UNIVERSITA' DEGLI STUDI DI PARMA

Dottorato di ricerca in
Scienze e Tecnologie Alimentari

Ciclo XXIV

PROBING QUALITY AND SAFETY OF TOMATO PRODUCTS BY GENOMIC AND PROTEOMIC TOOLS

Coordinatore:
Chiar.mo Prof. Davide Barbanti

Tutors:
Chiar.mo Prof. Rosangela Marchelli
Chiar.mo Prof. Stefano Sforza

Dottorando: Mariangela Bencivenni

Ai miei genitori, faro perenne e insostituibile della mia vita
A Dimer, esempio a cui da sempre e continuamente ambisco
A Valerio e Lavinia
Alle mie nonne

"Tu che sei come sei,
che non cambierai mai,
promettimi che
ci sarà sempre un posto
che tieni caldo per me"
L.L.

"And when you kiss me
on that midnight street
sweep me off my feet
singing
'Ain't this life so sweet?'"
D.G.

"Non c'è montagna più alta di quella che non scalerò"
L.J.

"Memento audere semper. Et ventis adversis"
G.d.A

Abstract

The present Ph.D. thesis deals with the use of advanced genomic and proteomic techniques for the assessment of quality and safety of one of the most important traditional and worldwide renowned Italian crop products: tomato.
In the first section, innovative proteomic approaches have been applied to the issue of tomato allergy. Classical immunological techniques, based on gel electrophoresis for the separation of proteins, along with advanced mass spectrometric systems, have been applied to a set of 12 tomato ecotypes, allowing the detection and the characterization of the main allergens. Moreover, several novel isoforms of allergenic tomato Lipid Transfer Protein (LTP) have been detected and characterized, both in peel and in seed fractions of common tomatoes, by means of high-resolution mass spectrometric techniques, coupled with classic and innovative purification strategies.
In the second section, advanced genomic techniques have been applied to the problem of tomato variety identification. In particular, the discrimination of seven tomato varieties has been achieved by means of specific probes, namely Peptide Nucleic Acids (PNAs), recognizing characteristic tomato SNPs. After that the performances, in terms of binding affinity and mismatch recognition, of different modified PNAs (2D- and 5L- chiral-box) have been tested and compared to fully 'achiral' PNA, both in solution and on solid surfaces, the best PNA model was exploited in the design of a PNA-based microarray and in a PNA-mediated PCR clamping system for the recognition of tomato specific SNPs.

Keywords: tomato, allergens, proteomics, food safety, mass spectrometry, genomics, single nucleotide polymorphisms, peptide nucleic acids, microarray, PCR clamping.

Riassunto

La presente tesi di dottorato riguarda l'utilizzo di tecniche avanzate di genomica e proteomica per la valutazione della qualità e della sicurezza di una delle più importati e rinomate colture tradizionali italiane: il pomodoro.
Nella prima sezione, approcci innovativi di proteomica sono stati applicati al problema delle allergie al pomodoro. Tecniche immunologiche classiche, basate sull'elettroforesi su gel per la separazione delle proteine, affiancate da sistemi di spettrometria di massa avanzati, sono state applicate ad un set di 12 ecotipi di pomodoro, consentendo l'identificazione e la caratterizzazione dei principali allergeni. Inoltre, sono state individuate e caratterizzate diverse isoforme della proteina allergenica del pomodoro Lipid Transfer Protein (LTP), sia negli estratti di buccia che di semi di pomodori comuni, mediante tecniche di spettrometria di massa ad alta risoluzione accoppiate a strategie di purificazione classiche ed innovative.
Nella seconda sezione, tecniche di genomica avanzata sono state applicate alla questione dell'identificazione varietale in pomodoro. In particolare, è stata effettuata la discriminazione di sette varietà di pomodoro mediante l'utilizzo di sonde specifiche, chiamate acidi peptido nucleici (PNAs), che riconoscono SNP caratteristiche del pomodoro. Dopo che le prestazioni di PNA diversamente modificati (2D- and 5L- chiral-box), in termini di affinità di legame e di riconoscimento di

mismatch, sono state testate e confrontati con quelle del PNA chirale, sia in soluzione che su superficie, il miglior modello di PNA è stato utilizzato per il design di un microarray basato sui PNA e per lo sviluppo di un sistema PNA-mediated PCR clamping per il riconoscimento di SNP specifiche del pomodoro.

Parole chiave: pomodoro, allergeni, proteomica, sicurezza alimentare, spettrometria di massa, genomica, polimorfismi del singolo nucleotide, qualità alimentare, acidi peptido nucleici, microarray, PCR clamping.

Table of Contents

Chapter I –Preface

The term 'quality' is defined according to DIN ISO 9000 as the totality of features relevant to the ability of a product to fulfil its requirements (International Institute for Standardization, 2005). However, the concept of food quality should be considered on a much broader basis as the different demands of the manufacture, the consumer, the surveillance and the legislative bodies must be taken into account in order to obtain healthy and safe products without neglecting the economic, ideological and ecological issues associated with food quality.

First, the consumer wants 'healthy' products with high nutritional value with regard to macronutrients like proteins, carbohydrates, fats and fibre as well as vitamins and trace elements, and also for compounds having functional properties in the body, like antioxidants. Second, these products should also accomplish the requirements for superior flavour and texture. Third, a typical geographic origin or the use of defined ingredients, which often result in higher market prices, are also important factors for evaluating the quality of food, not because of changes in nutritional value, but in terms of consumer expectation deception, selling cheap products as high-price premium food. Last, but not least, obnoxious compounds should be absent in quality foods or present under the limits set by the regulatory bodies. There is no doubt that a quality food should primarily be safe.

In this scenario, several directives and laws have been issued for the assessment of hazards, by means of systematic analysis systems dedicated to the fast, sensitive and selective identification of perils. In addition, the European Commission also adopted the framework regulation EC/178/2002 which mainly aims at preventing fraudulent or deceptive practises, adulteration of food and any practice that may mislead the consumer.

Whereas on one hand these measures guarantee consumer's expectations for information and protection, on the other hand they imply an increase in the management procedures and in their costs for food industries. The speed-up in the development of new technologies in the fields of biotechnology, chemistry and food science, and their perfect integration, have allowed the implementation of tools and procedures for the food industries and the supervising authorities, helping the analysis simplification, the time-saving and the cost-effectiveness. In particular, the enormous advances in genomic and proteomic techniques have allowed in the last years to gain deeper insights on the molecular composition of food, in turn allowing to gain more information on the quality and the safety of the foods themselves.

The overall aim of this PhD thesis is to explore the use of advanced genomic and proteomic techniques for the assessment of quality and safety of one of the most important traditional Italian crop products: tomato. In particular, genomic techniques will be applied to the problem of variety identification and proteomic techniques to the issue of allergenicity.

1. Tomato

1.1 Historical background and botanical description

The origin of tomatoes belongs to South America and precisely to the Andes (the modern Peru, Bolivia, Chile and Ecuador), where they grown wildly. Although in America tomato was already included in the diet of natives, this distinctive vegetable of Mediterranean area was imported to Europe only in the first half of the sixteenth century, when Spanish conquistadores landed to America, where it spread just as an ornamental plant.

At the beginning of 1600th, nearly a century after it has been discovered, it was still consider unhealthy, since it was believed to be poisonous like the other *Solanaceae,* due to its high content of tomatine. Moreover, tomato was thought to have aphrodisiac properties and, for this reason, it was used by alchemists for their magic potions. This belief could also explain the names that different European languages used to refer to it: *love apple* in English, *pomme d'amour* in French, *Libesapfel* in German and *pomo* (or mela) *d'oro* in Italian, all definitions which referred to love. Nowadays, all these names have been replaced by the original Aztec name, *tomatl*. It was just at the end of the eighteenth century that tomato spread for human nutrition in Europe, especially in France and in the southern of Italy, but while in France it embodied the royal meal, in Italy it became the basic food for the starving population. The first transformation process in puree and the potting of boiled material dates back to 1762.

