. Biomol. Struct. 34, 91–118, **2005**.

[3] Harmand T.J., Murar C.E., Bode J.W. New chemistries for chemoselective peptide ligations and the total synthesis of proteins. Curr. Opin. Chem. Biol. 22, 115-21, Oct **2014**.

[4] Merrifield R.B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J. Am. Chem. Soc. 85, 2149–2154, **1963**.

[5] Lee J.Y., Bang D. Challenges in the Chemical Synthesis of Average Sized Proteins: Sequential vs. Convergent Ligation of Multiple Peptide Fragments. Biopolymers. 94(4), 441-447, **2010**.

[6] Hackenberger C.P.R., Schwarzer D. Chemoselective Ligation and Modification Strategies for Peptides and Proteins Angew. Chem. Int. Ed. 47, 10030–10074, **2008**.

[7] Dawson P.E., Muir T.W., Clark-Lewis I., Kent S.B.H. Synthesis of proteins by native chemical ligation. Science 266, 776 – 779, **1994**.

[8] Dawson P.E., Churchill M.J., Ghadiri M.R., Kent S.B.H. Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives. J. Am. Chem. Soc. 119, 4325 – 4329, **1997**.

[9] Johnson E.C.B., Kent S.B.H. Insights into the mechanism and catalysis of the native chemical ligation reaction. J. Am. Chem. Soc. 128, 6640 – 6646, **2006**.

[10] Perlstein M.T., Atassi M.Z., Cheng S.H. Desulfurization of sulfur amino acids and proteins with Raney nickel. Biochim. Biophys. Acta Protein Struct. 236, 174–182, **1971**.

[11] Wan Q., Danishefsky S.J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew. Chem. Int. Ed. 46, 9248–9252, **2007**.

[12] Ellman G.L. Tissue sulfhydryl groups. Arch. Biochem. Biophys. 82(1), 70-77, May **1959**.

[13] Hojo H., Aimoto S. Polypeptide Synthesis Using the S-Alkyl Thioester of a Partially Protected Peptide Segment. Synthesis of the DNA-Binding Domain of c-Myb Protein (142–193)–NH_2. Bull. Chem. Soc. Jpn. 64, 111–117, **1991**.

[14] Li X.Q., Kawakami T., Aimoto S. Direct preparation of peptide thioesters using an Fmoc solid-phase method. Tetrahedron Lett. 39, 8669–8672, **1998**.

[15] Seitz O., Mende F. 9-Fluorenylmethoxycarbonyl-based solid-phase synthesis of peptide α-thioesters. Angew. Chem. Int. Ed. 50, 1232–1240, **2011**.

[16] Kenner G.W.M., McDermott J.R., Sheppard R.C. The safety catch principle in solid phase peptide synthesis. J. Chem. Soc. Chem. Commun. 636 – 637, **1971**.

[17] Blanco-Canosa J.B., Dawson P.E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. Angew. Chem. Int. Ed. 47, 6851–6855, **2008**.

[18] Harpaz Z., Siman P., Kumar K.S.A., Brik A. Protein synthesis assisted by native chemical ligation at leucine. Chem. Bio. Chem. 11, 1232–1235, **2010**.

[19] Tiefenbrunn T.K., Blanco-Canosa J., Dawson P.E. Alternative chemistries for the synthesis of thrombospondin-1 type 1 repeats. Biopolymers 94, 405–413, **2010**.

[20] Bang D., Kent S.B. Kinetically controlled ligation for the convergent chemical synthesis of proteins. Angew. Chem. Int. Ed. 45, 3985–3988, **2006**.

[21] Lee J., Kwon Y., Pentelute B.L., Bang D. Use of model peptide reactions for the characterization of kinetically controlled ligation. Bioconjugate Chem. 22, 1645–1649, **2011**.

[22] Fang G.M., Li Y.M., Shen F., Huang Y.C., Li J.B., Lin Y., Cui H.K., Liu L. Protein chemical synthesis by ligation of peptide hydrazides. Angew Chem Int Ed, 50, 7645–7649, **2011**.

6 Total chemical synthesis approach of the allergen Pru p3

6.1 Introduction

In the present work the total chemical synthesis of Pru p 3 was approached; two previously developed strategies for the convergent assembly of the target protein, starting from five peptide fragments to be produced by Fmoc-SPPS and to be bound by means of native chemical ligation (NCL) and peptide-hydrazide ligation, followed by desulfurization, were tested.
All the reaction conditions were set up and the five parts were bound together to obtain the whole protein stretch.
Finally, immunological tests and structural characterisation by means of circular dichroism were performed on the two final protein halves foreseen by the chosen synthetic strategy: peptides sPru p 3 (1-37) and sPru p 3 (38-91).
The immunological part was performed at the Department of Experimental Immunology, Academic Medical Center, Amsterdam. The research group of Prof. Ronald van Ree is gratefully acknowledged.

6.2 Results and discussion

6.2.1 Choice of the strategy

The two previously developed strategies for the assembly of the target protein in a convergent fashion (Chapter 5), were tested in the present work.
To avoid side reactions due to the double reactivity of the central fragments, temporary protection is needed for the terminals that are not required to react in the first ligation involving the fragment.
Protection of N-terminal Cys was achieved introducing the residue in the form of L-thiazolidine-4-carboxylic acid (Thz) during peptide synthesis[1], while C-terminal thioester peptides were masked as C-terminal hydrazide peptides[2].
As in preliminary tests we performed, the thiazolidine deprotection turned out to be straightforward and less cumbersome than the peptide-hydrazide to

peptide-thioester conversion, the strategy involving two NCL and only one peptide-hydrazide ligation was chosen for the assembly of the target protein (**Figure 6.1**).

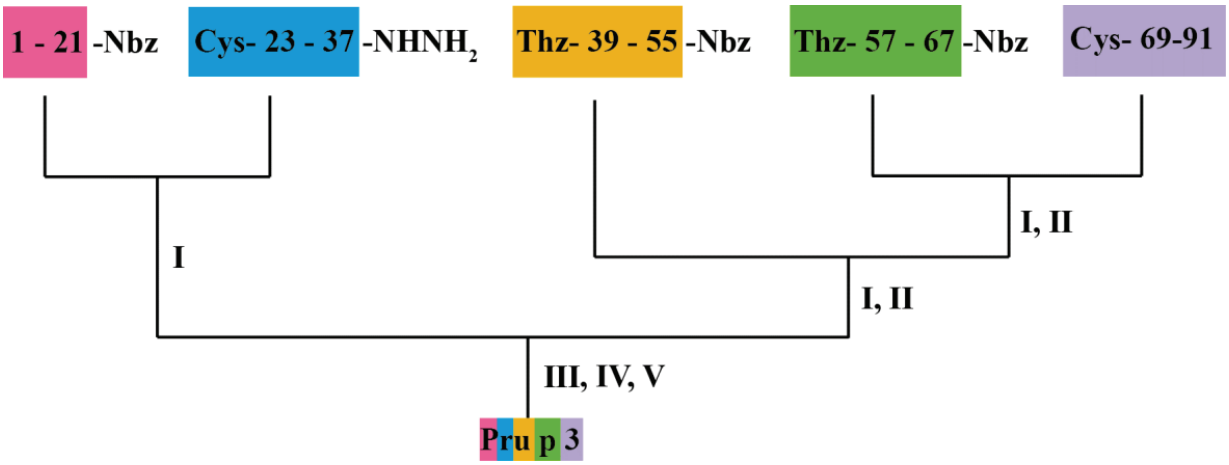


Figure 6.1 Chosen strategy for the convergent assembly of Pru p 3. I: NCL; II: Thiazolidine deprotection; III: peptide-hydrazine ligation; IV: Cys desulfurization; V: Acm deprotection.

The side chain of native cysteines was protected against desulfurization with the acetaminomethyl (Acm) group, which was shown to be stable in the reaction conditions used during the protein assembly and desulfurization of unprotected Cys[3, 4].

6.2.2 Peptide synthesis

Peptides were synthesized according to the Fmoc-SPPS protocol, using an automated peptide synthesizer.
The C-terminal fragment Cys-[69-91] was produced employing a Novasyn TGA resin preloaded with lysine; this kind of resin is suitable for the synthesis of long peptides, as it has a low loading (0.2 mmol/g) and is constituted of beads with large internal cavities. Multiple couplings were used to drive to completion the attachment of residues in "difficult" positions (i.e. following prolines). Previous trials of synthesis of the same peptide employing a standard preloaded Wang resin (0.5 mmol/g) or a manually preloaded HMPB

Chemmatrix resin (0.2 mmol/g), leaded to poor yields because of the production of big amounts of truncated fragments.
The peptide Cys-[23-37], needed to produce the protein half sPru p 3 (1-37), was synthesized in standard conditions for SPPS, employing a Wang resin preloaded with Leu.
The peptides [1-21]-Nbz, Cys-[23-37]-Nbz, Thz-[39-55]-Nbz and Thz-[57-67]-Nbz were synthesized using Dawson Dbz AM resin, on which the C-terminal amino acid was loaded manually, according to manufacturer's instructions. Amino acid solutions with low concentrations (0.15 M instead of 0.5 M) were used for the peptide-chain assembly, to avoid attachment to the free amino-group on Dawson's linker[5].
The on-resin conversion of the Dbz-group to Nbz has been achieved through a two-step treatment; according to manufacturer's instruction the first step of this process can be performed using dichloromethane (DCM) as solvent, however the conversion to Nbz was only partial when we tried this protocol for our peptides. This could be due to the formation of aggregates of the peptides in DCM, hindering the attack of the nucleophile to the linker. The same was already observed by others, who solved the problem using N,N-dimethylformamide (DMF) as a solvent for the first step of the process[6]; we tried this approach too, but without complete success. Conversely, the reaction went to completion when the first step was carried out in LiBr 2 M in tetrahydrofuran (THF); this solvent system effectively disrupts secondary structures that can be formed by peptides on resin[7], thus making the reaction site accessible to the incoming nucleophile.
The same solvent system had also efficiently been used to increase the product yield for the synthesis of C-terminal peptide thioesters on Kenner's "Safety-Catch" resins[8].

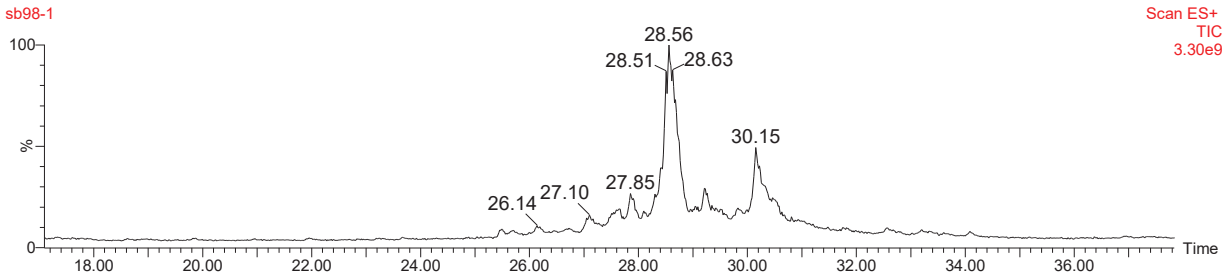
6.2.3 Cleavage of the peptides from the resin

The cleavage of the peptides from the solid support was accomplished through a treatment with TFA. Contextually, the acid-labile side-chain protecting group were removed from the peptides; to block the carbocations formed as a consequence of this process and to avoid their irreversible attachment to the

reaction product, scavengers like DTT, H_2O and TIS were added to the cleavage cocktail.
The synthesized peptides were characterized by means of UPLC-MS:

- Peptide [1-21]-Nbz

A



B

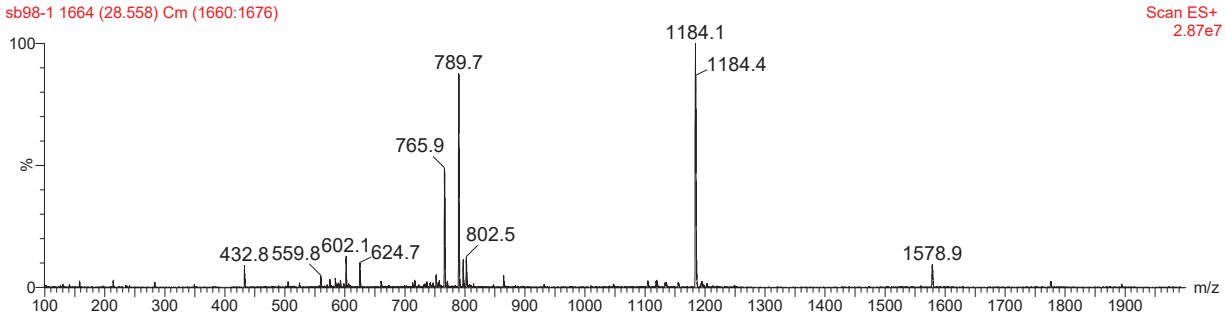
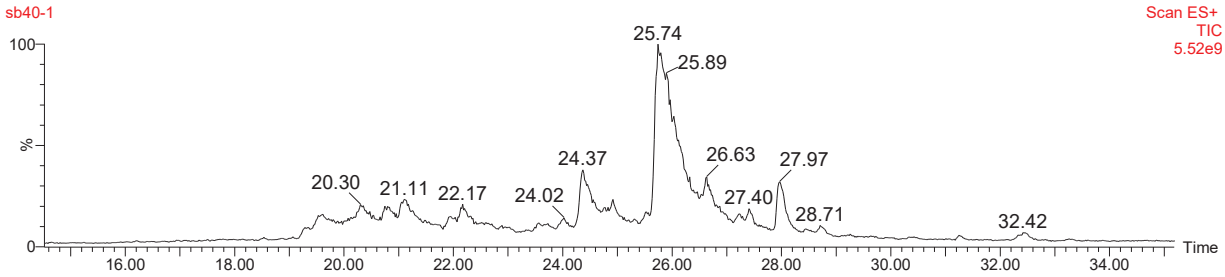


Figure 6.2 Chromatogram of the crude product obtained from the synthesis of the peptide [1-21]-Nbz (**A**); mass spectrum corresponding to the peak at rt = 28.56' (**B**).

The main peak in the reported chromatogram (**Figure 6.2 A**), at rt = 28.56', corresponds to the desired synthesis product ITC(Acm)GQVSSSLAPC(Acm) IPYVRGGG-Nbz (calculated avg MW= 2366.4; found ions: 1184.1 m/z $[M+2H]^{2+}$, 789,7 m/z $[M+3H]^{3+}$ (**Figure 6.2 B**)).

- Peptide Cys-[23-37]-Nbz

A



B

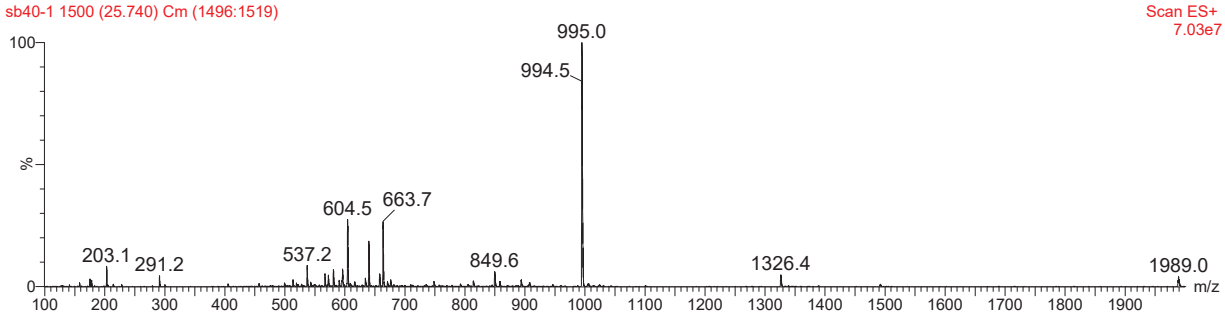
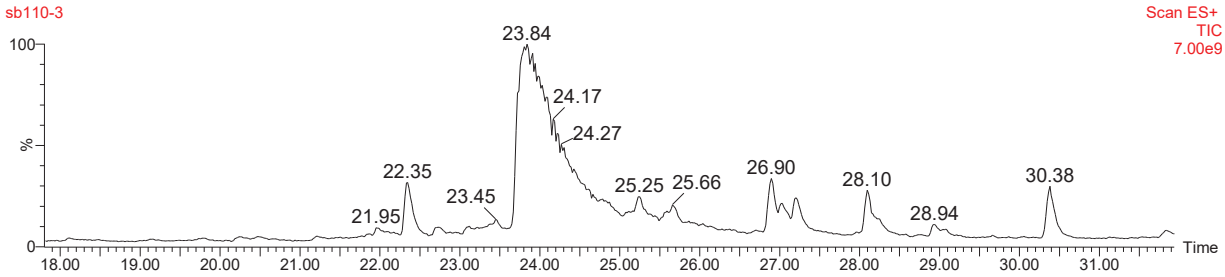


Figure 6.3 Chromatogram of the crude product obtained from the synthesis of the peptide Cys-[23-37]-Nbz (**A**); mass spectrum corresponding to the peak at rt = 25.74' (**B**).

The main peak in the reported chromatogram (**Figure 6.3 A**), at rt = 25.74', corresponds to the desired synthesis product CVPPAC(Acm)C(Acm) NGIRNVNNL-Nbz (calculated avg MW= 1988.0; found ions: 1989.0 m/z $[M+H]^+$, 995.0 m/z $[M+2H]^{2+}$, 663.7 m/z $[M+3H]^{3+}$ (**Figure 6.3 B**)).

- Peptide Cys-[23-37]

A



B

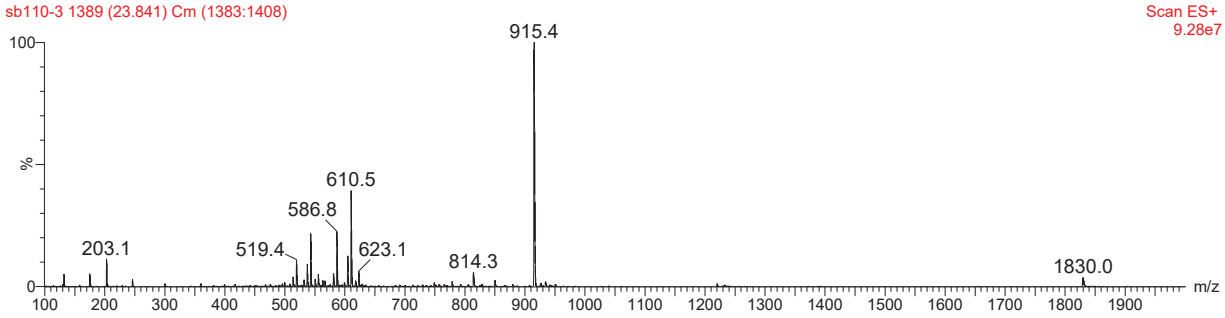
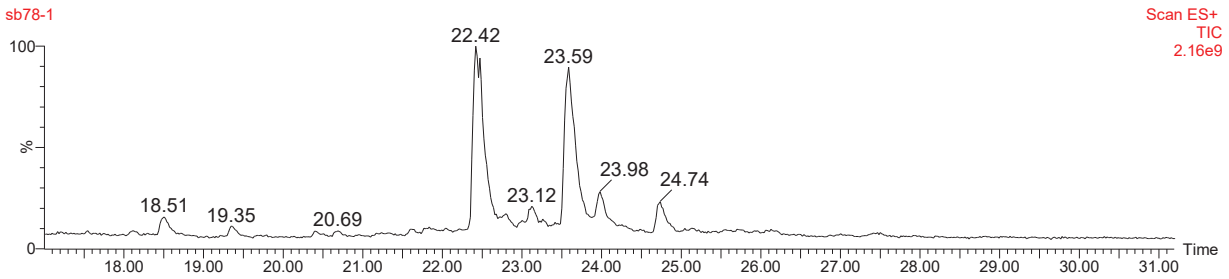


Figure 6.4 Chromatogram of the crude product obtained from the synthesis of the peptide Cys-[23-37] (**A**); mass spectrum corresponding to the peak at rt = 23.84' (**B**).

The main peak in the reported chromatogram (**Figure 6.4 A**), at rt = 23.84', corresponds to the desired synthesis product CVPPAC(Acm)C(Acm) NGIRNVNNL (calculated avg MW= 1829.0; found ions: 1830.0 m/z $[M+H]^+$, 915.4 m/z $[M+2H]^{2+}$ (**Figure 6.4 B**)).

- Peptide Thz-[39-55]-Nbz

A



B

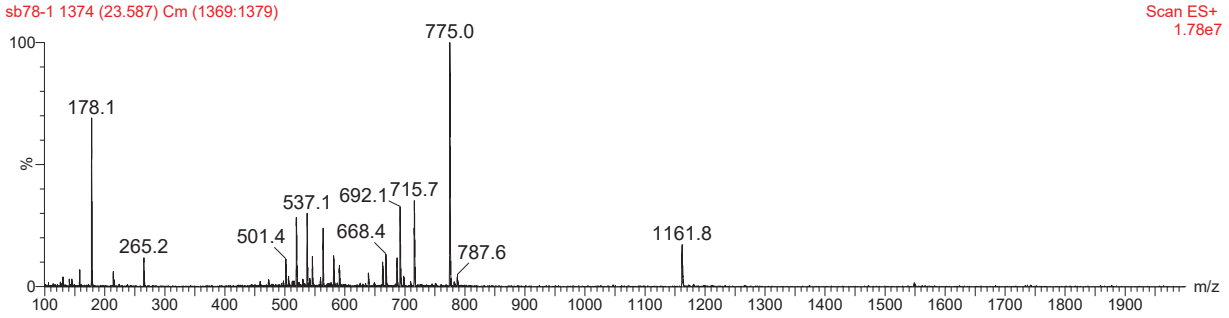
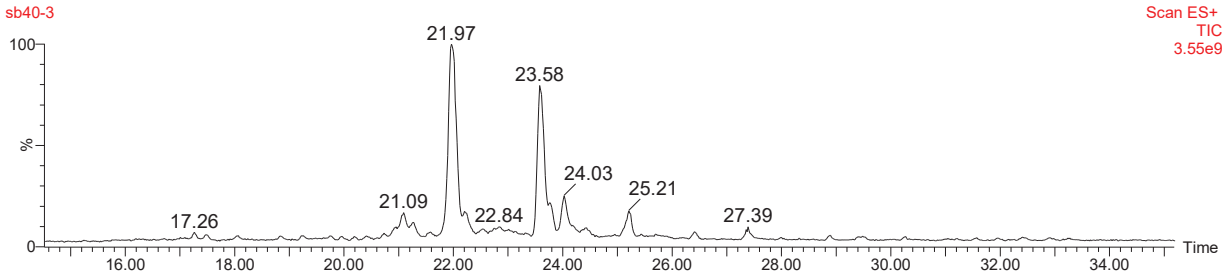


Figure 6.5 Chromatogram of the crude product obtained from the synthesis of the peptide Thz-[39-55]-Nbz (**A**); mass spectrum corresponding to the peak at rt = 23.59' (**B**).

The peak at rt = 23.59' in the reported chromatogram (**Figure 6.5 A**), corresponds to the desired synthesis product ThzRTTPDRQAAC(Acm) NC(Acm)LKQLS-Nbz (calculated avg MW= 2321.3; found ions: 1161.8 m/z $[M+2H]^{2+}$, 775.0 m/z $[M+3H]^{3+}$ (**Figure 6.5 B**)). The peak at rt = 22.42' is relative to the product peptide without the C-terminal Nbz moiety; this specie is produced by hydrolysis in the sample used for the UPLC-MS analysis, due to the lability of the Nbz group in H_2O. When used in ligation reactions, almost the whole amount of peptide is quickly converted in thioester by reaction with the MPAA contained in the buffer, avoiding the formation of significant amounts of the unreactive hydrolysed specie.

- Peptide Thz-[57-67]-Nbz

A



B

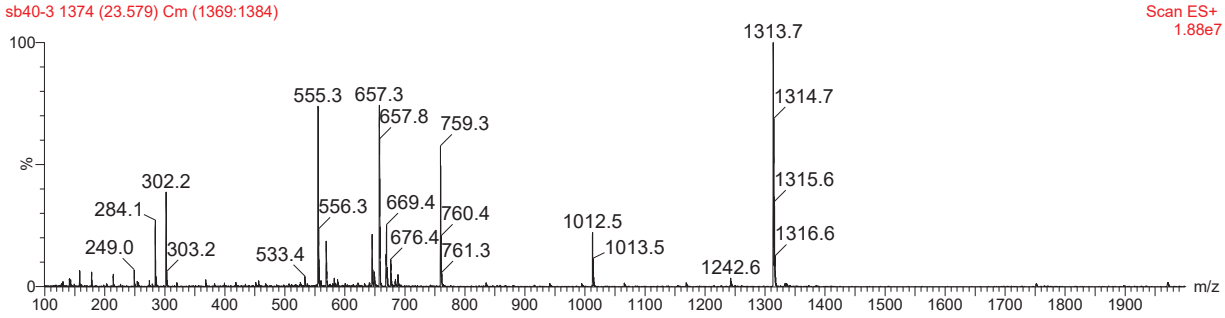
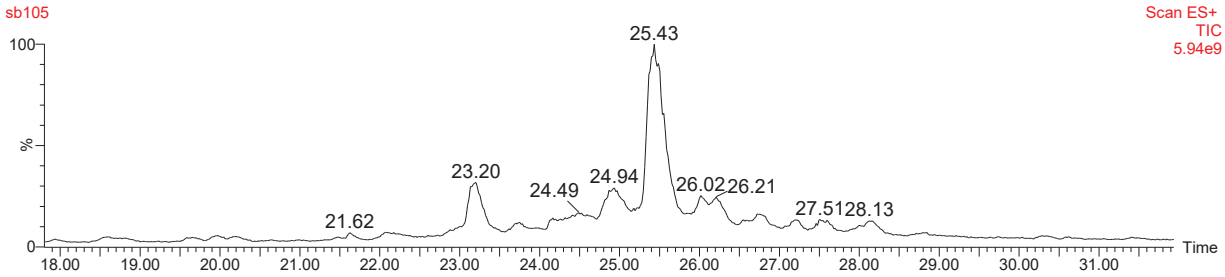


Figure 6.6 Chromatogram of the crude product obtained from the synthesis of the peptide Thz-[57-67]-Nbz (**A**); mass spectrum corresponding to the peak at rt = 23.58' (**B**).

The peak at rt = 23.58' in the reported chromatogram (**Figure 6.6 A**), corresponds to the desired synthesis product ThzSVPGVNPNNAA-Nbz (calculated avg MW= 1313.2; found ions: 1313.7 m/z $[M+H]^{+}$, 657.3 m/z $[M+2H]^{2+}$ (**Figure 6.6 B**)). The peak at rt = 21.97' corresponds to the product peptide without the C-terminal Nbz moiety, produced by hydrolysis in the sample used for the UPLC-MS analysis.

- Peptide Cys-[69-91]

A



B

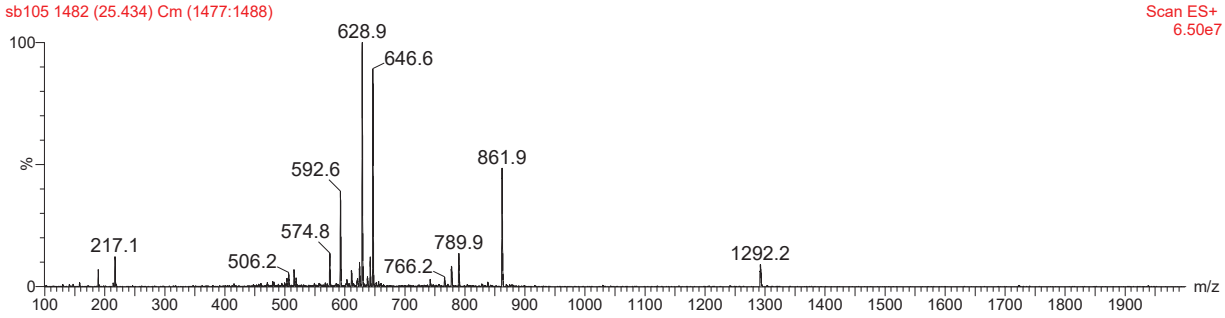


Figure 6.7 Chromatogram of the crude product obtained from the synthesis of the peptide Cys-[69-91] (**A**); mass spectrum corresponding to the peak at rt = 25.43' (**B**).

The main peak in the reported chromatogram (**Figure 6.7 A**), at rt = 25.43', corresponds to the desired synthesis product CLPGKC(Acm)GVSIPYKISA STNC(Acm)ATVK (calculated avg MW= 2582.9; found ions: 1292.2 m/z $[M+2H]^{2+}$, 861.9 m/z $[M+3H]^{3+}$, 646.6 m/z $[M+4H]^{4+}$ (**Figure 6.7 B**)).

6.2.4 Peptide-Nbz to peptide-$NHNH_2$ conversion

The Nbz moiety in peptide Cys-[23-37]-Nbz was readily transformed in C-terminal hydrazide through nucleophilic substitution in an aqueous buffer at neutral pH (**Figure 6.8**).

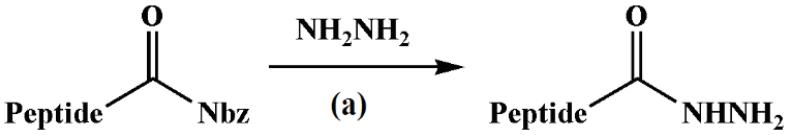
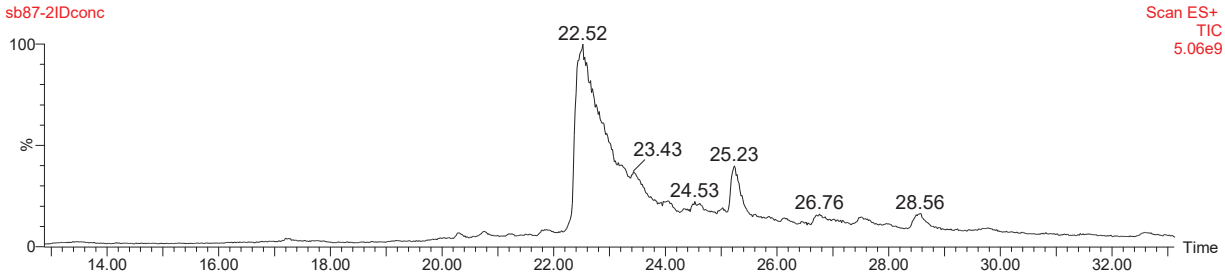


Figure 6.8 Peptide-Nbz to peptide-$NHNH_2$ conversion; (**a**): Gn-HCl 6 M, NaH_2PO_4 0.2 M, DTT 20 mM and NH_2NH_2-H_2O 0.2 M; pH=7.

The addition of DTT turned out to be crucial for the efficient and complete conversion, as we obtained complex mixtures in trials conducted without reducing agent or using as such TCEP-HCl, the reducing agent employed for NCL reactions.

An UPLC-MS analysis was performed to characterize the reaction product:

A



B

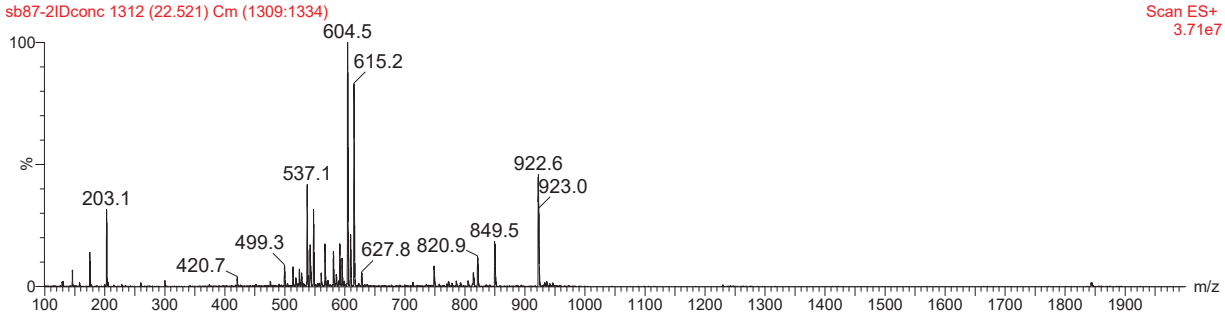


Figure 6.9 Chromatogram of the crude product obtained from the conversion of the peptide Cys-[23-37]-Nbz in Cys-[23-37]-$NHNH_2$ (**A**); mass spectrum corresponding to the peak at rt = 22.52' (**B**).