In the same period, Linnaeus placed the tomato in the genus *Solanum* as *Solanum lycopersicum*, which in Latin meant wolf-peach, reflecting the belief that tomato was poisonous.[1] Due to its substantial differences with potato and aubergine, Miller created for tomato another genus: from there the new name of *Lycopersicon esculentum*, which helped with the acceptability of tomato as a food. Phylogenetic relationships between *Solanum* and *Lycopersicon* have been the subject of a great debate for a long time, with many *Solanaceae* researchers recognizing *Lycopersicon* as a distinct genus while others suggesting its merger with *Solanum*. More recently, based on much molecular and morphological information, tomato has been moved again to the genus *Solanum*, making *Solanum lycopersicum* the correct name.

In addition to the cultivated species *L. esculentum* and its wild form *L. esculentum var. cerasiforme*, many other related wild species are know, all of them native to Western South America, mainly Peru. Only the cultivated species and wild cherry are found outside this range and are common throughout many parts of the world. However, the natural habitat of *Lycopersicon* is highly variable, from very dry to very wet and from coastal to mountainous areas of more than 3300 m elevation.[2] This diversity in habitat has undoubtedly contributed to the great variation that can be found in *Lycopersicon* genus. All species within *Lycopersicon* produce perfect, hermaphrodite flowers. A complete range of mating systems is found, from autogamous to obligately outcrossed self-incompatible biotypes. All tomato species are diploid (2n = 2x = 24) and are similar in chromosome number and structure.[3] They have yellow flowers and the stamens are joined to produce an anther cone that surrounds the carpels. Botanically, tomato fruit is classified as a berry with a thin epicarp, a fleshy mesocarp and an endocarp divided into locular cavities containing

several seeds. The number of carpels in the flowers corresponds to the number of locules in the fruits. Both leaves and stem are densely covered by hairs excreting the characteristic tomato smell. It is a perennial herbaceous plant, although in temperate climates it is grown as an annual. The growth habit can be erect or prostrate. According to the shape of fruits and leaves and to the growth habit, tomatoes differ in: 'cherry' tomatoes – small size, round-shaped with two cavities; 'pear' tomatoes – medium size, pear-shaped with three cavities; 'potato-leaved' tomatoes – big leaves with intact edges; 'compact-habit' tomatoes – bushy with upright stems. Besides the shape of fruits and leaves, tomato varieties may be classified according to the use in: fresh consumption tomato, round/globular shaped, smoothed or ribbed surfaces; peeled tomato, cylinder/pear- shaped; concentrate tomato, square/oblong shaped.
Tomato fruits contain 93–97% of water. The dry matter is composed by sugars (40–60%), proteins and amino acids (15–20%), organic acids (4–10%), minerals (in particular potassium), vitamins and pigments (vitamin A, C, lycopene and β-carotene), insoluble matter (cellulose, emicellulose and pectins). Seeds are rough, flattened and discoidal, with 1000 seeds weighing 2.5–3.5 g.[4] Fruit colour varies depending on the species, from red to orange to yellow to green.
Besides the variety, the growing conditions are the major factors influencing the colour of the fruits: a mild shining promotes the formation of lycopene (which confers the red colour), while β-carotene (yellow/orange colour) stores up under intensive light treatments.

1.2 Economic importance

The cultivated tomato is widely grown around the world and constitutes a major agricultural industry.
The global tomato production is approximately 1.25 x 10^8 t/year, produced on about 4.5 million hectares (FAOSTAT data, 2006, http://faostat.fao.org). The 10 leading fruit producing countries are, in descending order, China (3.2 x 10^7 t/year), United States, Turkey, India, Egypt, Italy, Iran, Spain, Brazil and Mexico (FAOSTAT data, 2006, http://faostat.fao.org). The highest production per hectare is reported in the Netherlands (473 t/ha year) where tomatoes are exclusively grown under greenhouse conditions (FAOSTAT data, 2006, http://faostat.fao.org). The intake of raw tomatoes differs substantially throughout Europe: 5.9 g/day are consumed in the Netherlands, 47.5 g/day in Germany and 112.5 g/day in Greece. Further, global tomato consumption has increased over the past two decades (USDA data 2007, http://static.globaltrade.net).
According to Italtrade.com, Italy produces more than 4 million tons of tomatoes annually, roughly 90% of which is meant for processing. Italy leads the EU in tomato production, accounting for 38% of the total EU-27. Of these, 5.2 million tons are meant for processing (53% of the whole volume transformed by the entire European Union) and 1.3 million are meant for fresh consumption. In this last segment of the market Italy plays a less important role, representing only 19% of European supply (source: AgroalimentareNews). Generally, conditions in Italy allow for production of high quality and quantity tomatoes throughout the year. The cultivation of processing tomatoes is possible in many parts of Italy, but three regions account for almost 90% of the production. Nearly 40% of the production is in the Puglia region, another 40%

is localized in Emilia-Romagna region and 10% in Campania region (USDA data 2007, http://static.globaltrade.net).
The data reported here showed that tomato is a symbol of the Italian food-industry and represents its most significant horticultural product.

1.3 Tomato food chain

The success of the tomato crop has always been connected to its industrial transformation. The development of the first tomato processing industries in Italy was due to the capitals provided by fruit and vegetable exporters, who realized the opportunity to expand the number of accessible markets by overcoming the problem of the perishable nature of fresh products. A considerable impetus to the cultivation of tomato was, in fact, recorded in Italy at the beginning of last century, in connection with technological development that allowed the rise of the canning and transformation industries.
Initially, tomatoes were simply canned and sterilized, according to the so-called "appertization". Later, mainly due to the opening of some processing plants in the South of Italy (especially in the Nocera-Sarno area), the "peeling" process of long-shaped tomatoes, typically grown in these areas, took place. The development of this industry sector caused the conversion of large agricultural areas to the cultivation of tomato, thus positively influencing the socio-economic aspects of local populations. Independently of the quality of the product, the worldwide diffusion of 'San Marzano' tomato variety was due to the enhancement of the tomato processing.[5]
The process of conservation begins straight after picking and it should comply with many rules to ensure quality: red fruit, uniform shape and size, high pulp content and a low number of seeds. The characteristics are further defined by the type of transformation that the tomato undergoes: peeled, puree, pulp or chopped, juice, dried flakes. Processing begins with tomato washing in tanks; tomatoes are then heated to a temperature of 90 °C, in order to remove the skin from the flesh, using mechanical or heat processes, and then they are transferred onto conveyor belts for sorting. In modern industrial production lines, the sorting process is performed by an optical sorter. The product is then submitted to a process of partial concentration. Lastly, the tins are filled with the fruit and juice and then vacuum sealed before pasteurisation.
Tomato processed products represent the major export of Italy's trade but, on the other side, growing imports have been needed in order to meet both domestic consumption requirements and a fairly stable export demand. Most of the imported material comes from non-European countries: this could affect the quality features of those tomato productions which are related to the creation of high-quality brand names, aiming both to protect typical food production, related to a specific geographical area, and to encourage those sections which include valuable and traditional products. For example, the legitimate and traditional production of 'San Marzano' has been protected by the certification mark protection PDO (Protected Designation of Origin) since 1999. The branding of protection by the European Union has proved a useful tool to protect such typical food and revitalize some productive sectors, including traditional products of value. So, in order to avoid an inappropriate use of these 'quality' claims, for example the fraudulent substitution of

premium varieties with others of lower commercial value, several analytical tools, mainly based on DNA analysis, have been developed to assess the authenticity of the raw material through all the stages of production and to avoid frauds which may damage both consumers and producers.