The main peak in the reported chromatogram (**Figure 6.9 A**), at rt = 22.52', corresponds to the desired reaction product CVPPAC(Acm)C(Acm)NGIRNV NNL-$NHNH_2$ (calculated avg MW= 1843.0; found ions: 922.6 m/z $[M+2H]^{2+}$, 615.2 m/z $[M+3H]^{3+}$ (**Figure 6.9 B**)).

6.2.5 Native Chemical Ligation

All the synthesized peptides were directly used in ligation reactions (**Figure 6.10**), without prior purification.

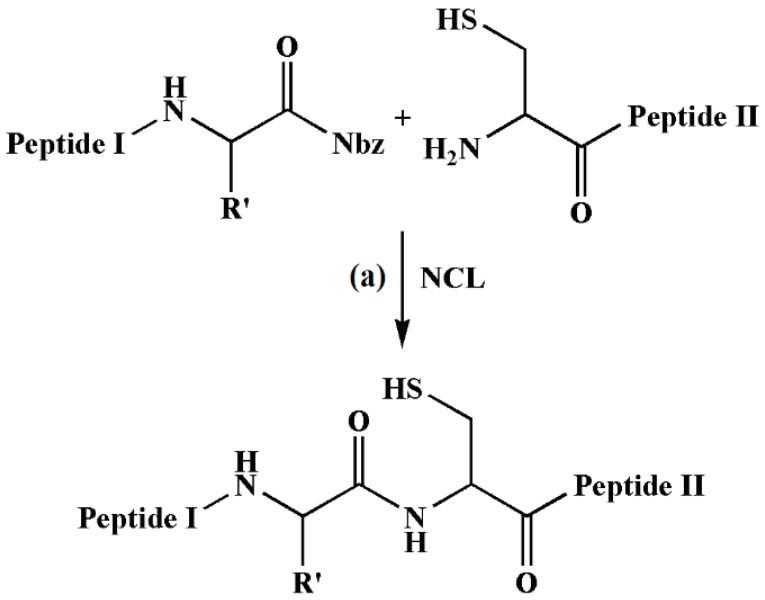


Figure 6.10 Native chemical ligation (NCL); (**a**): Gn-HCl 6 M, NaH_2PO_4 0.2 M, TCEP-HCl 20 mM and MPAA 0.2 M; pH=7.

Ellman's reagent was used to quantify peptides bearing a Cys at their N-terminal. Due to their lability in neutral aqueous medium, C-terminal Nbz peptides were roughly quantified by weight.

The peptide-Nbz to peptide-thioester conversion occurred in situ in the ligation buffer, where the thiol MPAA was present. To minimise the hydrolysis of the Nbz moiety prior to the conversion in C-terminal thioester, the peptide-Nbz was added to the reaction mixture only as the pH had already been adjusted to 7. As the hydrolysis could not be totally avoided, the peptide-Nbz was used in slight excess (1.2 equivalents).

All the reactions were complete after 5 h, as monitored by UPLC-MS.

Thz deprotection of the ligation products Thz-[57-67]-Cys-[69-91] and Thz-[39-55]-Cys-[57-67]-Cys-[69-91] was achieved in the NCL buffer, using a protocol previously reported in literature[9] (**Figure 6.11**).

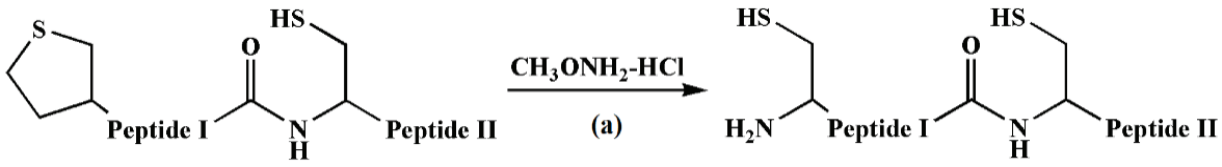


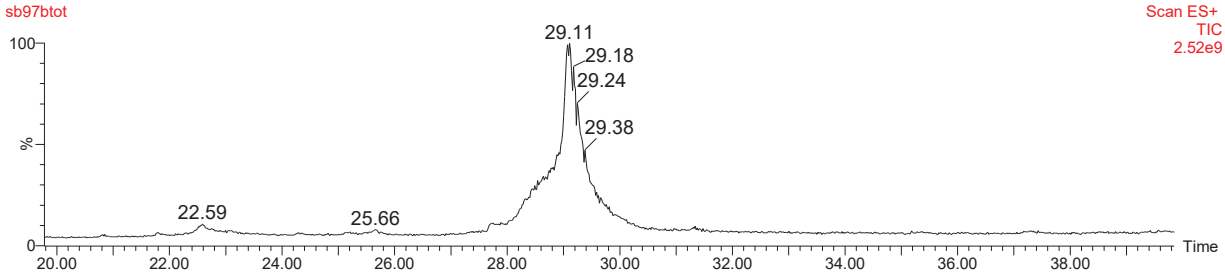
Figure 6.11 Thz deprotection of a ligation product; (**a**): Gn-HCl 6 M, NaH_2PO_4 0.2 M, TCEP-HCl 20 mM, MPAA 0.2 M and $MeONH_2$-HCl 0.4 M; pH=4.

To extract the ligation products from the ligation buffer and simplify their further purification by means of RP-HPLC, ultra-filtrations using filter devices with a nominal MW cut-off of 3 kDa were performed. Various trials were carried out using H_2O and aqueous solutions containing different concentrations of Gn-HCl for the washes of the retentate in the filters. Analysing the final retentates and the filtrates by means of UPLC-MS we established that two washes of the retentate with Gn-HCl 3 M, constituted the best way of getting rid of the low MW species (that were found only in the filtrates) while avoiding the loss of the reaction products (found exclusively in the retentates).
The used centrifugal filters were not compatible with hydrazine, therefore the peptide [1-21]-Cys-[23-37]-$NHNH_2$ was purified directly by RP-HPLC.
The gradients for the RP-HPLC purification were optimized for every single ligation product.

The purified peptides were characterized by means of UPLC-MS:

- Ligation [1-21]-Nbz + Cys-[23-37]-$NHNH_2$

A



B

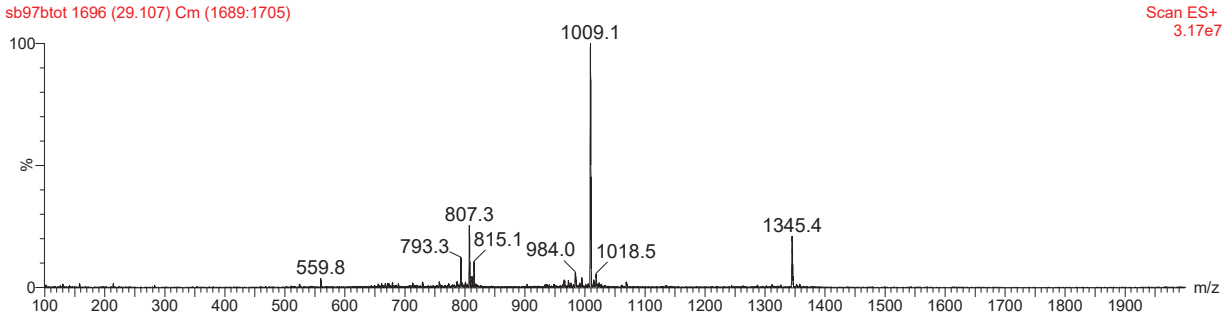
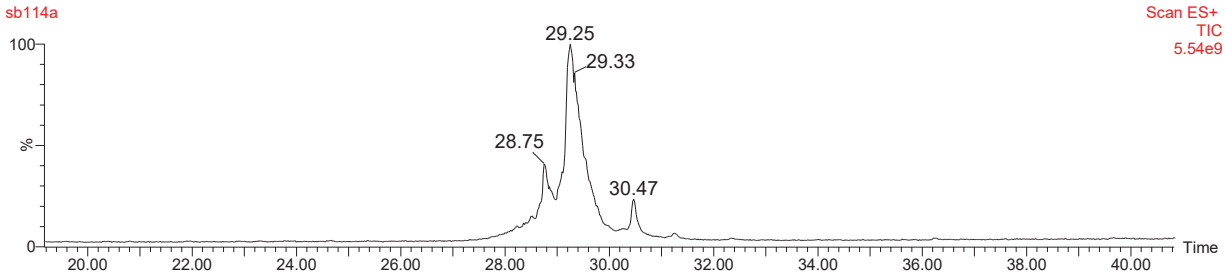


Figure 6.12 Chromatogram of the purified product obtained from the ligation [1-21]-Nbz + Cys-[23-37]-$NHNH_2$ (**A**); mass spectrum corresponding to the peak at rt = 29.11' (**B**).

The peak at rt = 29.11' in the reported chromatogram (**Figure 6.12 A**) corresponds to the desired ligation product ITC(Acm)GQVSSSLAPC(Acm) IPYVRGGGCVPPAC(Acm)C(Acm)NGIRNVNNL-$NHNH_2$ (calculated avg MW= 4032.3; found ions: 1345.4 m/z $[M+3H]^{3+}$, 1009.1 m/z $[M+4H]^{4+}$ (**Figure 6.12 B**)).

- Ligation [1-21]-Nbz + Cys-[23-37]

A



B

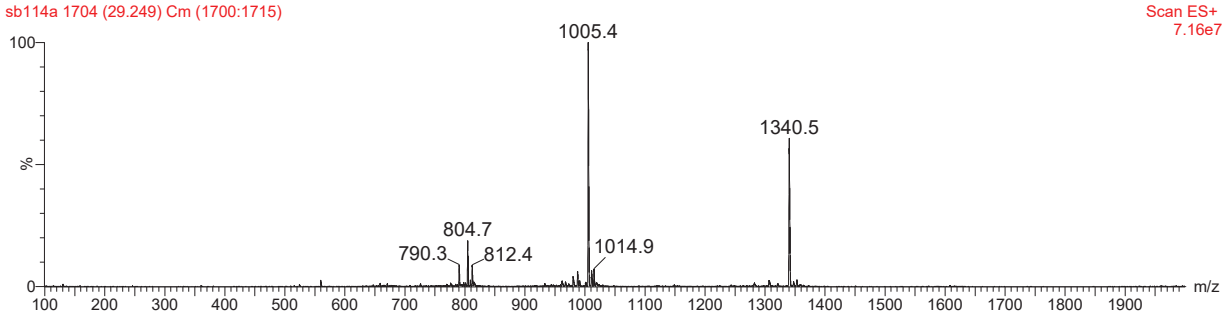
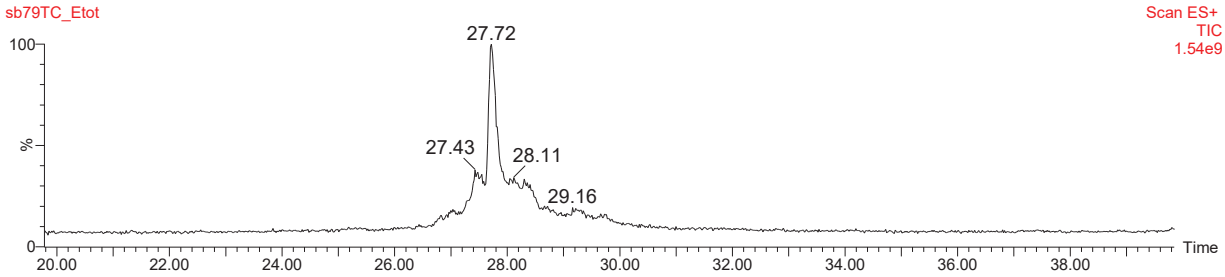


Figure 6.13 Chromatogram of the purified product obtained from the ligation [1-21]-Nbz + Cys-[23-37] (**A**); mass spectrum corresponding to the peak at rt = 29.25' (**B**).

The peak at rt = 29.25' in the reported chromatogram (**Figure 6.13 A**) corresponds to the desired ligation product ITC(Acm)GQVSSSLAPC(Acm) IPYVRGGGCVPPAC(Acm)C(Acm)NGIRNVNNL (calculated avg MW= 4018.3; found ions: 1340.5 m/z $[M+3H]^{3+}$, 1005.4 m/z $[M+4H]^{4+}$ (**Figure 6.13 B**)).

- Ligation Thz-[57-67]-Nbz + Cys-[69-91] and Thz deprotection

A



B

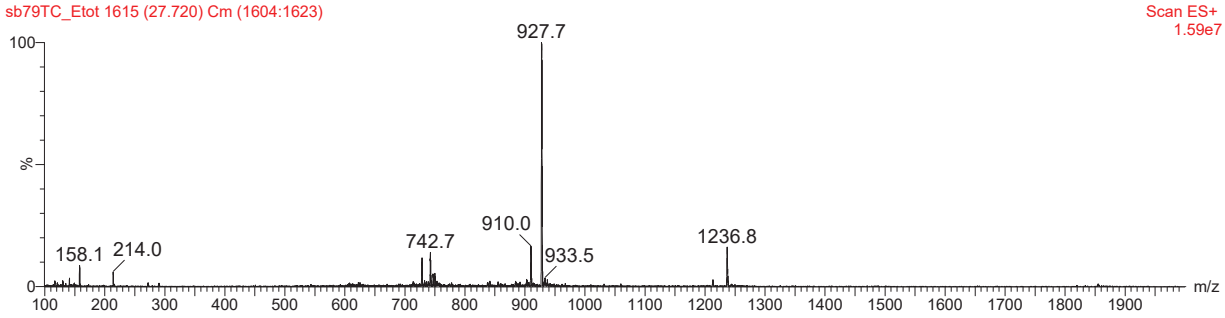
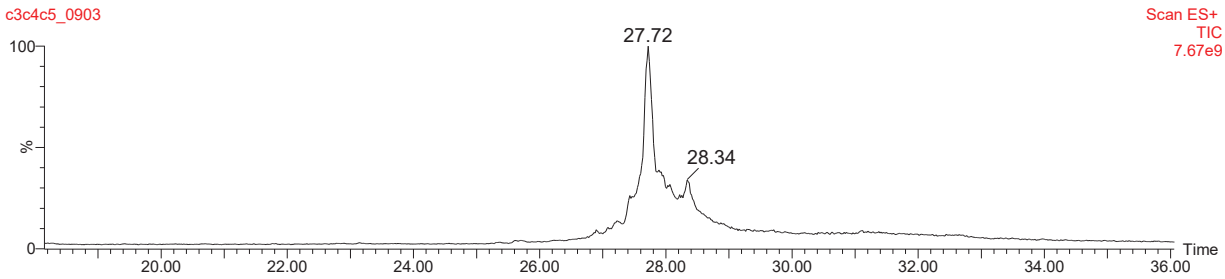


Figure 6.14 Chromatogram of the purified product obtained from the ligation Thz-[57-67]-Nbz + Cys-[69-91], followed by Thz deprotection (**A**); mass spectrum corresponding to the peak at rt = 27.72' (**B**).

The main peak in the reported chromatogram (**Figure 6.14 A**), at rt = 27.72', corresponds to the desired reaction product CSVPGVNPNNAACLPGK C(Acm)GVSIPYKISASTNC(Acm)ATVK (calculated avg MW= 3707.1; found ions: 1236.8 m/z $[M+3H]^{3+}$, 927.7 m/z $[M+4H]^{4+}$ (**Figure 6.14 B**)).

- Ligation Thz-[39-55] + Cys-[57-67]-Cys-[69-91] and Thz deprotection

A



B

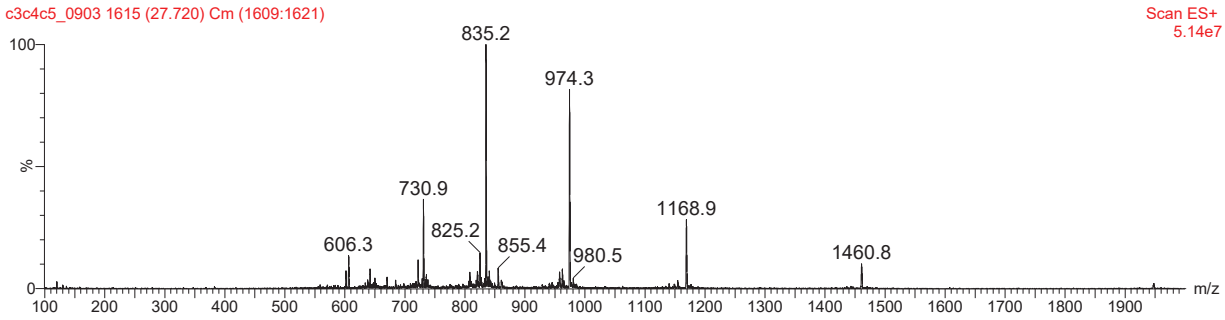


Figure 6.15 Chromatogram of the purified product obtained from the ligation Thz-[39-55] + Cys-[57-67]-Cys-[69-91], followed by Thz deprotection (**A**); mass spectrum corresponding to the peak at rt = 27.72' (**B**).

The main peak in the reported chromatogram (**Figure 6.15 A**), at rt = 27.72', corresponds to the desired reaction product CRTTPDRQAAC(Acm)NC(Acm) LKQLSCSVPGVNPNNAACLPGKC(Acm)GVSIPYKISASTNC(Acm)ATVK (calculated avg MW= 5839.4; found ions: 1460.8 m/z $[M+4H]^{4+}$, 1168.9 m/z $[M+5H]^{5+}$, 974.3 m/z $[M+6H]^{6+}$, 835.2 m/z $[M+7H]^{7+}$, 730.9 m/z $[M+8H]^{8+}$ (**Figure 6.15 B**)).

6.2.6 Peptide-hydrazide ligation

The last two parts of the protein were joined through a peptide-hydrazide ligation (**Figure 6.16**); in this process, the peptide-hydrazide is transformed in situ in a C-terminal thioester, which subsequently binds to the N-terminal Cys peptide.

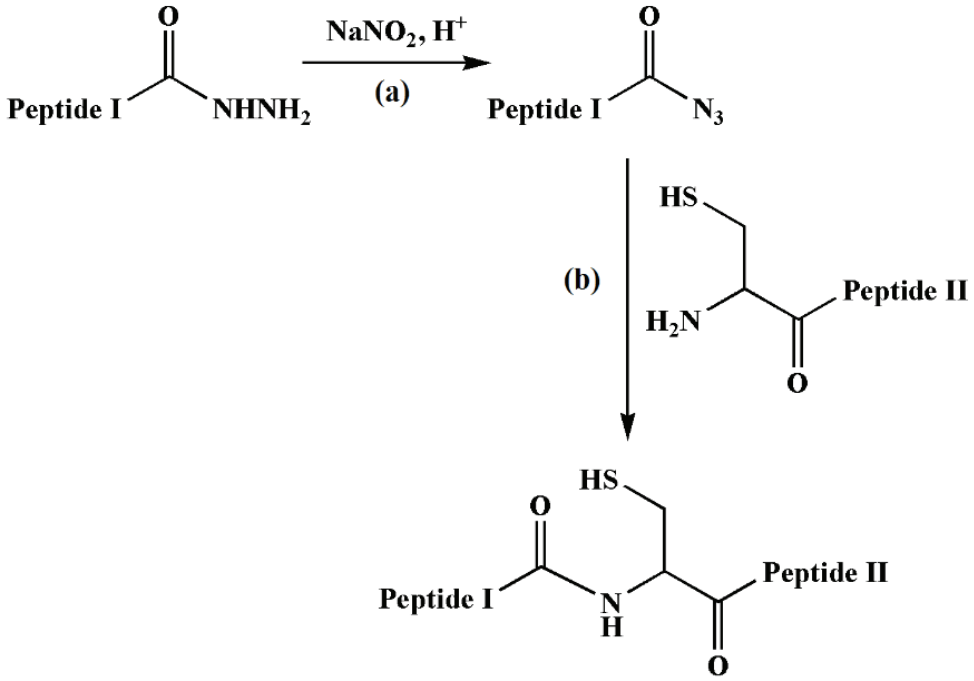
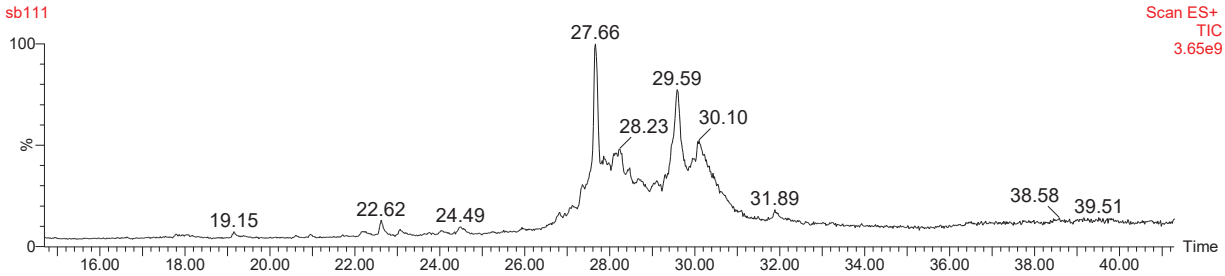


Figure 6.16 Peptide-hydrazide ligation; (**a**): Gn-HCl 6 M, NaH_2PO_4 0.2 M and $NaNO_2$ (10 eq.); pH = 3, -10°C; (**b**): + MPAA 0.2 M + TCEP-HCl 20 mM; pH = 7, RT.

The two peptides were quantified measuring the light absorbance at 280 nm of aqueous solutions containing the same, exploiting the presence of a Tyr residue in each.

The reaction mixture was analysed by means of UPLC-MS:

A



B

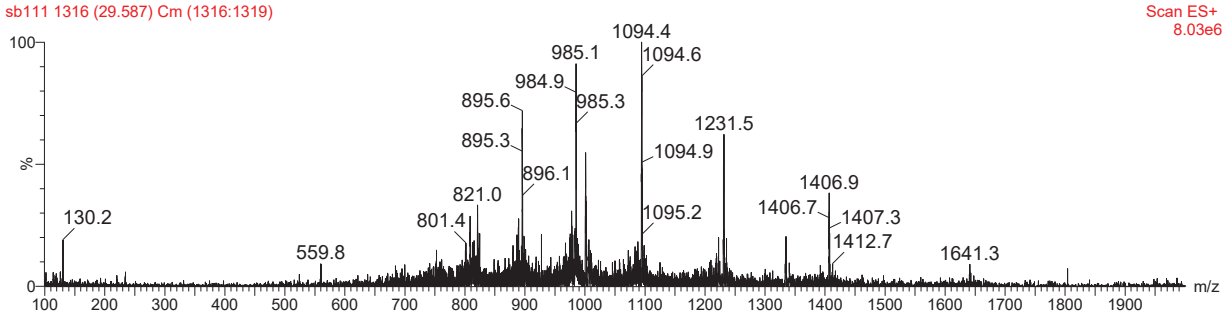


Figure 6.17 Chromatogram of the crude product obtained from the ligation [1-21]-Cys-[23-37]-$NHNH_2$ + Cys-[39-55]-Cys-[57-67]-Cys-[69-91] (**A**); mass spectrum corresponding to the peak at rt = 29.59' (**B**).

The peak at rt = 29.59' in the reported chromatogram (**Figure 6.17 A**), corresponds to the desired ligation product ITC(Acm)GQVSSSLAPC(Acm)IPYVRGGGCVPPAC(Acm)C(Acm)NGIRNVNNLCRTTPDRQAAC(Acm)NC(Acm)LKQLSCSVPGVNPNNAACLPGKC(Acm)GVSIPYKISASTNC(Acm)ATVK (calculated avg MW= 9839.8; found ions: 1641.3 m/z $[M+6H]^{6+}$, 1406.9 m/z $[M+7H]^{7+}$, 1231.1 m/z $[M+8H]^{8+}$, 1094.4 m/z $[M+9H]^{9+}$, 985.1 m/z $[M+10H]^{10+}$, 895.6 m/z $[M+11H]^{11+}$, 821.0 m/z $[M+12H]^{12+}$ (**Figure 6.17 B**)). The peak having a retention time of 27.66' corresponds to unreacted peptide Cys-[39-55]-Cys-[57-67]-Cys-[69-91]. Apparently, the quantitation of the reacting peptides needs to be further optimized for this ligation.

6.2.7 Cysteine desulfurization

The two final deprotection steps (i.e. cysteine desulfurization and Acm deprotection) were applied to the two protein halves foreseen by the chosen synthetic strategy, to obtain the peptides sPru p 3 (1-37) and sPru p 3 (38-91) in their native form.

The conversion of the Cys residues, used for the NCL reactions, to alanines was conveniently achieved through a metal-free, radical process (**Figure 6.18**). The latter can take place in neutral aqueous buffers in the presence of a radical initiator (e.g. AAPH) and a hydrogen donor such as reduced glutathione (GSH)[10].

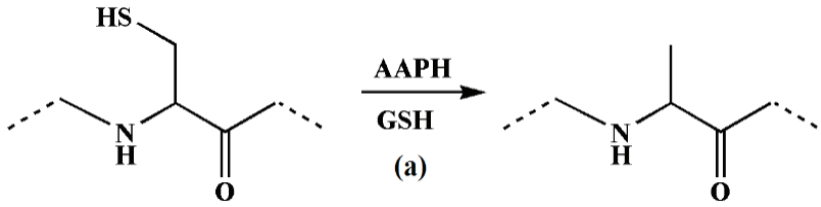
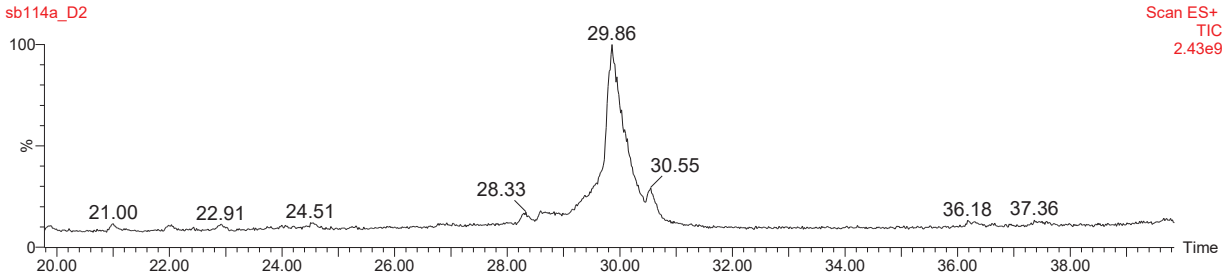


Figure 6.18 Radical Cysteine desulfurization; (**a**): Gn-HCl 6 M, NaH_2PO_4 0.2 M, TCEP-HCl 0.5 M, AAPH (40 eq.) and GSH (10 eq.); pH=6.5.

The reaction products were purified by ultrafiltration, as described in the experimental section and analysed through UPLC-MS:

- Desulfurization of the peptide [1-21]-Cys-[23-37]

A



B

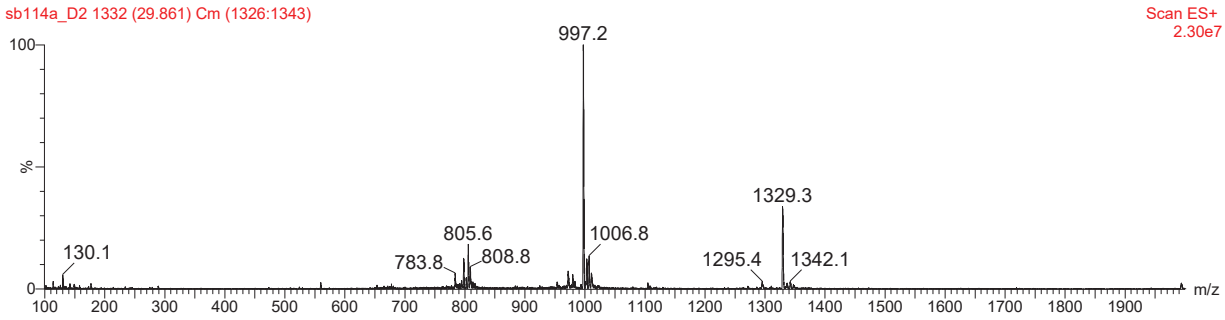
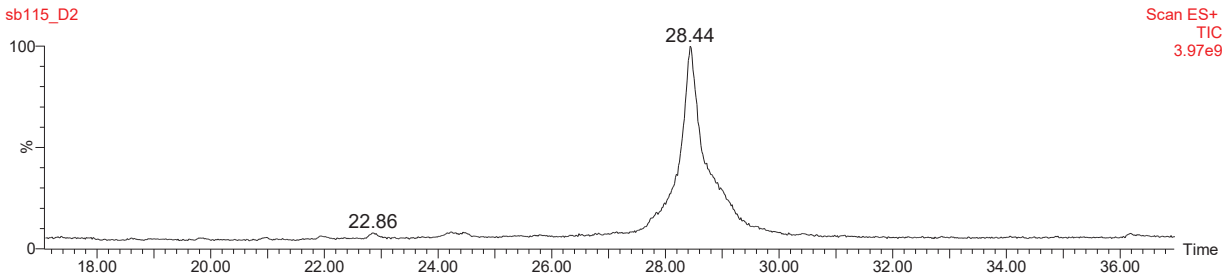


Figure 6.19 Chromatogram of the purified product obtained from the desulfurization of the peptide [1-21]-Cys-[23-37] (**A**); mass spectrum corresponding to the peak at rt = 29.86' (**B**).

The main peak in the reported chromatogram (**Figure 6.19 A**), at rt = 29.86', corresponds to the desired reaction product ITC(Acm)GQVSSSLAPC(Acm) IPYVRGGGAVPPAC(Acm)C(Acm)NGIRNVNNL (calculated avg MW= 3986.3; found ions: 1329.3 m/z $[M+3H]^{3+}$, 997.2 m/z $[M+4H]^{4+}$ (**Figure 6.19 B**)).

- Desulfurization of the peptide Cys-[39-55]-Cys-[57-67]-Cys-[69-91]

A



B

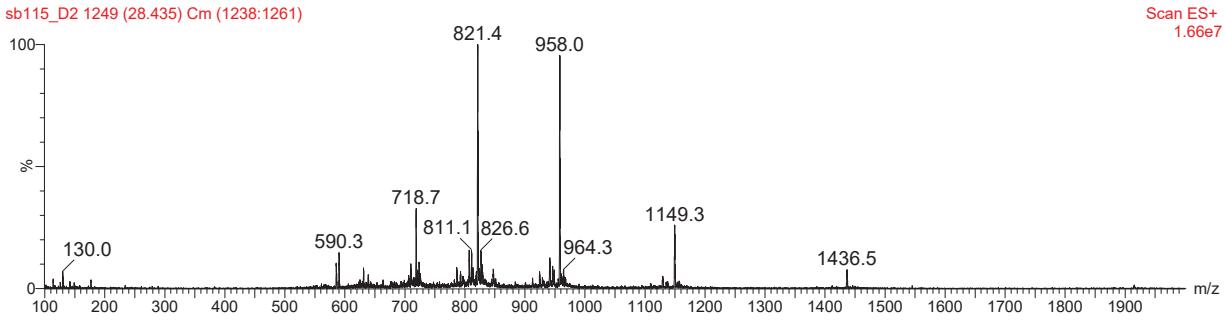


Figure 6.20 Chromatogram of the purified product obtained from the desulfurization of the peptide Cys-[39-55]-Cys-[57-67]-Cys-[69-91] (**A**); mass spectrum corresponding to the peak at rt = 28.44' (**B**).

The main peak in the reported chromatogram (**Figure 6.20 A**), at rt = 28.44', corresponds to the desired reaction product ARTTPDRQAAC(Acm)NC(Acm) LKQLSASVPGVNPNNAAALPGKC(Acm)GVSIPYKISASTNC(Acm)ATVK (calculated avg MW= 5743.3; found ions: 1436.5 m/z $[M+4H]^{4+}$, 1149.3 m/z $[M+5H]^{5+}$, 958.0 m/z $[M+6H]^{6+}$, 821.4 m/z $[M+7H]^{7+}$, 718.7 m/z $[M+8H]^{8+}$ (**Figure 6.20 B**)).