1.4 A model organism

Tomato is considered an excellent model system for studying genetic plant, both in basic and applied research. This is due to different reasons,[4] including ease of culture under a wide range of environments, short life cycle, photoperiod insensitivity, high self fertility and homozygosity, great reproductive potential, easy control of pollination and hybridization, diploid species with a rather small genome (950 Mbp),[6,7] lack of gene duplication, amenability to asexual propagation and whole plant regeneration,[8,9] ability to develop haploids,[10] and availability of a wide array of mutants,[11] and genetic stocks.[12] Recent availability of high molecular weight insert genomic libraries, including both YAC,[13] and BAC,[14] libraries, has facilitated map-based or positional cloning. Furthermore, members of *Lycopersicon* are easily transformed, and transgenic tomatoes are routinely produced using co-cultivation with *Agrobacterium tumefaciens*.[8,15] Tomato was the first food crop in the U.S. for which a genetically engineered variety was marketed,[16] and also for which a disease resistance gene was positionally coloned.[17,18] Currently, the euchromatic portions of the 12 tomato chromosomes are being sequenced, which will make tomato even more of ideal crop plant system for genomic studies (see below). For all these reasons tomato has long served as a model system for plant genetics, development, physiology, pathology, and fleshy fruit ripening, resulting in the accumulation of substantial information regarding the biology of this economically important organism. Many genomic tools are now available on this *Solanaceous* species and have rapidly generated a great amount of genomic resources, including mapping populations, mapped DNA markers, bacterial artificial chromosomes expressed sequence tag (EST) collections, mutants concerning fruit development and ripening,[19] and gene expression profiling of fruit development and maturation.[20]

Recently a remarkable amount of data have been gathered in the field of tomato proteomic, mainly thanks to the technological achievements which allowed high throughput screening of proteomic states. They largely deal with the study of proteome variations associated with the fruit ripening,[21,22] or with the seed germination.[23] A proteomic approach was also used to describe the proteome of tomato pollen,[24] and to identify differentially expressed proteins in the wild type and mutant anthers with the objective of analyzing their roles in normal pollen development and in male sterility.[25] Several attempts have been addressed to the proteomic study of tomato protein changes in response to different extreme growing conditions, such as heat treatment,[26] salt,[27,28] aluminium,[29] and iron stresses,[30] or to characterize 2S storage albumin from the seeds that cross-reacted with antiserum to the fruit lectin,[31] responsible of some adverse effects after their ingestion. Different studies have also carried out to investigate the molecular and biochemical mechanisms potentially involved in tomato fruit defense against pathogenic attacks.[32,33] These findings represent a significant contribution towards the construction of a comprehensive tomato proteome database, encompassing many

different species, and which could serve as a valuable resource for understanding the biology of the *Solanaceae* family.

1.5 Tomato genome sequencing

The nuclear genome of tomato has 12 chromosomes and approximately 950 Mbp of DNA, containing 59% of non-coding regions, 28% of sequences coding, 11% of transposon sequences and 2% of organelle sequences.[34] Approximately 77% of the chromosomal DNA is placed in the centromeric heterochromatin, which lacks in genes.[34] The tomato genome encodes about 35,000 genes, most of which are placed in the distal euchromatic regions of chromosomes,[35] with a density of 6.7 kb / gene, similar to that of Arabidopsis and rice.[34]

The 12 tomato chromosomes have been sequenced by an International consortium composed of 10 countries, which different chromosomes have been assigned to: China (3), France (7), India (5), Italy (12), Japan (8), Korea (2), Netherlands (6), Spain (9), United Kingdom (4) and USA (1,10,11) (http://www.sgn.cornell.edu/help/about/tomato_sequencing.pl.).

Tomato genomics has generated a huge amount of data resulting in the need for accessible databases, such as the database of the International Solanaceae Genomics Project (SOL), the SOL Genomics Network (SGN; http://sgn.cornell.edu). SGN gives free access to genetic and physical tomato maps and to DNA sequences of tomato genes, expressed sequence tags (ESTs), bacterial artificial chromosomes (BACs) and molecular markers. The International Solanaceae Genomics Project also includes the EU-SOL and Lat- SOL programmes. EU-SOL focuses on the development of high-quality and healthy tomato and potato varieties with improved consumer-, processor- and producer-directed traits. Lat-SOL aims at joining efforts in Latin America and promoting information and resource flow between laboratories working in basic and applied aspects of Solanaceae species.[36]

On October 23rd 2009 Keygen delivered to the Consortium the physical map of the tomato genome, built using the technology "Whole Genome Profiling" (WGP) (www.keygene.com). The entire sequence of 950 Mb that constitutes the tomato genome was covered by assembling BAC clones under conditions of high stringency. This map will be used as a skeleton to produce a 'higher quality' draft sequence of the tomato genome in the next years. Started in 2003, the sequencing project has ambitious research targets, including developmental and functional genomic studies of the *Solanaceae* family.[19]

1.6 Genetic variability and breeding

The cultivated tomato has limited variability, largely because of several population bottlenecks in the forms of founder events and natural and artificial selections that occurred during domestication and evolution of modern cultivars.[37] For example, tomatoes that were first introduced to Europe by Spanish explorers furnished the entire genetic base for the modern cultivars and consequently the current European and U.S. cultivars are highly similar to each other. It is estimated that only ~5% of the total genetic variation within *Lycopersicon* can be found within *L. esculentum*,[38,39] and genes for many desirable agricultural characteristics do not exist in this species. The related wild tomato species, however, are a rich source of desirable genes and

characteristics for crop improvement, though they remain largely under exploited, mainly due to difficulties that inter-specific crosses may represent.

Breeding new cultivars of tomato with improved characteristics started more than 200 years ago in Europe, mainly in Italy. Until 1950s, tomato breeding included development of multipurpose cultivars to meet several needs, including fresh market and processing industries. Subsequently, breeding objectives have depended upon method of culture, that is field or greenhouse grown, and whether the product has to be used as fresh or processed.[40] Today, fresh market and processing cultivars are quite distinct, largely as a result of the different quality requirements for intended use. However, the universal goal of tomato breeding for both fresh market and processing purposes has been to increase fruit yield per unit area. Other essential characteristics common to both industries include disease resistance, broad adaptability, earliness in maturity, ability to set fruit at adverse temperatures, resistance to rain-induced cracking, tolerance to major ripe-fruit rots, adequate vine cover, fruit firmness, and several other fruit quality characteristics. Specific traits needed in processing cultivars include compact, determinate plant habit and concentrated flowering and fruit set suitable for onceover machine harvest, ease of fruit separation from the vine (joint less characteristic), and specific fruit quality characteristics such as colour, pH, total acidity, soluble solids, total solids, and viscosity. Specific traits of interest in fresh market cultivars include large, round fruits with adequate firmness and shelf-life, uniform fruit size, shape and colour, appearance, freedom from external blemishes or abnormalities, texture, taste and flavour.[40]

Moreover, accessions of nearly all wild tomato species have been successfully used to introduce valuable traits for crop improvement, especially monogenic sources conferring resistance to fungal, nematode, bacterial and viral diseases.[36]

1.7 Tomato nutrients and their benefits on human health

Tomato and tomato-based products are important sources of many established nutrients and are predominant sources of some phytochemicals that may have health benefits. Based on the volume of consumption per person, tomatoes are relatively rich sources of folate, vitamin C, vitamin A, and potassium. The relative importance of tomatoes as contributors of these nutrients varies across populations, according to their consumption. For example, since they are highly consumed in Italy, tomatoes have been estimated as the second most important source of vitamin C after oranges.[41] Despite their low content in proteins and organic acids, tomatoes are a predominant source of several carotenoids, particularly lycopene and β-carotene, and of other potentially beneficial phytochemicals, including phenylpropanoids (phenolic acids), phytosterols, and flavonoids.[42,43]

The biological activities of carotenoids, such as β-carotene, are related in general to their ability to form vitamin A within the body. Because lycopene is not converted to vitamin A,[44] it may be entirely available for other properties. Two major hypotheses have been proposed to explain the anti-carcinogenic and anti-atherogenic activities of lycopene: non-oxidative and oxidative mechanisms. Among the non-oxidative mechanisms, lycopene is hypothesized to suppress carcinogen-induced phosphorylation of regulatory proteins such as p53 and Rb anti-oncogenes and stop cell division at the G0–G1 cell cycle phase.[45] Preliminary in vitro evidence also

indicates that lycopene reduces cellular proliferation induced by insulin-like growth factors, which are potent mitogens, in various cancer cell lines.[46] Lycopene also has been shown to act as a hypocholesterolemic agent by inhibiting HMG–CoA (3-hydroxy-3-methylglutaryl– coenzyme A) reductase.[47] Among the oxidative mechanisms, lycopene has been hypothesized to prevent carcinogenesis and atherogenesis by protecting critical cellular biomolecules, including lipids, lipoproteins, proteins and DNA.[48,49] In healthy human subjects, lycopene- or tomato-free diets resulted in loss of lycopene and increased lipid oxidation,[50] whereas dietary supplementation with lycopene increased serum lycopene levels and reduced endogenous levels of oxidation of lipids, proteins, lipoproteins and DNA.[49] The proposed mechanisms for the role of lycopene in the prevention of chronic diseases are summarized in Fig 1.

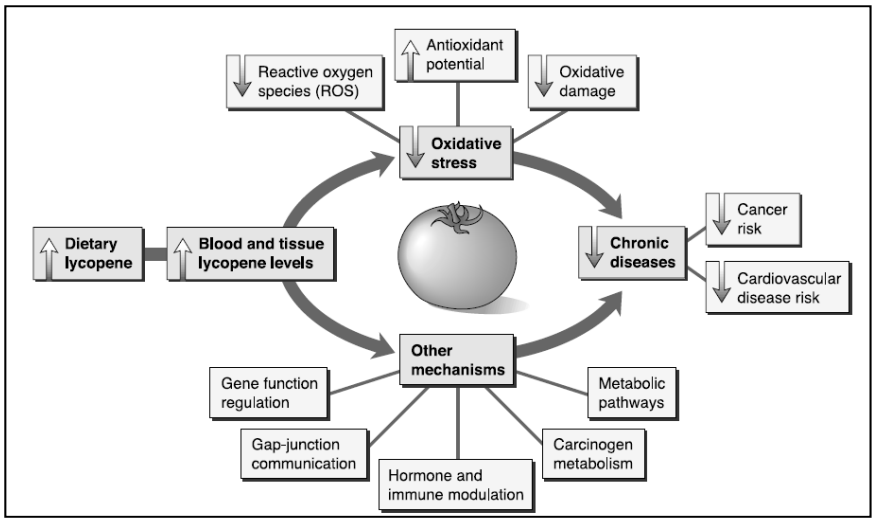


Fig 1. Proposed mechanisms for the role of lycopene in preventing chronic diseases. Dietary lycopene may increase the lycopene status in the body and, acting as an antioxidant, may trap reactive oxygen species, increase the overall antioxidant potential or reduce the oxidative damage to lipid (lipoproteins, membrane lipids), proteins (important enzymes) and DNA (genetic material), thereby lowering oxidative stress. This reduced oxidative stress may lead to reduced risk for cancer and cardiovascular disease. Alternatively, the increased lycopene status in the body may regulate gene functions, improve intercell communication, modulate hormone and immune response, or regulate metabolism, thus lowering the risk for chronic disease. These mechanisms may also be interrelated and may operate simultaneously to provide health benefits. Adapted from reference 51.

Tomatoes have also been recognized as an important source of dietary flavonoids because of a worldwide high consumption. Besides their relevance in plants, flavonoids are important for human health because of their high pharmacological activities as radical scavengers, their high antioxidant capacity and their ability to induce human protective enzyme systems, and their protective effects against cardiovascular diseases, cancers, and other age-related diseases.[52] The occurrence of flavonoids in tomato fruits is almost exclusively restricted to their skin, leaving only negligible quantities in the remaining parts of the fruit.[53] The main flavonoids in tomato fruits identified in previous literature have been reported to be rutin, naringenin,[54,55,56] and chalconaringenin.[57,58] Moreover, some minor flavonoids have been identified from tomato fruits as reviewed by Moco and co-workers,[59] among them kaempferol 3-rutinoside and naringenin 7-glucoside. By the way, flavonoid

levels were found to be different in the different varieties of fresh tomatoes: these differences can be attributed either to a variety-related metabonomic difference or to the different ripening levels at which the different varieties may have been harvested. Besides, it was evident that the harsh thermal treatments used in tomato processing caused a strong decrease for all the compounds considered, leading to level off the contents of these secondary metabolites.[56]

Due to its nutritional and healthy properties related to its content in antioxidant molecules, tomato is among the major components of diets around the world and thus it is intensively consumed. Nevertheless, several public concerns dealing with tomato consumption, are related to this growing and intensive uptake. Different studies reported fresh and processed tomatoes as elicitors of food allergies, among fruits and vegetables.[60,61,62] Clinical reactions towards tomatoes are mostly local symptoms in the oral mucosa, but tomatoes can also elicit systemic reactions and can even cause serious illnesses.[63,64] Moreover, several studies proved that some allergens may be still present in tomato products even after technological treatments, increasing the risk of adverse reactions when they occur in different meals as additional component.[65,66] Therefore, the study of tomato allergies, helped by the characterization and the identification of the main and the most common allergens, could represent an headway in the tackling of this public health issue.

2. References

[1] Rick CM. The tomato. *Science American*, 1978, **23**:76–87.

[2] Warnock SJ. A review of taxonomy and phylogeny of the genus Lycopersicon. *HortScience*, 1988, **23**:669–673.

[3] Rick CM, Butler L. Cytogenetics of the tomato. *Advances in Genetics*, 1956, **8**:267–382.

[4] Rick CM, Yoder JI. Classical and molecular genetics of tomato: highlights and perspectives. *Annual Review of Genetics*, 1988, **22**:281–300.

[5] Porretta S. Quality evaluation of canned whole tomatoes. *Food Science and Technology*, 1995, **1**:97-104.

[6] Peterson DG, Pearson WR, Stack SM. Characterization of the tomato (Lycopersicon esculentum) genome using in vitro and in situ DNA reassociation. *Genome*, 1998, **41**:346–356.

[7] Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 1991, **9**:208–218.

[8] McCormick S, Niedermeyer J, Fry J, Barnason A, Horsch R, Fraley R. Leaf disc transformation of cultivated tomato (L. esculentum) using Agrobacterium tumefaciens. *Plant Cell Reports*, 1986, **5**:81–84.

[9] Fillatti JJ, Kiser J, Rose R, Comai L. Efficient transfer of glyphosate tolerance gene into tomato using binary Agrobacterium tumefaciens vector. *BioTechnology*, 1987, **5**:726–730.

[10] Zagorska NA, Shtereva A, Dimitrov BD, Kruleva MM. Induced anderogenesis in tomato (Lycopersicon esculentum Mill.). Influence of genotype on androgenetic ability. *Plant Cell Reports*, 1998, **17:**968–973.

[11] Menda N, Semel Y, Peled D, Eshed Y, Zamir D. In silico screening of a saturated mutation library of tomato. *Plant Journal*, 2004, **38**:861–872.