6.2.8 Acm deprotection

Native cysteines were deprotected from the acetaminomethyl group following a procedure previously reported in the literature[11] (**Figure 6.21**).

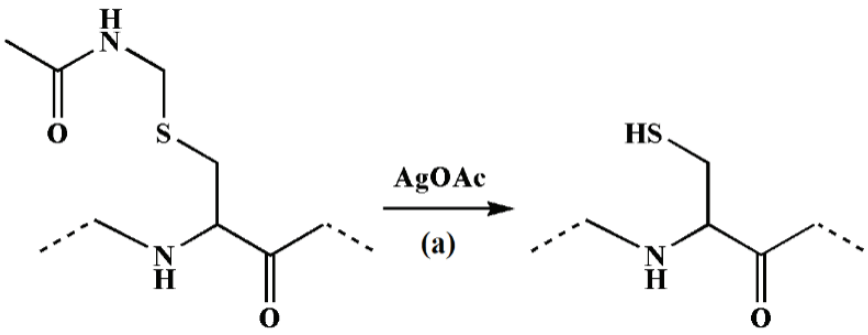
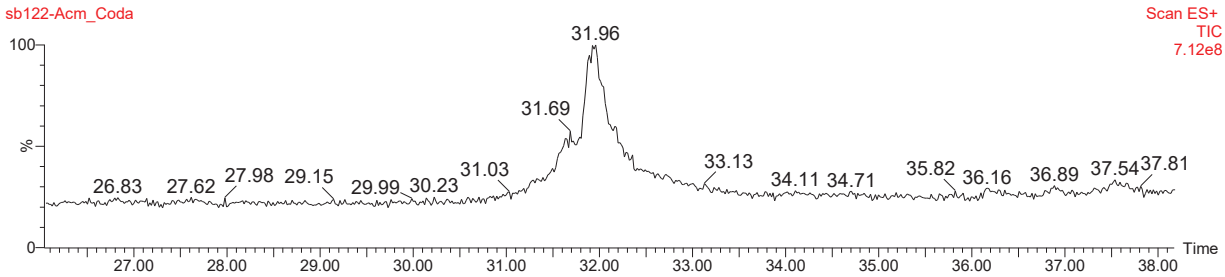


Figure 6.21 Acm deprotection; (**a**): AgOAc (35 eq.).

The reaction products were purified through ultrafiltration and characterized by means of UPLC-MS:

- Deprotection of the peptide [1-21]-Ala-[23-37]

A



B

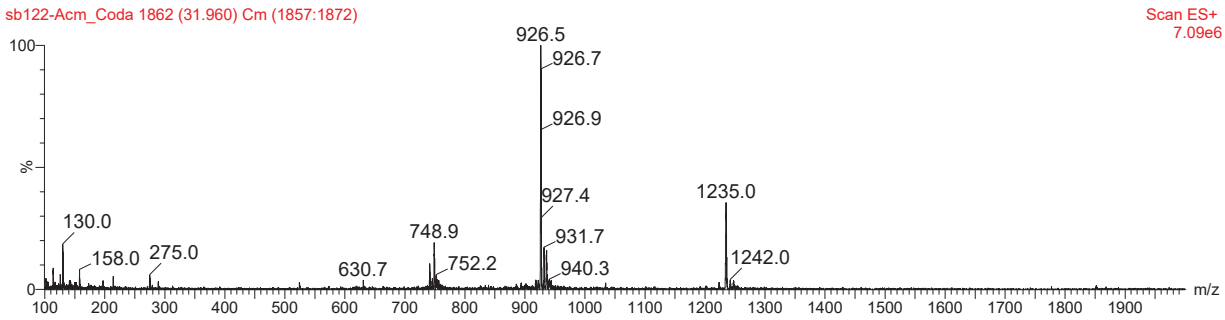


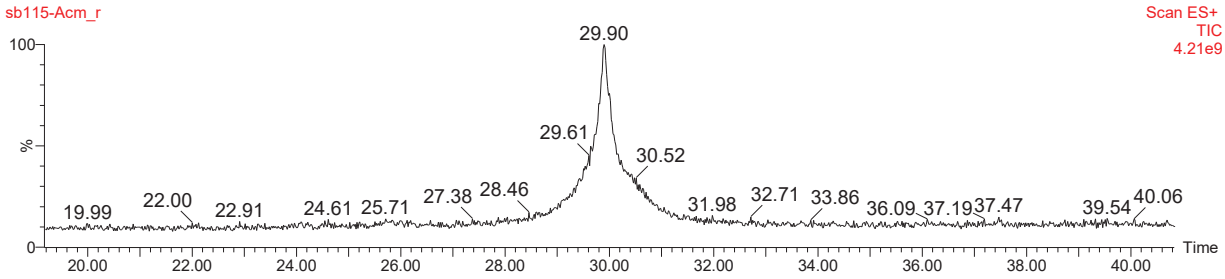
Figure 6.22 Chromatogram of the purified product obtained from the deprotection of the peptide [1-21]-Ala-[23-37] (**A**); mass spectrum corresponding to the peak at rt = 31.96' (**B**).

The main peak in the reported chromatogram (**Figure 6.22 A**), at rt = 31.96', corresponds to the desired reaction product ITCGQVSSSLAPCIPYVRGGG

AVPPACCNGIRNVNNL (sPru p 3 (1-37)) (calculated avg MW= 3702.3; found ions: 1235.0 m/z $[M+3H]^{3+}$, 926.5 m/z $[M+4H]^{4+}$ (**Figure 6.22 B**)).

- Deprotection of the peptide Ala-[39-55]-Ala-[57-67]-Ala-[69-91]

A



B

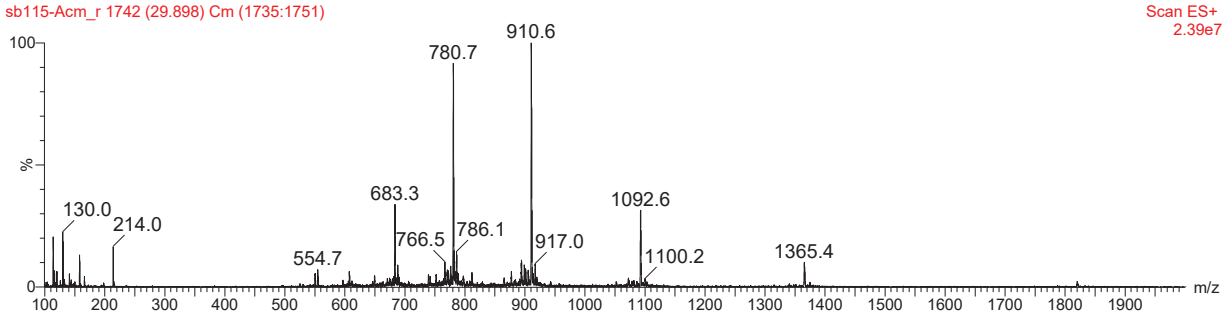


Figure 6.23 Chromatogram of the purified product obtained from the deprotection of the peptide Ala-[39-55]-Ala-[57-67]-Ala-[69-91] (**A**); mass spectrum corresponding to the peak at rt = 29.90' (**B**).

The main peak in the reported chromatogram (**Figure 6.23 A**), at rt = 29.90', corresponds to the desired reaction product ARTTPDRQAACNCLKQLSA SVPGVNPNNAAALPGKCGVSIPYKISASTNCATVK (sPru p 3 (38-91)) (calculated avg MW= 5459.3; found ions: 1365.4 m/z $[M+4H]^{4+}$, 1092.6 m/z $[M+5H]^{5+}$, 910.6 m/z $[M+6H]^{6+}$, 780.7 m/z $[M+7H]^{7+}$, 683.3 m/z $[M+8H]^{8+}$ (**Figure 6.23 B**)).

6.2.9 Immunological tests

6.2.9.1 Dot blot

The ability of LTP specific antibodies to bind the peptides sPru p 3 (1-37) and sPru p 3 (38-91), was pointed out through dot blot analyses; a recombinant and a natural Pru p 3 standard were used as controls.
The results obtained after incubation with sera of two rabbits immunized with the apple LTP Mal d 3[12] are shown in **Figure 6.24**.

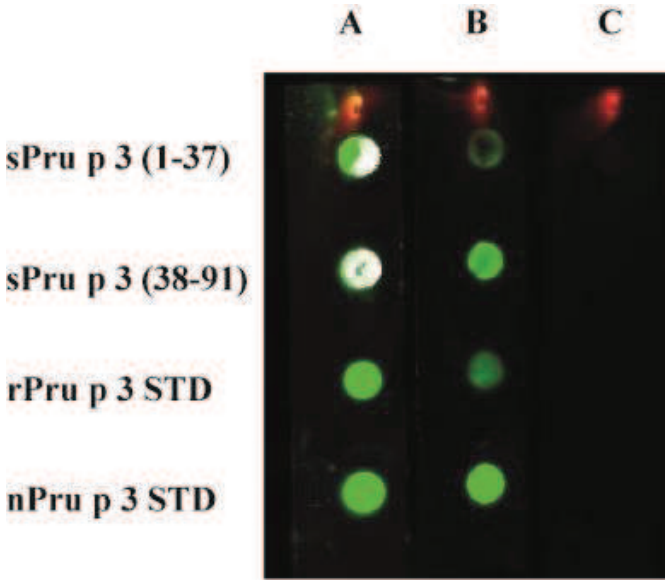


Figure 6.24 Dot blot results produced by incubation with LTP specific polyclonal IgG rabbit antibodies; **A**: rabbit-anti LTP IgG 125487; **B**: rabbit-anti LTP IgG 126/41; **C**: label control.

The antibodies contained in the tested sera show a very good binding ability to both the synthetic peptides; no nonspecific binding of the secondary antibodies to the analytes can be detected (**Figure 6.24 C**).

Two different Pru p 3 specific monoclonal mouse antibodies were obtained from Gao and co-workers[Unpublished data] and were tested for binding with the synthetic peptides; results are shown in **Figure 6.25**.

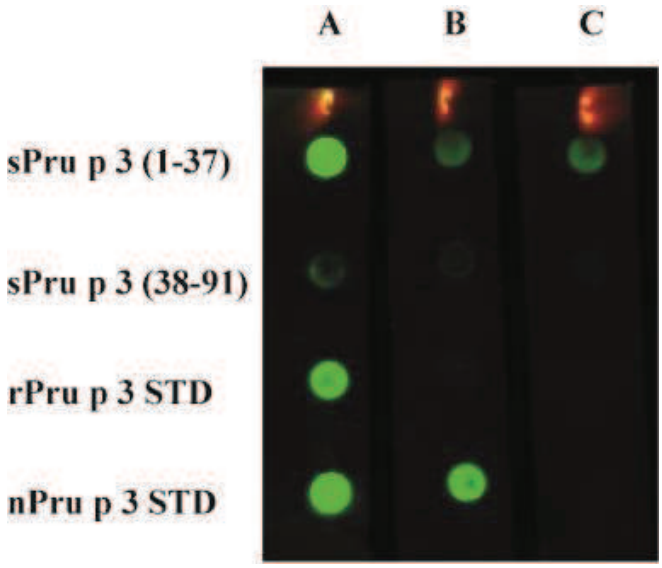


Figure 6.25 Dot blot results produced by incubation with Pru p 3 specific monoclonal IgG mouse antibodies; **A**: mAb A7-1 (IgG1 isotype); **B**: mAb 4-1 (IgG2a isotype); **C**: label control.

The secondary antibodies give a certain degree of nonspecific binding to the peptide sPru p 3 (1-37) (**Figure 6.25 C**), however it can be stated that a positive result is obtained for the binding of mAb A7-1 to the same peptide, as the intensity of the detected dot is much higher than in the control. No relevant binding of the tested antibodies with the peptide sPru p 3 (38-91) can be observed.

Sera of 18 peach allergic patients (**Table 6.1**) were used to assess the binding ability of the IgE antibodies contained in them to the synthetic peptides (**Figure 6.26**).

Table 6.1 Sera of peach allergic patients used for the dot blot analyses.

Serum	Origin	Peach specific IgE (kUA/l)
A	Spain	11.8
B	Spain	20.1
C	Spain	7.3
D	Spain	11.7
E	Spain	15.4
F	Spain	25.2
G	Spain	9.4
H	Spain	14.6
I	Netherlands	10.4
J	Netherlands	31.0
K	Spain	17.4
L	Spain	26.6
M	Spain	20.1
N	Spain	21.2
O	Italy	41.1
P	Italy	160.0
Q	Italy	28.4
R	Netherlands	11.1

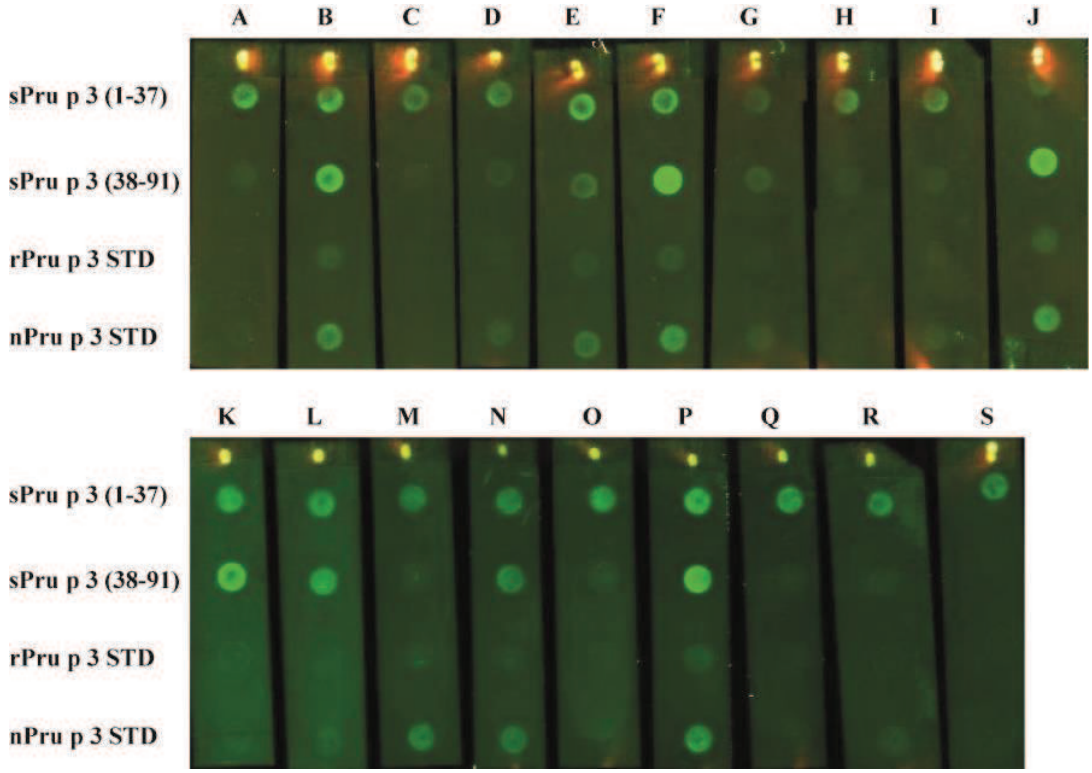


Figure 6.26 Dot blot results produced by incubation with sera of peach allergic patients; detection with anti-Human IgE. **A-R**: sera (**Table 6.1**); **S**: label control.

The IgE antibodies contained in 7 of the 18 tested sera display a good binding ability to the peptide sPru p 3 (38-91) (**Figure 6.26 B**, **F**, **J**, **K**, **L**, **N**, **P**). The apparently bigger binding ability of these antibodies to the synthetic peptide than to the natural Pru p 3 standard can partially be ascribed to the different molar amounts of the two species on the blot, as each dot contains the same mass of analyte. Another explanation for the higher intensity of the signals relative to the interaction between the peptide sPru p 3 (38-91) and the mentioned antibodies, could be the fact that the latter are specific for epitopes that are more exposed on the blot in the case of the synthetic peptide.
Unfortunately, the secondary antibodies used for detection give high nonspecific binding to the peptide sPru p 3 (1-37) (**Figure 6.26 S**), so the binding ability of the IgE antibodies contained in the tested sera to this peptide cannot be established by dot blot.
For none of the tested sera, the used recombinant Pru p 3 standard proved to be an efficient tool for the detection of specific IgE antibodies against this protein.