[12] (http://tgrc.ucdavis.edu; http://www.sng.cornell.edu)
[13] Bonnema G, Hontelez J, Verkerk R, Zhang YQ, van Daelen R, van Kammen A, Zabel P. An improved method for partially digesting plant megabase DNA suitable for YAC cloning: application to the construction of a 5.5 genome equivalent YAC library of tomato. *Plant Journal*, 1996, **9**:125–133.
[14] Budiman MA, Mao L, Wood TC, Wing RA. A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Research*, 2000, **10**:129–136.
[15] Patil RS, Davery MR, Power JB, Cocking EC. Effective protocols for *Agrobacterium*-mediated leaf disc trasformation in tomato (*Lycopersicon esculentum* Mill.). *Indian Journal of Biotechnology*, 2002, **1**:339–343.
[16] Breuning G, Lyons JM. The case of the FLAVR SAVR tomato. *California Agriculture*, 2000, **54**:6–7.
[17] Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science*, 1993, 262:1432–1436.
[18] Martin GB, Vicente MC, Tanksley SD. High resolution linkage analysis and physical chraracterization of the *Pto* bacterial resistance locus in tomato. *Molecular Plant-Microbiology Interactions*, 1993, **6**:26–34.
[19] Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD. The SOL genomics network. A comparative resource for Solanaceae biology and beyond. *Plant Physiology*, 2005, **138**:1310–1317.
[20] Lemaire-Chamley M, Petit J, Garcia V, Just D, Baldet P, Germain V, Fagard M, Mouassite M, Cheniclet C, Rothan C. Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. *Plant Physiology*, 2005, **139**:750–769.
[21] Rocco M, D'Ambrosio C, Arena S, Faurobert M, Scaloni A, Marra M. Proteomic analysis of tomato fruits from two ecotypes during ripening. *Proteomics*, 2006, **6**:3781–3791.
[22] M Faurobert, C Mihr, N Bertin, T Pawlowski, Luc Negroni, N Sommerer, M Causse. Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiology*, 2007, **143**:1327–1346.
[23] Sheoran IS, Olson DJH, Ross ARS, Sawhney VK. Proteome analysis of embryo and endosperm from germinating tomato seeds. *Proteomics*, 2005, **5**:3752–3764.
[24] Sheoran IS, Ross ARS, Olson DJH, Sawhney VK. Proteomic analysis of tomato (Lycopersicon esculentum) pollen. *Journal of Experimental Botany*, 2007, **58**:3525–3535.
[25] Sheoran IS, Ross ARS, Olson DJH, Sawhney VK. Differential expression of proteins in the wild type and *7B-1* male-sterile mutant anthers of tomato (Solanum lycopersicum): a proteomic analysis. *Journal of Proteomics*, 2009, **71**:624-636.
[26] Polenta GA, Calvete JJ, Gonzalez CB. Isolation and characterization of the main small heat shock proteins induced in tomato pericarp by thermal treatment. *Federation of European Biochemical Societies Journal*, 2007, **274**:6447-6455.
[27] Amini F, Ehsanpour AA, Hoang QT, Shin JS. Protein pattern changes in tomato under in vitro salt stress. *Russian Journal of Plant Physiology*, 2007, **54**:464–471.

[28] Fish T, Thannhauser TW. Salt-induced and salt-suppressed proteins in tomato leaves. *Journal of the American Society for Horticultural Science*, 2009, **134**:289-294
[29] Zhou S, Sauve R, Thannhause TW. Proteome changes induced by aluminium stress in tomato roots. *Journal of Experimental Botany*, 2009, **60**:1849–1857.
[30] Li J, Wu XD, Hao ST, Wang XJ, Ling HQ. Proteomic response to iron deficiency in tomato root. *Proteomics*, 2008, **8**:2299–2311.
[31] Oguri S, Kamoshida M, Nagata Y, Momonoki YS, Kamimura H. Characterization and sequence of tomato 2S seed albumin: a storage protein with sequence similarities to the fruit lectin. *Planta*, 2003, **216**:976–984.
[32] Casado-Vela J, Sellés S, Bru Martínez R. Proteomic analysis of tobacco mosaic virus-infected tomato (Lycopersicon esculentum M.) fruits and detection of viral coat protein. *Proteomics*, 2006, **6**:S196–S206.
[33] Houterman PM, Speijer D, Dekker HL, De Koster CG, Cornelissen BJC, Rep M. The mixed xylem sap proteome of Fusarium oxysporum-infected tomato plants. *Molecular Plant Pathology*, 2007, **8**:215–221.
[34] Wang Y, van der Hoeven R S, Nielsen R, Mueller LA, Tanksley SD. Characteristics of the tomato nuclear genome as determined by sequencing undermethylated EcoRI digested fragments. *Theoretical and Applied Genetics*, 2005, **112**:72–84.
[35] van der Hoeven RS, Ronning C, Giovannoni JJ, Martin G, Tanksley SD. Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell*, 2002, **14**:1441-1456.
[36] Pavan S, van Heusden AW, Yuling B. Solanum lycopersicum (Tomato). In: Encyclopedia of Life Sciences (ELS). *John Wiley & Sons*, 2009.
[37] Rick CM. Tomato, Lycopersicon esculentum (Solanaceae). In: Evolution of Crop Plants. *N. W. Simmonds*, 1976.
[38] Miller JC, Tanksley SD. RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theoretical and Applied Genetics*, 1990, **80**:437–448.
[39] Rick CM, Fobes JF. Allozyme variation in the cultivated tomato and closely related species. *Bulletin of the Torrey Botanical Club*, 1975, **102**:376–384.
[40] Stevens MA, Rick CM, Genetics and breeding. In: The tomato crop: a scientific basis for improvement. *Chapman and Hall*, 1986.
[41] Krogh V, Freudenheim JL, D'Amicis A, Scaccini C, Sette S, Ferro-Luzzi A, Trevisan M. Food sources of nutrients of the diet of elderly Italians: II. Micronutrients. *International Journal of Epidemiology*, 1993, **22**:869–877.
[42] Martinez-Valverde I, Periago MJ, Provan G, Chesson A. Phenolic compounds, lycopene and antioxidant activity in commercial varieties of tomato (Lycopersicum esculentum). *Journal of the Science of Food and Agricultural*, 2002, **82**:323–330.
[43] Beecher GR. Nutrient content of tomatoes and tomato products. *Proceedings of the Society for Experimental Biology and Medicine,* 1998, **218**:98–100.
[44] Stahl W, Sies H. Lycopene: a biologically important carotenoid for humans? *Archives of Biochemistry and Biophysics*, 1996, **336**:1-9.
[45] Matsushima NR, Shidoji Y, Nishiwaki S, Yamada T, Moriwaki H, Muto Y. Suppression by carotenoids of microcystin-induced morphological changes in mouse hepatocytes. *Lipids*, 1995, **30**:1029-1034.

[46] Levy J, Bosin E, Feldmen B, Giat Y, Miinster A, Danilenko M. Lycopene is a more potent inhibitor of human cancer cell proliferation than either α- carotene or β-carotene. *Nutrition and Cancer*, 1995, **24**:257-266.
[47] Fuhramn B, Elis A, Aviram M. Hypocholesterolemic effect of lycopene and β-carotene is related to suppression of cholesterol synthesis and augmentation of LDL receptor activity in macrophage. *Biochemical and Biophysical Research Communications,* 1997, **233**:658-662.
[48] Agarwal S, Rao AV. Tomato lycopene and low density lipoprotein oxidation: a human dietary intervention study. *Lipids*, 1998, **33**:981-984.
[49] Rao AV, Agarwal S. Bioavailability and in vivo antioxidant properties of lycopene from tomato products and their possible role in the prevention of cancer. *Nutrition and Cancer,* 1998, **31**:199-203.
[50] Rao AV, Agarwal S. Effect of diet and smoking on serum lycopene and lipid peroxidation. *Nutrition Research*, 1998, **18**:713-721.
[51] Agarwal S, Rao AV. Tomato lycopene and its role in human health and chronic diseases. *Canadian Medical Association Journal*, 2000, **163**:739-744.
[52] Cook NC, Samman S. Review: Flavonoids—chemistry, metabolism, cardioprotective effects, and dietary sources. *Journal of Nutritional Biochemistry*, 1996, **7**:66–76.
[53] Davies JN, Hobson GE. The constituents of tomato fruit– The influence of environment, nutrition and genotype. *Critical Reviews in Food Science* & *Nutrition,* 1981, **15**:205–280.
[54] Rivas N, Luh BS. Polyphenolic compounds in canned tomato pastes. *Journal of Science,* 1968, **33**:358–363.
[55] Slimestad R, Fossen T, Verheul MJ. The flavonoids of tomatoes. *Journal of Agricultural and Food Chemistry*, 2008, **56**:2436–2441.
[56] Dall'Asta C, Falavigna C, Galaverna G, Sforza S, Dossena A, Marchelli R. A multiresidual method for the simultaneous determination of the main glycoalkaloids and flavonoids in fresh and processed tomato (Solanum lycopersicum L.) by LC-DAD-MS/MS. *Journal of Separation Science*, 2009, **32**:3664–3671.
[57] Hunt GM, Baker EA. Phenolic constituents of tomato fruit cuticles. *Phytochemistry* 1980, **19**:1415–1419.
[58] Krause M, Galensa R. Determination of naringenin and naringenin-chalcone in tomato skins by reversed phase HPLC after solid-phase extraction. *European Food Research and Technology,* 1992, **194**:29–32.
[59] Moco S, Bino RJVorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoot J, de Vos CHR. A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiology*, 2006, **141**:1205–1218.
[60] Petersen A, Vieths S, Aulepp H, Schlaak Mand Becker WM. Ubiquitous structures responsible for IgE cross-reactivity between tomato fruit and grass pollen allergens. *Journal of Allergy and Clinical Immunology*, 1996, **98**:805–815.
[61] Ortolani C, Ispano M, Pastorello EA, Ansaloni R and Magri GC. Comparison of results of skin prick tests (with fresh foods and commercial food extracts) and RAST in 100 patients with oral allergy syndrome. *Journal of Allergy and Clinical Immunology*, 1989, **83**:683–690.
[62] Larramendi C, Ferrer A, Huertas AJ, Garcìa-Abujeta JL, Andreu C, Tella R, Cerdà MT, Bartra J, Lavìn JR, Pagàn JA, Lòpez-Matas MA, Fernàndez-Caldas E, Carnès J. Sensitization to tomato peel and pulp extracts in the Mediterranean Coast of Spain:

prevalence and co-sensitization with aeroallergens. *Clinical and Experimental Allergy*, 2007, **38**:169–177.

[63] Asero R, Mistrello G, Roncarolo D, Amato S, Arcidiacono R, Fortunato D. Detection of a novel allergen in raw tomato. *Journal of Investigational Allergology and Clinical Immunology*, 2008, **18**:397-400.

[64] Asero R, Mistrello G, Amato S. Anaphylaxis caused by tomato lipid transfer protein. *European Annals of Allergy and Clinical Immunology*, 2011, **43**:125-126.

[65] Bässler OY, Weiss J, Wienkoop S, Lehmann K, Scheler C, Dölle S, Schwarz D, Franken P, George E, Worm M, Weckwerth W. Evidence for novel tomato seed allergens: IgE-reactive legumin and vicilin proteins identified by multidimensional protein fractionation–mass spectrometry and in silico epitope modeling. *Journal of Proteome Research*, 2009, **8**:1111–1122.

[66] Pravettoni V, Primavesi L, Piantanida M, Brenna OV, Farioli L, Scibilia J, Mascheri A, Pastorello EA. Tomato industrial derivatives: Maillard reaction and residual allergenicity. *Clinical and Translational Allergy*, 2011, **1**:P19.

Chapter II - SECTION I

Proteomics in food analysis

1. The proteomic approach

Most foods are produced from living organisms and tissues and, for this reason, they bear all the complexity of the biological system from which they derive.[1]
In particular, proteins and the proteolysis products, like large and small peptides and free amino acids, are the major constituents of many foods and play a prominent role in determining their nutritional, rheological, sensorial, functional and biological properties. Belonging to extremely complex biochemical systems, food proteins are found in elaborate mixtures, made up of components which underwent to many biological processes. Besides, food technological processes, such as cheese-making or pasteurization, baking or fermentation, and food storing and ripening, contribute to increase the complexity of these systems, due to the chemicals and enzymatic reactions on proteins that may occur.
As a consequence, the proteome, that is the totality of proteins present in a biological sample (cell, organelle, tissue or food matrix) is a dynamic system, well defined in a certain moment and conditions, but changing along time, according not only to the genes, but also on external factors. The field of proteome analysis, also called proteomics, aimed at describing the proteome of a biological sample, including foods, due to the complexity of the systems, is faced by several analytical challenges.
As described above, a food matrix may contain up to tens of thousands of different proteins, requiring highly efficient protein separation and identification methods and also bio-informatic tools that allow the handling of the large amount of data accumulating during the analysis. Furthermore, food proteomes usually differ in the different expression levels of certain proteins, rather than in their presence or absence. This means that highly reliable and precise methods are necessary for the quantification of thousands of proteins in food samples. Finally, protein separation and identification must be linked to structural and functional information on proteins of interest.

1.1 Protein extraction

Virtually, every proteomic workflow study includes the extraction of the protein components from the food of interest. Depending on the characteristics of the starting material and on the protocol used, different patterns of protein could be obtained, even from the same food matrix. In particular, plant tissue have a relatively low protein concentration, if compared with animal and bacterial tissues. Besides, due to their significant amount of interfering substances, such as pigments, carbohydrates, polyphenols, polysaccharides and starch, the protein extract composition could be negatively influenced and, consequently, the following steps of separation and analysis may be strongly affected.[2] Moreover, the protein profiles depends on many factors, such as environmental conditions, ripening stage and postharvest handling. So, no definitive and comprehensive protein extraction protocol is actually available and suitable for a large variety of food matrix sources; instead, specific protocols should be refined according to the starting material.

1.2 Protein separation

The protein component of the extract should be separated before characterization. Nowadays, two complementary proteomic methodologies are available and most widely applied in proteins analysis: separation on gel and separation by chromatography.[3] These approaches are complementary, since they focus on specific subsets of proteins that are only partially overlapping. They differ in the way how proteins or peptides are isolated, separated, and detected.

1.2.1 Gel-based technique

The gel-based techniques underpin on gel electrophoresis, based on the movement of charged molecules in an electric field. The generally used support medium is cellulose or thin gels made up of either polyacrylamide or agarose. Cellulose is used as support medium for low molecular weight compounds, such as amino acids, whereas agarose and polyacrylamide gels are widely used for larger molecules like proteins. The general electrophoretic techniques can not be used to measure the molecular weight of the proteins because the mobility of a substance in the gel is influenced by both charge and size. In order to overcome this limit, the biological samples are treated in a way to give a uniform charge to proteins, so that electrophoretic mobility then depends primarily on size. The molecular weight of protein may be estimated if they are subjected to electrophoresis in the presence of sodium dodecyl sulfate (SDS) and a reducing agent (usually mercaptoethanol). SDS disrupts the secondary, tertiary and quaternary structure of the protein to produce a linear polypeptide chain coated with negatively charged SDS molecules, yielding denatured proteins having the same density of charge. In this way in SDS-PAGE (or 1D gel electrophoresis) the proteins get separated in accordance to their size (Fig 1).

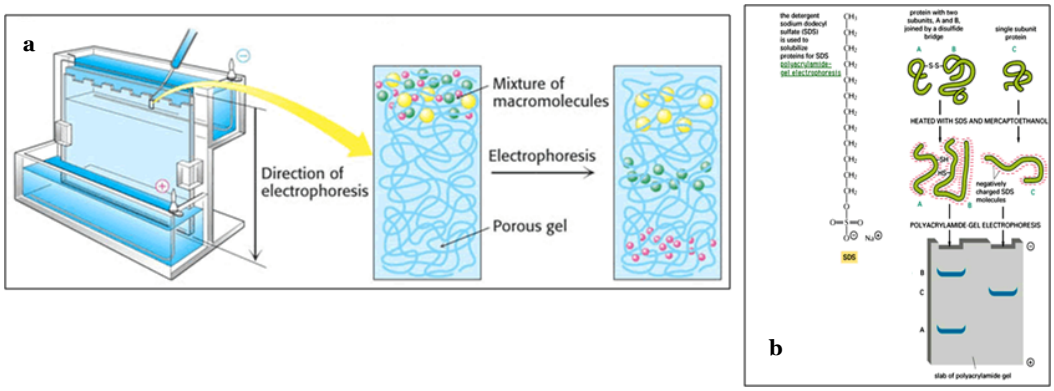


Fig 1. a, schematic representation of SDS PAGE system; **b**, mechanism of action of denaturant agent on proteins before SDS PAGE.