6.2.9.2 CAP-inhibition assays

The sera **F** and **P**, for which the response was particularly high on the dot blot assay with the peptide sPru p 3 (38-91) (**Figure 6.26**), were used to perform CAP-inhibition tests employing commercial CAPs coated with recombinant Pru p 3. This kind of test allows to effectively compare the ability of different antigens in their native conformation (i.e. in solution) to interact with specific IgE antibodies. In particular, we tested the ability of the peptides sPru p 3 (1-37), sPru p 3 (38-91) and a mixture of them to inhibit the binding between the antibodies contained in the used sera and the recombinant Pru p 3 present in the CAPs. The tests were also performed with a natural Pru p 3 standard, to obtain a reference value in the used experimental conditions.
Percentages of inhibition were calculated comparing the produced data with those achieved running the analyses without inhibition. The obtained results are summarized in **Table 6.2**.

Table 6.2 Results of the CAP-inhibition assay.

Added antigen		**Serum**	
		F (1:6 in PBS)	**P** (1:20 in PBS)
None (uninhibited)	**IgE conc. (kUA/l)**	4.13	7.95
sPru p 3 (1-37)	**IgE conc. (kUA/l)**	3.72	7.70
	Inhibition	9.9 %	3.1 %
sPru p 3 (38-91)	**IgE conc. (kUA/l)**	3.70	7.19
	Inhibition	10.4 %	9.6 %
sPru p 3 (1-37) + sPru p 3 (38-91)	**IgE conc. (kUA/l)**	2.64	6.00
	Inhibition	36.1 %	24.5 %
nPru p 3	**IgE conc. (kUA/l)**	0.12	0.87
	Inhibition	97.1 %	98.9 %

As reported in **Table 6.2**, only little inhibition is obtained when the sera are pre-incubated individually with the peptides, while a more relevant inhibition can be observed after treatment of the sera with a mixture of the two peptides. As

the inhibition percentages obtained with the mixture are bigger than the sum of the same values calculated for the single peptides, the two protein halves are apparently able to interact with a single IgE antibody in a concerted manner.

6.2.10 Circular Dichroism (CD) Spectroscopy

The peptides sPru p 3 (1-37) and sPru p 3 (38-91) were analysed by CD spectroscopy to verify if they assume a well-defined secondary structure in solution.
Tests were performed in H_2O and in the presence of the anionic detergent sodium dodecyl sulfate (SDS; 2 mM), to mimic a more hydrophobic environment, as the one which can be found in cells.
The obtained spectra are reported in **Figure 6.27**.

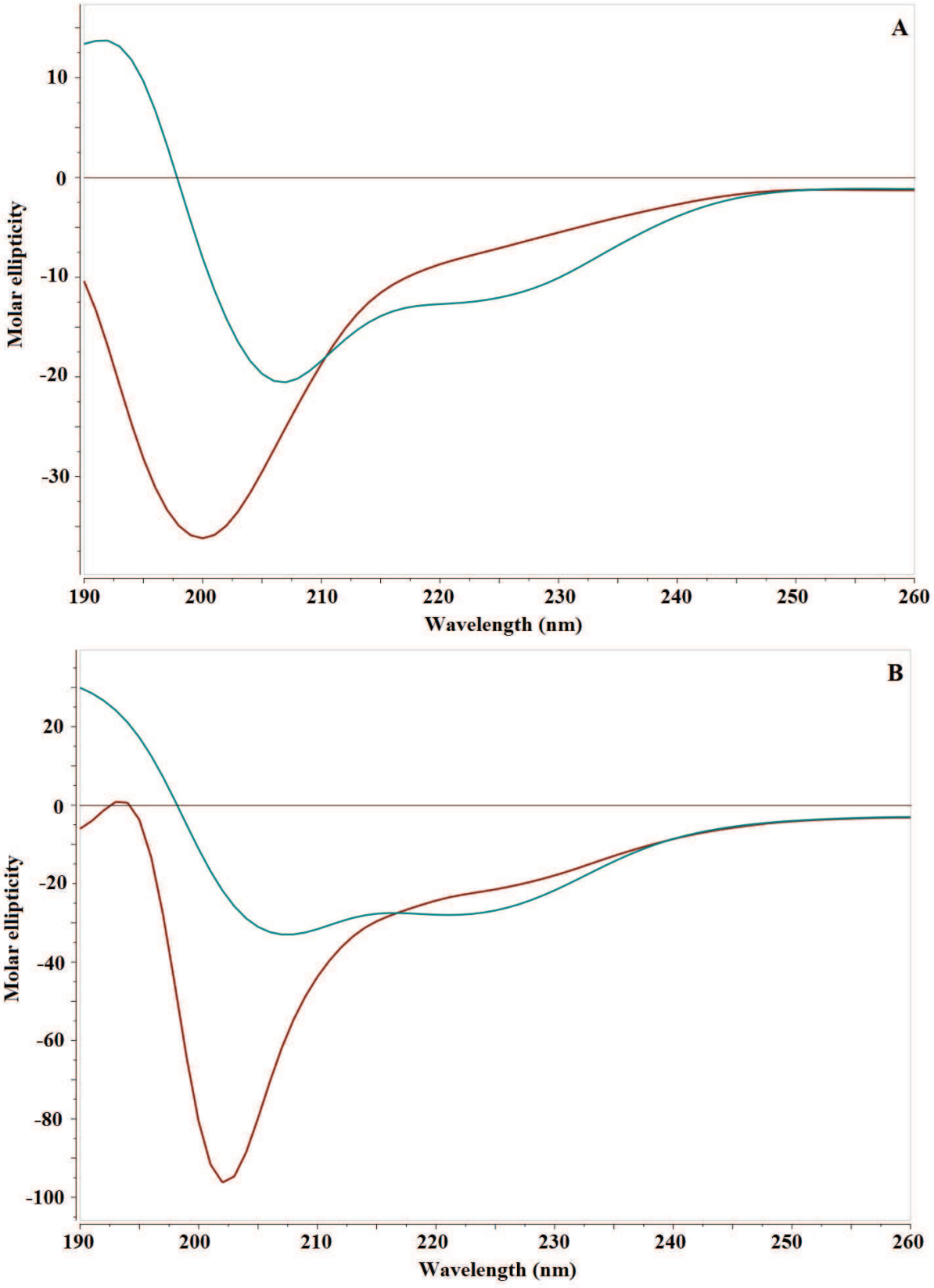


Figure 6.27 CD spectra of the peptides sPru p 3 (1-37) (**A**) and sPru p 3 (38-91) (**B**) in H_2O (red line) and in SDS 2 mM (blue line).

The software K2D3 was used for the deconvolution of the CD spectra, to assess the secondary structure contents of the analysed peptides (**Table 6.3**).

Table 6.3 Secondary structures found in the peptides sPru p 3 (1-37) and sPru p 3 (38-91).

Peptide	Environment	% α-helix	% β-sheet
sPru p 3 (1-37)	H_2O	8.2	8
	SDS 2 mM	77.5	-
sPru p 3 (38-91)	H_2O	72.5	-
	SDS 2 mM	94.8	-

According to the obtained data, the peptide sPru p 3 (1-37) has a rather unordered conformation in H_2O, while the structure of the peptide sPru p 3 (38-91) has a high α-helix content in the same conditions. Both peptides assume a conformation with a very high α-helix content, as characteristic for LTPs, in the presence of SDS.

6.3 Conclusions

A strategy for the total chemical synthesis of the allergenic protein Pru p 3 has been developed and tested. The whole protein sequence was ideally divided, at the level of four alanine residues, into five peptide fragments, to be produced by Fmoc-SPPS and to be bound through native chemical ligation (NCL) and peptide-hydrazide ligation, followed by desulfurization. All the experimental conditions for the needed processes were set up.

The five starting fragments were bound together to obtain the entire protein stretch. The synthesis has to be repeated on a bigger scale and the product of the peptide-hydrazide ligation needs to be purified, to perform the final desulfurization and deprotection steps on it and obtain the protein in its native sequence.

The two final protein halves foreseen by the developed synthetic strategy were desulfurized and deprotected to produce the peptides sPru p 3 (1-37) and sPru p

3 (38-91). The immunogenicity of these fragments was confirmed through tests with antibodies raised in animals against LTPs or contained in sera of patients allergic to peach. Moreover, the secondary structures of these peptides were investigated through CD spectroscopy and it was found that, in conditions mimicking the degree of hydrophobicity that can be found in cells, they both have conformations with the high α-helix content typical for LTPs.
Future work will be focused on the production of the entire protein in its native form; once obtained, the product will be compared with natural Pru p 3 from the structural and immunological points of view.

6.4 Experimental part

6.4.1 Chemicals

2,2'-Azobis(2-amidinopropane) dihydrochloride (AAPH), 4-Mercaptophenylacetic acid (MPAA), 4-Nitrophenyl chloroformate, 5,5′-Dithiobis(2-nitrobenzoic acid) (DTNB, Ellman's Reagent), Acetonitrile (ACN), Dichloromethane (DCM), Di-isopropylethylamine (DIPEA), Dithiotreitol (DTT), Ethylenediaminetetraacetic acid (EDTA), Ethyl ether, Formic acid (FA), Hydrazine monohydrate (NH_2NH_2-H_2O), L-Glutathione reduced (GSH), Lithium bromide (LiBr), Methanol (CH_3OH), Methoxyamine hydrochloride ($MeONH_2$-HCl), N,N-Diisopropylethylamine (DIPEA), N,N-Dimethylformamide (DMF), Piperidine, Silver acetate (AgOAc), Sodium dodecyl sulfate (SDS), Sodium phosphate monobasic (NaH_2PO_4), Sodium nitrite ($NaNO_2$), Tetrahydrofuran (THF), Trifluoroacetic acid (TFA), Triisopropylsilane (TIS), Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) and Tris(hydroxymethyl)aminomethane hydrochloride (TRIS-HCl) were purchased from Sigma Aldrich (Sigma, St. Louis, MO, USA).
Boc-protected amino acids (Boc-Cys(Trt)-OH, Boc-Ile-OH-0.5H_2O, Boc-Thz-OH), Fmoc-protected amino acids (Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OBut)-OH, Fmoc-Cys(Acm)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gly-OH, Fmoc-Gln(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Pro-OH, Fmoc-Ser(^{t}But)-OH, Fmoc-Thr(^{t}But)-OH, Fmoc-Tyr(^{t}But)-OH, Fmoc-Val-OH), Guanidine hydrochloride (Gn-HCl), N,N,N′,N′-

Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU), resins for peptide synthesis (Dawson Dbz AM resin (100-200 mesh), Fmoc-Lys(Boc) NovaSyn TGA resin, Fmoc-Leu Wang resin (100-200 mesh)) and Tween 20 were purchased from Merck-Millipore (Merck KGaA, Darmstadt, Germany).
Blotting-Grade Blocker was purchased from Bio-Rad (Hercules, CA, USA).
Buffered sodium chloride solution, pH=7.4 (PBS) was purchased from Fresenius Kabi GmbH (Graz, Austria).
NuPage Transfer buffer was purchased from Novex (Thermo Fisher Scientific, Waltham, MA, USA).
Pierce^{TM} BCA Protein Assay Kit was purchased from Thermo Scientific (Thermo Scientific-Pierce Biotechnology, Rockford, IL, USA).
Secondary IgG antibodies labelled with IRDye® 800 CW (IRDye® 800CW Goat anti-Rabbit IgG (H + L), 0.5 mg; IRDye® 800CW Donkey anti-Mouse IgG (H + L), 0.5 mg) were purchased from LI-COR Biotechnology (Lincoln, NE, USA), while IRDye® 800 CW Conjugated monoclonal Mouse anti-Human IgE were obtained from Rockland (Gilbertsville, PA, USA).

6.4.2 Procedure

6.4.2.1 Attachment of the first amino acid to the Dawson Dbz resin

Swelling of the resin was achieved by suspension in DCM (1 h); the resin was washed with DMF, then the Fmoc protecting group was removed by treatment with 20% piperidine in DMF (2 x 10'). The first amino acid was loaded onto the resin by the addition of 6 equivalent of amino acid, 6 equivalents of HBTU and 9 equivalents of DIPEA, to the initial loading of the resin (0.49 mmol/g), in DMF (1 h with slight shaking). The resin was washed with DMF and DCM and was finally dried under vacuum.

6.4.2.2 Peptide synthesis

Peptides were synthesized using a Syro I automated synthesizer (Biotage, Uppsala, Sweden), according to the Fmoc-SPPS protocol. Prior to the synthesis

the resin was allowed to swell in DCM (3 washes of 10'). Fmoc-deprotection was achieved by treatment with piperidine 40% in DMF (5') followed by a second wash with 20% piperidine in DMF (7'); amino acidic couplings were performed in DMF using 4 equivalents of amino acid and HBTU and 8 equivalents of DIPEA, to the initial loading of the resin (40'). Extensive washes with DMF were performed after each deprotection and coupling step. Following the peptide-chain assembly, the resin was washed with DCM (3 x 10') and dried under vacuum.

6.4.2.2.1 Synthesis of the peptide Cys-[69-91]

The peptide Cys-[69-91] was synthesized with the procedure described above using a Fmoc-Lys(Boc) NovaSyn TGA resin; amino acid solutions with concentrations of 0.5 M in DMF were used. A double coupling step was performed for Ile[77], while triple couplings were used for the amino acids Leu[69] and Cys[68]. Cysteines in positions 73 and 87 were introduced as Fmoc-Cys(Acm)-OH.

6.4.2.2.2 Synthesis of the peptide Cys-[23-37]

The peptide Cys-[23-37] was synthesized with the procedure described above using a Fmoc-Leu-Wang resin; amino acid solutions with concentrations of 0.5 M in DMF were used. Cysteines in positions 27 and 28 were introduced as Fmoc-Cys(Acm)-OH.

6.4.2.2.3 Peptide-Nbz synthesis

Peptides [1-21]-Nbz, Cys-[23-37]-Nbz, Thz-[39-55]-Nbz and Thz-[57-67]-Nbz were synthesised as follows:
Peptide chain assembly on preloaded Dawson Dbz resin was carried out with the procedure outlined above; amino acid solutions with concentrations of 0.15 M in DMF were used and the N-terminal amino acid was introduced as Boc-protected. Cysteines in positions 3, 13, 27, 28, 48 and 50 were introduced as Fmoc-Cys(Acm)-OH.

The obtained Boc-peptide-Dbz-resin was allowed to swell in DCM (1 h) and was then treated with 5 equivalents of 4-Nitrophenyl chloroformate in LiBr 2 M in THF (1 h). Treatment with 0.5 M DIPEA in DMF (30') produced Boc-peptide-Nbz-resin that was washed with DCM and dried under vacuum.

6.4.2.3 Cleavage of the peptides from the resin

The dry resin was suspended in a cleavage cocktail made of TFA (95%), H_2O (2.5%), TIS (2.5%), in the case of peptides without Cys(Trt) and TFA (94%), DTT (2.5%), H_2O (2.5%), TIS (1%) for peptides containing Cys(Trt). After 3 h the acidic solution containing the cleaved peptide was recovered, then the resin was washed twice with TFA and the washes were pooled together with the first solution.
The volume of the liquid was reduced under N_2 flux, then ethyl ether was added; after 2 h at -18ºC the crude peptide was recovered by centrifugation, washed twice with ethyl ether and finally dried under vacuum.
A small amount of each product was dissolved in H_2O and analysed by UPLC-MS.

6.4.2.4 UPLC-MS analysis of the synthesis products

The chromatographic separation was performed using an Acquity WATERS UPLC system (Waters,Milford, MA, USA) equipped with a C18 column (150 x 2.1, 130 Å) (Waters) kept at 35 °C; 2 µL of sample were injected and the flow rate was set at a 0.2 mL/min.
Eluents:

- A: H_2O + 0.1% FA + 0.2% ACN
- B: ACN + 0.1% FA

Gradient: 7' isocratic elution with 100% A followed by a 43' linear gradient from 0% to 50% of B.