SDS-PAGE is a very easy and widespread technique used to obtain a fast separation of a protein mixture.[4] Furthermore, 1D gel electrophoresis offers the advantage that virtually all proteins are soluble in SDS and also extremely acidic and basic proteins are easily visualized.

In general, cells contain more proteins than genes. Moreover, nobody knows any precise figures or the extent of post translational modifications (PTMs) that may confer additional complexity to the protein composition. Considering that SDS PAGE

is a low resolution technique (less than 100 proteins may be resolved), highly resolving and sensitive identification techniques must be applied to enable differentiation of all these protein species.

In this scenario, two-dimensional (2D) gel electrophoresis technique is the most powerful method available for this purpose, since it allows to resolve up to 10000 proteins in the optimal conditions.[5] It consists in first separating sample proteins by isoelectric focusing, according to their isoeletric point by means of gel strip containing an immobilized pH gradient (IPG). When an electric field is applied, proteins migrate in the IPG until they assume a net charge of zero, that is as they reach a pH point equal to their isoeletric point. After their immobilization, proteins are further separated according to their apparent molecular weight by SDS-PAGE (Fig 2). Proteins spots are then detected by Coomassie Brilliant Blue, silver, fluorescence dyes, or radio labeling. The staining also allows to identify qualitative differences in the protein patterns of quite simple samples visually, whereas the more complex samples require informatics tools.

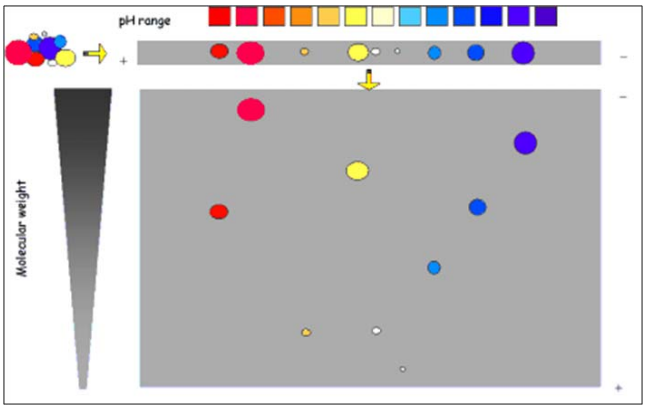


Fig 2. Principles of 2D electrophoresis. In the first dimension isoelectric focusing (IEF) the proteins are separated according to their isoelectric points (*pI*). In the second dimension, proteins are separated according to their molecular weights (*MW*) in a SDS-polyacrylamide gel.

Although the protein separation by 2D gel electrophoresis is a potential powerful tool for this kind of investigation, we can find in the available literature few studies focused on this topic. This may be due to several factors and, at least partially, to the variable reproducibility of the 2D gel electrophoresis protein profiles. Actually, there are still various aspects of the technique that need to be improved, including sample preparation, the resolution and detection of protein components, since the results are still dependent on the extract quality and the researcher's skill.[6] Among the major bottlenecks is the protein solubility of the extracts, which is very often affected by the presence of a wide variety of non-protein substances, which may also interfere with the resolution of the separation and/or the visualization of the results. Sample preparation is a critical step in 2D gel electrophoresis, especially when the extract is derived from a plant source.[7] Another well-known disadvantage of the standard 2D gel technique is the loss of very hydrophobic proteins and proteins with extreme pI values. Besides, highly-abundant proteins, which are often not interesting for the proteome studies, may spoil the analysis of the less abundant ones, which are of high importance.

1.2.2 Chromatography-based techniques

Liquid chromatography (LC) is a very dynamic methodology that, depending on the type of chromatography selected, can separate the protein components as a function of properties such as the molecular mass, pI, hydrophobicity and capacity to bind a ligand. LC allows good separations of acidic and basic, low- and high-molecular-weight proteins in a very short analysis time. By choosing different systems, either the separation of high amounts of proteins (preparative chromatography) or high resolution using chromatographic systems working at high pressure (fast-protein LC [FPLC] and high-pressure LC [HPLC,]) can be obtained.[8] Furthermore, a better resolution in the separation of a protein mixture can be obtained by combining different types of chromatography (Multidimensional LC [MDLC]), in which the sample is applied to a first column and eluted onto a second column allowing different stationary phases and therefore different separation mechanisms to be combined and facilitating greater resolution than a single HPLC column alone (Fig 3). Various combinations of separation methods include size-exclusion chromatography (SEC) with reversed-phase LC (RPLC), RPLC with capillary electrophoresis (CE), strong cation-exchange chromatography (SCX) with RPLC, SCX with avidin affinity chromatography (AC) to select specifically biotinylated peptides, and isoelectric focusing (IEF) with RPLC.[9] Recent advances in MultiID include peptide separation by ultra-high-pressure LC (UHPLC) and anion-and-cation mixed-bed ion exchange techniques.[10]

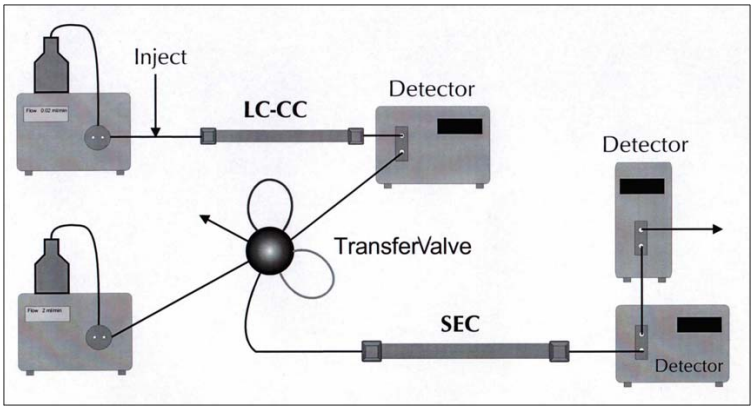


Fig 3. Schematic representation of an automated 2D chromatographic system.

The protein profiles are generally very reproducible, thus allowing comparative studies of the proteome variations in tissues. Except for size-exclusion chromatography, generally, LC offers a high flexibility regarding the volume of the sample to be treated and usually does not require concentration/precipitation steps that are sometimes a cause of protein component loosening.

Gel-free approaches have the disadvantage that qualitative and quantitative information on protein isoforms and differential post-translational modifications are often lost.[11] Besides that, cross-species identification for poorly sequenced genomes, as is the case of most plant-based foods, is not possible.

1.3 Detection and identification of separated proteins

Whatever the separation methodology is, it will require at least partial information about amino acid sequence in order to identify a protein of interest. Sequence data can be obtained by mass spectrometry technologies or by Edman degradation.

1.3.1 MS technologies for proteome analysis

Recently mass spectrometry (MS) has emerged as an indispensable tool to analyze protein and peptide mixtures arising from their proteolysis. Among all the available techniques commonly used for the analysis of a full proteome, MS has incrementally improved especially because it provides specificity, speed and reliability of the analytical response and it also gives the chance to handle the complexity associated to the biological system. For these reasons the employment of MS in the study of food proteomes is rapidly increasing.