ESI-MS spectrometer (WATERS-ACQUITY-Ultraperformance- Waters) conditions were the following: positive ion mode, capillary voltage 1.93 kV, cone voltage 150 V, source temperature 100 °C, desolvation temperature 200°C, spraying gas (N_2) 100 L/h, desolvation gas (N_2) 650 L/h, full scan acquisition from 100 to 2000 m/z in continuum mode and 1 sec of scan time.

6.4.2.5 Peptide Cys-[23-37]-Nbz to Cys-[23-37]-hydrazide conversion

The crude peptide-Nbz was dissolved in 5 ml of an aqueous buffer made of: Gn-HCl 6 M, NaH_2PO_4 0.2 M, DTT 20 mM and NH_2NH_2-H_2O 0.2 M; pH=7.
After 3 h at room temperature with magnetic stirring, the mixture was desalted using C18 sep pak cartridges (Waters) and analysed through UPLC-MS, as outlined above. The solution was then dried under N_2 flux.

6.4.2.6 Native Chemical Ligation (NCL)

[1-21]-Nbz + Cys-[23-37]-$NHNH_2$, [1-21]-Nbz + Cys-[23-37], Thz-[57-67]-Nbz + Cys-[69-91] and Thz-[39-55] + Cys-[57-67]-Cys-[69-91] ligations were performed according to the following protocol:
Peptide quantification: Peptides bearing an N-terminal cysteine were quantified using Ellman's Reagent: the crude peptide was dissolved in Gn-HCl 6 M and diluted 1:100 in Riddle's buffer (Gn-HCl 6 M, TRIS-HCl 50 mM, EDTA 1 mM; pH=8.3) to obtain 3 ml; to the produced solution 30 µl of DTNB buffer (TRIS-HCl 0.1 M, DTNB 10 mM; pH=7.6) were added.
A blank sample was prepared by mixing 3 ml of Riddle's buffer with 30 µl of DTNB buffer.
Light absorption at 412 nm was measured using a Jasco V-530 UV-Vis Spectrophotometer (Jasco Inc, Easton, USA); the concentration of the peptide in the starting solution was calculated considering a value of ε_{412}= 13600 $cm^{-1}M^{-1}$ for the extinction coefficient of the reagent.
C-terminal Nbz peptides were quantified by weight.

Ligation: NaH_2PO_4, TCEP-HCl and MPAA were added to the Gn-HCl 6 M solution containing the N-terminal Cys peptide, to obtain final concentrations of 0.2 M, 20 mM and 0.2 M respectively. The pH value was adjusted to 7, then 1,2 equivalents of the peptide-Nbz were added.
The reaction was allowed to go on at room temperature with magnetic stirring for 5 h, then the solution was analysed by means of UPLC-MS as outlined above; the mass spectrum was acquired starting from the 7^{th} minute to prevent the salts from entering into the ion source.

6.4.2.7 Thz deprotection

Thz-[57-67]-Cys-[69-91] to Cys-[57-67]-Cys-[69-91] and Thz-[39-55]-Cys-[57-67]-Cys-[69-91] to Cys-[39-55]-Cys-[57-67]-Cys-[69-91] conversions were obtained according to the following protocol:
Once the NCL was complete, an equal volume of an aqueous solution containing $MeONH_2$-HCl 0.4 M was added to the reaction mixture and the pH value was adjusted to 4. After 3 h, the solution was desalted by ultrafiltration according to the following procedure:
Filter pre-rinsing: the ultra-filtration membranes in Amicon Ultra devices (Millipore, Bedford, MA, USA), having a MW cut-off of 3 kDa, were pre-rinsed 5 times: each wash was performed with a CH_3OH:H_2O (1:1) solution, at 5000 rpm and at room temperature for 15'. Devices were stocked in a CH_3OH:H_2O (5:95) solution at 4 °C until use.
Sample desalting: the reaction mixture was loaded onto the pre-rinsed filter devices and centrifuged at 7000 rpm and at 4°C until the depletion of the whole solution. Two washes with Gn-HCl 3 M in H_2O of the retentates were performed using the same centrifugation settings. The retentate was finally recovered using H_2O.
An UPLC-MS analysis of the obtained solution was performed as previously described.

6.4.2.8 Reaction products purification by RP-HPLC

The ligation products were purified by means of semipreparative-HPLC (Waters 1525 Binary HPLC Pump equipped with a 998 detector (Waters)) using a Jupiter C18 column (250 x 10 mm, 300 Å) (Phenomenex, Torrance, CA, USA); flow rate was set at 4 ml/min. The UV absorption spectrum of the eluate was registered at 214 nm.
Eluents:

- A: H_2O + 0.1% TFA
- B: ACN + 0.1% TFA

Gradients were optimized for each reaction product. Manually collected fractions were dried at reduced pressure (rotavapor) and analysed by means of UPLC-MS as outlined above, to check the purity of the peptides.

6.4.2.9 Peptide-hydrazide ligation

Peptide [1-21]-Cys-[23-37]-$NHNH_2$ was bound to peptide Cys-[39-55]-Cys-[57-67]-Cys-[69-91] according to the following procedure:
Peptide quantification: As both peptides contained a Tyr residue, UV-absorption at 280 nm was used for quantification; the peptides were dissolved in aqueous solutions containing Gn-HCl 6 M and NaH_2PO_4 20 mM, pH=6.5.
Light absorbance at 280 nm was measured and the concentration of the peptides in the solutions was calculated considering a value of ε_{280}= 1280 $cm^{-1}M^{-1}$ for the extinction coefficient of the Tyr residue.
Ligation: Volumes of the solutions containing the same molar amounts of the two peptides were pooled together; NaH_2PO_4 was added to a final concentration of 0.2 M and the pH value was lowered to 3.
The solution was cooled to -10 ºC in an ice-salt bath, then 10 equivalents of $NaNO_2$ dissolved in the minimum quantity of H_2O were added and the reaction was allowed to go on at -10 ºC with magnetic stirring for 30'.
100 equivalents of MPAA were added to the reaction mixture and the pH value was adjusted to 7; after 2 h at room temperature with magnetic stirring, TCEP-HCl was added to a final concentration of 20 mM and the pH was readjusted to 7.

The ligation was allowed to go on for further 3 h, then an UPLC-MS analysis of the reaction mixture was performed using an Acquity UPLC BEH 300, C8 column (150 x 2.1, 130 Å) (Waters); all the other parameters were kept as those previously described for the peptide characterization. The mass spectrum was acquired starting from the 7th minute of chromatographic run.

6.4.2.10 Cysteine desulfurization

The ligation products [1-21]-Cys-[23-37] and Cys-[39-55]-Cys-[57-67]-Cys-[69-91] were desulfurized according to the following procedure:
The purified peptide was dissolved in an aqueous buffer containing Gn-HCl 6 M and NaH_2PO_4 20 mM at pH=6.5 and quantified measuring the UV-absorption at 280 nm, as described above.
NaH_2PO_4 and TCEP-HCl were added to a final concentration of 0.2 M and 0.5 M respectively and the pH was readjusted to 6.5. 40 equivalents of AAPH and 10 equivalents of GSH were added and the mixture was stirred at room temperature for 24 h. An equal volume of TFA 10% in H_2O was added to the system to quench the reaction.
Desalting was achieved by ultrafiltration using filters with a 3 kDa MW cutoff: the solution was loaded onto pre-rinsed filter devices and centrifuged at 7000 rpm, 4°C until the depletion of the whole solution. A wash with Gn-HCl 1 M in H_2O of the retentates, followed by two washes with H_2O, were performed using the same centrifugation settings. The retentate was finally recovered using H_2O.
An UPLC-MS analysis of the obtained solution was performed as previously described.

6.4.2.11 Acm deprotection

Acm deprotection of the peptides [1-21]-Ala-[23-37] and Ala-[39-55]-Ala-[57-67]-Ala-[69-91] was achieved through the following procedure:
The peptide was dissolved in H_2O and 35 equivalents, relative to the quantity of cysteines (determined using Ellman's reagent, as previously described), of AgOAc were added. The mixture was left at room temperature with magnetic stirring for 5 h, then the reaction was quenched by adding two volumes of an

aqueous solution containing Gn-HCl 6 M and DTT 0.2 M. After mixing for 10', the suspension was centrifuged and the clear supernatant was recovered.
The obtained solution was diluted with an equal volume of H_2O and desalted by ultrafiltration, as previously described for the desulfurization products.
An UPLC-MS analysis of the obtained solution was performed as previously described.

6.4.2.12 Immunological tests

The peptides sPru p 3 (1-37) and sPru p 3 (38-91) were tested for immunogenicity; a natural and a recombinant Pru p 3 standard were used as controls.
Peptide and standard quantification: the concentration of aqueous solutions containing the peptides and the standards was determined using the Pierce™ BCA Protein Assay Kit, following the instruction of the producer.
Dot blot: a nitrocellulose membrane (Nitrocellulose membrane filter paper sandwich, 0.2 μm pores (Novex-Thermo Fisher Scientific)) was wetted (10') in a solution made of 1 part of NuPage transfer buffer, 2 parts of CH_3OH and 7 parts of H_2O and placed into a dot blot press.
The solutions containing the synthetic peptides and the two Pru p 3 standards were spotted on the membrane so that each dot contained 0.5 μg of peptide or protein. The membrane was allowed to air dry for 1 h and then it was incubated for 1 h in a blocking buffer made of PBS containing 5% (w/v) Blotting-Grade Blocker.
The membrane was washed for 3 times with PBS containing 0.1% (v/v) Tween 20 and then it was cut in strips containing one dot of each peptide and protein standard.
The binding to polyclonal IgG antibodies raised in rabbits immunized with Mal d 3[12], monoclonal IgG antibodies obtained from the sera of mice immunized with Pru p 3 and IgE antibodies contained in sera of patients allergic to peach, was tested.
Each strip was incubated overnight at 4 ºC in 3 ml of PBS containing 0.1% (v/v) Tween 20, 0.5% (w/v) Blotting-Grade Blocker and 1 μl of solution containing IgG antibodies or 150 μl of human serum.

The strips were washed for 3 times with PBS containing 0.1% (v/v) Tween 20 and then they were incubated for 5 h at 4 ºC in 3 ml of PBS containing 0.1% (v/v) Tween 20, 0.5% (w/v) Blotting-Grade Blocker and 1 µl of solution containing respectively goat-anti-rabbit IgG, donkey-anti-mouse IgG or mouse-anti-human IgE antibodies labelled with IRDye® 800 CW.
Three strips that had not been treated with the primary antibodies were incubated with the three secondary antibodies respectively, as described above, to detect possible nonspecific binding of the latter with the analytes.
The strips were washed for 3 times with PBS containing 0.1% (v/v) Tween 20 and then they were visualized using an infrared fluorescence detection Odyssey Imager and software (LI-COR Biotechnology).
CAP-inhibition assays: ImmunoCAP (Phadia, Uppsala, Sweden) assays were performed according to manufacturer's instructions. Prior to incubation in CAPs coated with recombinant Pru p 3 (Product code: f420 - Phadia), sera were inhibited for 1 h, at room temperature, with a 100-fold molar excess, in respect to the amount of Pru p 3 in the CAP, of the peptides sPru p 3 (1-37) and sPru p 3 (38-91), a mixture of the 2 peptides, or the natural Pru p 3 standard. An uninhibited serum sample, diluted with PBS, was analysed to get the reference value.

6.4.2.13 Circular Dichroism (CD) Spectroscopy

CD spectra were measured on a Jasco J-715 CD spectrometer (Jasco Inc), at 20°C, using a 1 mm path length quartz cell and solutions with a concentration of 37 µM in the case of the peptide sPru p 3 (1-37), while for sPru p 3 (38-91) a 5 mm path length quartz cell and solutions with a concentration of 15 µM were employed.
The analyses were performed in H_2O and an aqueous solution containing SDS 2 mM.
K2D3 software (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/) was used for the deconvolution of the spectra.

6.5 References

[1] Villain M., Vizzavona J., Gaertner H. Chemical ligation of multiple peptide fragments using a new protection strategy. Proc. Second Int. Seventeenth Am. Peptide Symp. 107–108, **2001**.

[2] Zheng J.S., Tang S., Huang Y., Liu L. Development of New Thioester Equivalents for Protein Chemical Synthesis. Acc. Chem. Res. 46, 1, 12475–2484, **2013**.

[3] Pentelute L.B., Kent S.B.H. Selective Desulfurization of Cysteine in the Presence of Cys(Acm) in Polypeptides Obtained by Native Chemical Ligation. Org. Lett. 9, 4, **2007**.

[4] Wan Q., Danishefsky S.J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. Angew. Chem. Int. Ed. 46, 9248–9252, **2007**.

[5] Blanco-Canosa J.B., Dawson P.E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. Angew. Chem. Int. Ed. 47(36), 6851-6855, **2008**.

[6] Siman P., Blatt O., Moyal T., Danieli T., Lebendiker M., Lashuel H.A., Friedler A., Brik A. Chemical synthesis and expression of the HIV-1 Rev protein. Chembiochem. 12(7), 1097-1104, May **2011**.

[7] Hendrix J.C., Halverson K.J., Jarrett, J.T., Lansbury P.T. A novel solvent system for solid-phase synthesis of protected peptides: the disaggregation of resin-bound antiparallel beta-sheet. J. Org. Chem. 55, 4517-4518, **1990**.

[8] Quaderer R., Hilvert D. Improved Synthesis of C-Terminal Peptide Thioesters on “Safety-Catch” Resins Using LiBr/THF. Org. Lett. 3, 20, **2001**.

[9] Bang D., Kent S.B.H. A One-Pot Total Synthesis of Crambin. Angew. Chem. Int. Ed. 43, 2534–2538, **2004**.

[10] Haase C., Rohde H., Seitz O. Native Chemical Ligation at Valine. Angew. Chem. Int. Ed. 47, 6807 –6810, **2008**.

[11] Durek T., Torbeev V.Y., Kent S.B.H. Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme. PNAS 104 (12), 4846–4851, **2007**.

[12] Zuidmeer L., van Leeuwen W.A., Budde I.K., Cornelissen J., Bulder I., Rafalska I., Besolí N.T., Akkerdaas J.H., Asero R., Fernandez Rivas M., Gonzalez Mancebo E., van Ree R. Lipid transfer proteins from fruit: cloning, expression and quantification. Int. Arch. Allergy Immunol. 137(4), 273-281, Aug **2005**.

7 General discussion and conclusions

Food allergy is an issue of major concern for human health, as it can lead to life threatening symptoms and affects about 6-8% of young children and 3-4% of the global adult population[1].

This kind of immune mediated reaction is generally triggered by proteins, which are defined as class 1 or "true" food allergens if they are able to sensitize directly for allergic reactions through the gastrointestinal tract.

Non-specific Lipid Transfer Proteins (nsLTP) are widely distributed throughout the plant kingdom and were recognized as being relevant class 1 food allergens in many commonly consumed fruits and vegetables, especially for sensitized people in the Mediterranean area[2].

Allergy to members of the nsLTP family is often associated with systemic and severe clinical symptoms and the fact that they present highly conserved sequences and tridimensional structures that enable IgE recognition, promotes cross-reactivity among this type of proteins[3].

Once a food allergy has been diagnosed, the exclusion of the triggering food from the diet constitutes the best way to manage this condition.

Anyway, desensitization through immunotherapy (a process sometime referred to, quite inappropriately, as "vaccination") is also emerging as potential way of managing food allergy and avoiding excessive adverse reactions in the case of an accidental ingestion. Especially oral immunotherapy (OIT), consisting in the regular ingestion of small amounts of the pure triggering allergenic protein or of a food product containing the latter, is being extensively investigated[4].

To avoid allergic side effects of the therapy, many new developed treatments involve the use of hypoallergenic isoforms of the allergens or of specific fragments of the latter. These species are produced in such a way that they display reduced allergenicity, because they lack IgE-epitopes, but show good T cell antigenicity, a prerequisite of tolerance induction[5].

Methods employed for the diagnosis of food allergies are nowadays increasingly moving from the use of crude allergen extracts towards component-resolved techniques, in which the specific allergen responsible for sensitization or adverse reaction is searched[6]. This kind of application requires pure allergens that are also needed as reference materials for the calibration and

standardization of quantification methods among different laboratories and operators, as well as for risk assessment in the food industry to detect contamination. In addition, also the study of allergic reactions at a molecular level, which comprehension would ease the development of possible cures, could take advantage from the availability of pure allergens. Finally, as previously mentioned, pure allergens or parts of them are also useful tools for food allergy management through desensitisation.
Traditionally, allergenic proteins are obtained by extraction from biological tissues or by means of recombinant DNA techniques.
A third, not yet explored, possible source of allergens could be their total chemical synthesis: in the last decades, several techniques allowing the complete chemical production of proteins have been developed and proteins and enzymes of more than 200 amino acids have been successfully synthesized chemically. The biological properties of the obtained macromolecules were proven to be identical to those of their natural occurring homologues[7].
In the present thesis all these three approaches were exploited to try to obtain pure allergenic Lipid Transfer Proteins, defined as being a model of true food allergens[8], from different fruits.

An optimized procedure efficiently combining extraction, ultrafiltration and chromatography allowed to purify and identify a previously unknown LTP in almond (*Prunus dulcis*). A defatted flour was obtained from these nuts and used to produce a total protein extract. The protein fraction in the right mass range for LTPs (7-10 kDa) was isolated from this extract by means of two sequential ultrafiltration processes and analysed through SDS-PAGE. In-gel tryptic digestion, followed by LTQ-Orbitrap mass analyses of the produced peptide mixtures and alignment of the experimental data on protein databases for green plants, revealed the presence of a band in the gel corresponding to a fragment of a potential LTP. Reverse phase liquid chromatography was used to purify the corresponding protein that was then characterized at a molecular level using advanced mass spectrometry techniques. The exact mass of the entire protein was determined and its amino acidic sequence was established after tryptic and chymotryptic digestion followed by LTQ-Orbitrap analyses of the produced

peptide mixtures, alignment of the produced experimental data on protein databases for green plants and partial *de novo* sequencing.
Finally, the potential allergenicity of the identified amino acidic sequence was confirmed using *in silico* approaches.
In this work, ultrafiltration was shown to be a quick and simple manner to obtain sample desalting and protein fractioning according to molecular weight, substituting the more widely used size exclusion chromatography and simplifying or even avoiding the subsequent purification by liquid chromatography.
The identification and characterization of allergens contained in almond is of crucial importance, since the use of this nut is largely diffused in sweet bakery and many other food products could contain hidden traces of almond due to cross-contamination.
Together with other tree nuts, almond represents one of the most common sources of food allergens and its presence in food products has to be mandatory declared on their label, according to the EU Food Information for Consumers Regulation No. 1169/2011.
The same experimental procedure used to isolate the novel almond LTP was applied to pistachio, but any attempt to detect the presence of an LTP in these nuts did not yield the expected results.

The production of Jug r 3, a recognized allergenic LTP from walnut (*Junglas regia*)[9], was approached through recombinant DNA techniques. The gene coding for this protein was amplified through PCR, starting from walnut cDNA and inserted into a vector, which was then transformed into *E. coli* cells for expression.
The allergen was produced as a fusion protein, having a SUMO-tag bearing a 6 His stretch at its N-terminal and this allowed its facile purification by means of affinity chromatography. The used tag had also been reported to improve expression, folding and solubility of the target protein[10].
Once the fusion protein had been isolated from the host proteins, the SUMO-tag was removed by enzymatic cleavage, employing a SUMO specific protease.
The digested protein was purified again chromatographically and analysed by means of advanced mass spectrometry techniques to determine its exact mass;

this analysis allowed to easily highlight a mutation in the product, thus demonstrating to be an important tool to verify the outcome of a recombinant protein production.
The sequencing of the recombinant plasmid used for the synthesis of the walnut LTP evidenced that an error was already present at level of the nucleotide sequence, thus it was not introduced during the protein expression that worked properly.
Despite the presence of the mutation, the produced and purified recombinant allergen displayed immunogenic activity, as was confirmed by means of immunoblot, using LTP specific IgG antibodies raised in rabbits and sera of patients allergic to LTPs contained in various fruits.
Especially for nuts, the use of recombinant allergens as diagnostic tools is favourable, since their naturally occurring counterparts could be underrepresented in protein extracts or even totally lost during defatting processes performed prior to protein extraction. This is particularly true in the case of allergens that, like LTPs, are only minor constituents of the total protein fraction.

Two strategies for the total chemical synthesis of Pru p 3, the major allergen of peach (*Prunus persica*) in the Mediterranean area and best characterized member of the LTP family[11, 12], have been finally developed.
In the prospect of using Native Chemical Ligation (NCL) and peptide-hydrazide ligation, followed by desulfurization, to join peptide fragments composing the entire protein, the latter was retro-synthetically divided into five parts at level of four alanine residues.
Preliminary tests on the experimental conditions needed for the foreseen reactions, allowed to choose one of the strategies developed for the protein assembly.
The five fragments were produced through Solid Phase Peptide Synthesis (SPPS) according to the Fmoc protocol and assembled in a convergent fashion, to yield the whole protein stretch.
All the reaction products were characterized using Ultra Performance Liquid Chromatography coupled with Mass Spectrometry (UPLC-MS) and purified by means of ultrafiltration and/or LC.

The experimental conditions for the desulfurization process and the deprotection of the native cysteines were set up on the two final halves of the protein foreseen by the developed synthetic strategy. The obtained peptides, sPru p 3 (1-37) and sPru p 3 (38-91), were tested for immunogenicity using LTP specific IgG antibodies raised in rabbits and mice and sera of patients allergic to peach. Good binding ability of the antibodies, particularly to the peptide sPru p 3 (38-91), was detected by immunoblot. ImmunoCAP (Phadia) inhibition assays were performed to point out the ability of the two peptides to hinder the binding of specific IgE antibodies to a Pru p 3 standard; no relevant results were achieved using the species singularly, while a mixture of the two peptides allowed obtaining a certain grade of inhibition. Thus, the two protein halves are apparently able to interact with a single IgE antibody in a concerted manner.
Moreover, the structure of the two peptides was characterized by means of Circular Dichroism (CD) spectroscopy and it was found that, in conditions mimicking the degree of hydrophobicity which can be found in cells, they both have conformations with the high α-helix content typical for LTPs.
To produce the whole protein in its native sequence, the synthesis has to be repeated on a bigger scale and the product of the last ligation needs to be purified, to perform the final desulfurization and deprotection steps on it. Once synthesized, the protein will be compared with the natural occurring allergen, concerning its structure and its immunological activity.
Since all the members of the LTP family show a high degree of sequence homology, the developed strategy could easily be modified to produce LTPs contained in plant-sources different than peach.

Concluding, all the three available methodologies for the obtainment of pure proteins were efficiently applied to produce allergenic LTPs.
The extraction from biological tissues is certainly the most straightforward approach, but its applicability can be limited in the case of allergens that constitute only a minor part of the total protein fraction, as demonstrated in case of pistachios. Moreover, this technique does not allow the modification of the protein sequence, which would be useful to study consequent variations in

reactivity or to obtain hypoallergenic or differently allergenic isoforms, which could be employed for immunotherapy.
These problems can be overcome by the use of recombinant techniques, which potentially allow to produce the desired amount of any protein, starting from the gene coding for it, and to introduce mutations in its primary structure. Anyway, the production of recombinant proteins, although a very mature technique, does not allow to introduce non-coded unnatural amino acids in the sequence, which could be used to study in deep molecular details the influence of the structure on allergenicity or to obtain materials usable for desensitation.
The complete control on the protein sequence can only be obtained by its total chemical synthesis, which allows to easily incorporate unnatural amino acids, post-translational modifications or labelling agents, opening new opportunities for the understanding of protein molecules. Moreover, an optimized total synthetic approach could give accessibility to consistent amount of protein (in mg scale) in an affordable way and in a quite pure form.
The production of an allergen by total chemical synthesis, approached for the first time in this thesis, would constitute an important step forward in the research field of food allergy, since chemical synthesis facilitates the study of the structure-property relationship of protein functions at atomic resolution. Moreover, the biological mechanism of action of the synthesized allergens could be tightly controlled, thus allowing the development of efficient therapics.
Finally, the chemical production of allergenic proteins would avoid contamination and other issues that can arise from their obtainment from natural or recombinant sources[13].
In all the three approaches, mass spectrometry turned out to be a crucial tool for the analysis of proteins and peptides, to identify or confirm their sequence, and should be implemented in all case as the routine quality check technique in every production of pure allergens.

7.1 References

[1] Husain Z., Schwartz R.A. Food allergy update: more than a peanut of a problem. Inter. J. Dermatol. 52, 286–294, **2013**.

[2] Borghesan F., Mistrello G., Roncarolo D., Amato S., Plebani M., Asero R. Respiratory allergy to lipid transfer protein. Int. Arch. Allergy Immunol. 147(2), 161-165, **2008**.

[3] Hauser M., Roulias A., Ferreira F., Egger M. Panallergens and their impact on the allergic patient. Allergy, Asthma, Clin. Immunol. 6, 1−14, **2010**.

[4] Nowak-Węgrzyn A., Albin S. Oral immunotherapy for food allergy: mechanisms and role in management. Clin. Exp. Allergy. 45(2), 368-83, Feb **2015**.

[5] Ferreira F., Ebner C., Kramer B., Casari G., Briza P., Kungl A.J., Grimm R., Jahn-Schmid B., Breiteneder H., Kraft D., Breitenbach M., Rheinberger H.J., Scheiner O. Modulation of IgE reactivity of allergens by site-directed mutagenesis: potential use of hypoallergenic variants for immunotherapy. FASEB J. 12(2), 231-242, Feb **1998**.

[6] Tuano K.S., Davis C.M. Utility of Component-Resolved Diagnostics in Food Allergy. Curr. Allergy Asthma Rep. 15(6), 534, Jun **2015**.

[7] Kent S.B.H. Total chemical synthesis of proteins. Chem. Soc. Rev. 38, 338–351, **2009**.

[8] Salcedo G., Sánchez-Monge R., Díaz-Perales A., Pacios L.F. Review. Plant food allergens: peach non-specific lipid transfer protein Pru p 3 as a model. Span. J. Agric. Res. 6 (Special issue J. M. Malpica), 30-37, **2008**.

[9] Pastorello E.A., Farioli L., Pravettoni V., Robino A.M., Scibilia J., Fortunato D., Conti A., Borgonovo L., Bengtsson A., Ortolani C. Lipid transfer protein and vicilin are important walnut allergens in patients not allergic to pollen. J. Allergy Clin. Immunol. 114(4), 908-914, Oct **2004**.

[10] Butt T.R., Edavettal S.C., Hall J.P., Mattern M.R. SUMO fusion technology for difficult-to-express proteins. Protein Expr. Purif. 43(1), 1-9, Sep **2005**.

[11] Pastorello E.A., Farioli L., Pravettoni V., Ortolani C., Ispano M., Monza M. The major allergen of peach (Prunus Persica) is a lipid transfer protein. J. Allergy Clin. Immunol. 103, 520-526, **1999**.

[12] Fernandez-Rivas M., Gonzalez-Mancebo E., Rodriguez-Perez R., Benito C., Sanchez-Monge R., Salcedo G., Alonso D., Rosado A., Tejedor M.A., Vila C., Casas M.L. Clinically relevant peach allergy is related to peach lipid transfer protein, Pru p 3, in the Spanish population. J. Allergy Clin. Immunol. 112, 789-795, **2003**.

[13] Harmand T.J., Murar C.E., Bode J.W. New chemistries for chemoselective peptide ligations and the total synthesis of proteins. Curr. Opin. Chem. Biol. 22, 115-21, Oct **2014**.

PERSONAL INFORMATION

Sofie Buhler

Via I. di Alessandria 15, 43013 Langhirano (PR) Italia

+3905211890924 +393479225824

sofie.buhler@yahoo.com

Sex Female

Date of birth 21/04/1987

Nationality Italian

WORK EXPERIENCE

January 2013 - Present

PhD student in Food Science and Technology

University of Parma, Italy

- Solid Phase Peptide Synthesis and production of polypeptides through chemical ligation
- Protein extraction from food matrices and their identification by means of electrophoresis, enzymatic digestion and MS/MS analysis of the produced peptides
- Purification of peptides and proteins by means of chromatography and their analysis through mass spectrometry

May – September 2015

Visiting PhD student

Academisch Medisch Centrum – University of Amsterdam, Netherlands

- Characterisation of allergenic proteins through immunological tests
- Production of proteins by means of recombinant DNA techniques

February 2012 – February 2013

Chemical analyst

Laterlite SpA, Fornovo di Taro (PR), Italy

- Analysis of exhausted oils and emulsions used as alternative fuels in the production of expanded clays

EDUCATION AND TRAINING

October 2009 – October 2011

Master Degree in Organic Chemistry (110/110 cum laude)

University of Parma

- Organic synthesis
- Mass spectrometry
- Gas-chromatography
- NMR

October 2006 – July 2009	**Bachelor Degree in Chemistry** (110/110 cum laude) University of Perugia, Italy
September 2001 – June 2006	**Scientific High School Diploma** (100/100) ITAS G. Bruno, Perugia, Italy

PERSONAL SKILLS

Mother tongues: Italian, german

Other languages

Understanding	Speaking		Writing
	Spoken interaction	Spoken production	
English C1	B2	C1	C1

Levels: A1/2: Basic user - B1/2: Independent user – C1/2 Proficient user
Common European Framework of Reference for Languages

Communication skills: During my period as PhD student I held many general and organic chemistry lessons for bachelor students in Food Science

Organisational / managerial skills:

- I was correlator of master degree theses in Pharmaceutical Chemistry and Technology and in Biotechnology
- During my period as PhD student I helped to organize the educational chemistry laboratories for the bachelor students in Food Science

Job related skills:

- Solid phase peptide synthesis and chemical ligation reactions
- Protein extraction from natural matrices
- Liquid chromatography
- Mass spectrometry
- Gel electrophoresis and western-blotting
- Immunological tests to verify the allergenicity of proteins

Computer skills:

- Excellent command of Microsoft Office tools (European Computer Driving Licence- ECDL)
- Good ability in the use of software for chromatographic analysis (Empower) and mass spectrometry (Masslynx, Excalibur)

Driving licence: Category: B

ADDITIONAL INFORMATION

Pubblications

- S. Buhler, T. Tedeschi, A. Faccini, C. Garino, M. Arlorio, A. Dossena, S. Sforza. Isolation and full characterisation of a potentially allergenic lipid transfer protein (LTP) in almond. Food Addit Contam Part A , 32 (5), 648–656, 2015.
- C.I. Butré, S. Buhler, S. Sforza, H. Gruppen, P. Wierenga. Spontaneous, non-enzymatic breakdown of peptides during enzymatic protein hydrolysis. Biochim Biophys Acta. 1854(8), 987-994, 2015.
- C. G. Piscopo, S. Buhler, G. Sartori, R. Maggi. Supported sulfonic acids: reusable catalysts for more sustainable oxidative coupling of xanthene-like compounds with nucleophiles. Catal. Sci. Technol. 2, 2449–2452, 2012.

Conferences

Oral presentation entitled: Isolation and full characterisation of a potentially allergenic lipid transfer protein (LTP) in almond. XVIII Eurofoodchem Congress, Madrid, Spain, 13-16 October 2015.

Poster presentations

- Poster entitled: Characterization of LTP allergen by proteomic approaches and total chemical synthesis. XIX Workshop on the Developments in the Italian PhD Research on Food Science, Technology and Biotechnology, Bari, Italy 24-26 September 2014.
- Poster entitled: Characterization of LTP protein allergen by total chemical synthesis and proteomic approches. V EuCheMS Congress, Istanbul, Turkey, 31 August-4 September 2014.
- Poster entitlled: Extraction and characterization of novel allergenic proteins from tree nuts. VIII European Summer School in Advanced Proteomics, Bressanone, Italy, 3-9 August 2014.

Courses

- Basic course on scientific writing, research organization, oral and poster presentation – Parma, Italy, 6-10 October 2014.
- .VIII European Summer School in Advanced Proteomics, Bressanone, Italy 3-9 August 2014.
- XVII School in mass spectrometry, Pontignano, Siena, Italy 18-22 March 2013.

Awards

My master's degree thesis was awarded by the Italian Chemical Society for research in the field of catalysis applied to organic synthesis (Milan, Italy 21 November 2011).