For a long time MS has been restricted just for the analysis of small and thermostable compounds. The introduction of new types of mass spectrometers, which allowed the formation of ions from formation of molecular ions of intact biomolecules, as electrospray ionization (ESI),[12] and matrix assisted laser desorption/ionization (MALDI),[13] made peptides and also polypeptides able to be analyzed by MS.

In MS technologies, ionization is fundamental, as the physics of mass spectrometry relies upon the molecule of interest being charged, resulting in the formation of either positive ions or negative ions. In this way a molecular ion species is formed and, depending on the ionization method, fragment ions may also be created. These ion species are then separated according to their mass-to-charge (m/z) ratio and the masses are assigned from the measurement of some physical parameter. Finally, the measurement of ion abundance, based on peak height or peak area, is made leading to a semi-quantitative or quantitative answer.

Ionization sources. As concerns ionization, electrospray ionization probably is the preferred ionization methods for proteomics.[14] Here the sample is dissolved in a solvent mixture (e.g., acetonitrile-water) and then injected into a capillary held at a potential of 3-4 kV. As a result, a very fine spray of solvent droplets containing ions are formed. Multiple charged gas-phase ions are subsequently formed during the desorption process due to the evaporation of the solvent,[15] which will then enter the mass analyzer (Fig 4).

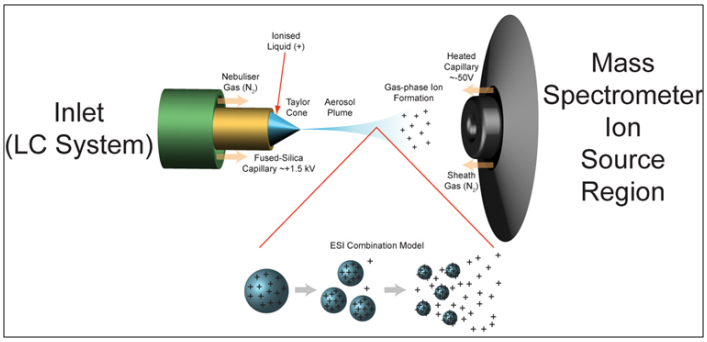


Fig 4. Electrospray ionization (ESI) and ion source overview.

For ESI, there are several ways to deliver the sample to the mass spectrometer. In the simplest method, electrospray sources have been connected in line with LC systems that automatically purify and deliver the sample to the mass spectrometer. Examples of this method are LC and reversed phase LC (RP-LC).

The second major ionization technique, matrix-assisted laser desorption ionization (MALDI), relies on a laser which is fired at a sample plate containing a dried mixture of matrix (α-cyano-4-hydroxycinammic acid) and sample. The matrix absorbs radiation from the laser resulting in excitation of the matrix molecules. As a result, a dense plume containing both the matrix and the analyte molecules is produced. The analyte molecules interact with protons from the matrix to form mainly single charged ions that enter the mass analyzer (Fig 5). The formed ions are separated in a mass analyzer according to their mass-to-charge ratio (m/z).[15]

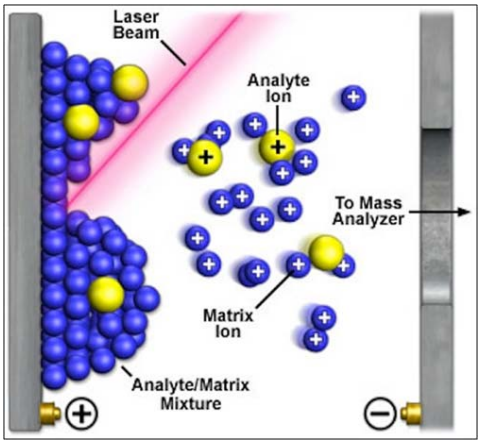


Fig 5. A schematic representation of MALDI source.

Analyzers. The analyzer is an instrument able to separate or differentiate introduced ions. Both positive and negative ions (as well as uncharged, neutral species) form in the ion source. However, only one polarity is recorded at a given moment. Basically, four types of mass analyzers have been deeply applied in the study of proteomes: quadrupole (Q), time-of-flight (TOF), ion trap, and Fourier-transform ion cyclotron resonance (FT-ICR), which strongly differ in both the physical principles of ion separation and the analytical performance.[16]

Quadrupole uses oscillating electrical fields to selectively stabilize or destabilize ions. By passing ions through a radio frequency quadrupole field, single mass/charge ratios can be measured. Only ions within a particular mass range, exhibiting oscillations of constant amplitude, could collect at the detector (Fig 6). Single quadrupole mass spectrometers require a clean matrix to avoid the interference of unwanted ions, and they exhibit very good sensitivity.

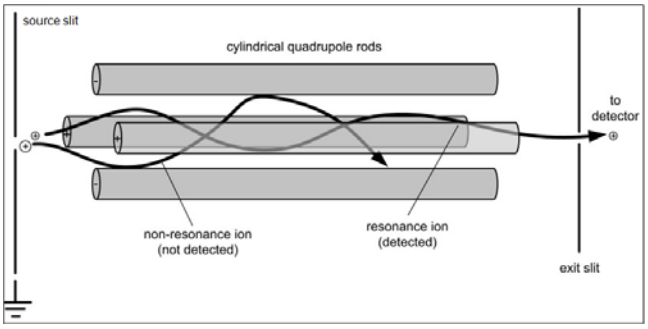


Fig 6. A schematic representation of the quadrupole analyzer.

In the time-of-flight analyzer, the ions enter a field-free drift range where they are not accelerated further and thus travel with a speed they have reached at the moment when passing the electrode. This speed, in turn, depends on the mass of the ions with heavier molecules having a higher moment of inertia and hence a lower velocity; ions with smaller *m/z* values will reach the detector first. For the resolution of the mass spectral analyses the length, L, of the field-free drift range is essential and in modern machines it measures about one meter. Further increase in resolution can be reached in the reflector mode: after having passed a distance in the drift range the ions enter another electromagnetic field and are accelerated in a nearly reversed direction towards an ion detector (Fig 7).

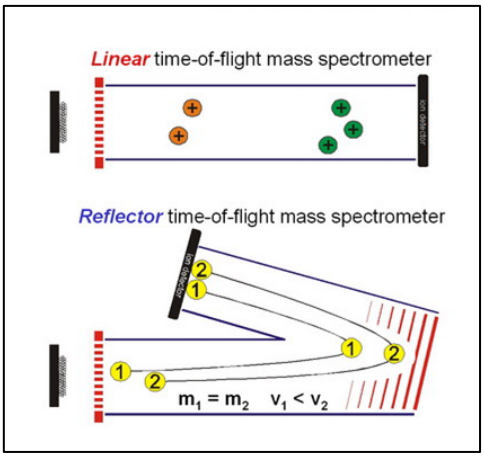


Fig 7. A schematic representation of the time-of-flight (TOF) analyzer.

The quadrupole ion trap,[17] is the three dimensional analogue of the linear quadrupole mass filter. In this device too, ions are subjected to forces applied by a radio frequency field but the forces occur in all three, instead of just two, dimensions. Stable motion of ions in the linear quadrupole allowed ions freedom of motion in one dimension. In the ion trap, stable motion allows no degrees of freedom. Hence, ions are trapped within the system. These features confer high sensitivity and high resolution to the system.

The first high-performance mass analyzer to employ ion trapping in an electrostatic field, the Orbitrap mass spectrometer, easily achieves ultra-high resolution (>100,000) with high mass accuracy (<1 ppm), a wide dynamic range (up to 5,000), fast scanning, and uncompromised sensitivity. Ions injected from the ion source are

trapped in the linear ion trap where ions of interest can be isolated and fragmented, scanned out and detected by an independent set of detectors. For high accuracy measurements, achieved with hybrid instruments such as linear trap quadrupole and Orbitrap (LTQ-Orbitrap), the ions are axially ejected from the linear trap into the C-Trap where they are captured again and 'cooled' by collisions with nitrogen gas. The ions are then squeezed into a smaller cloud within the C-Trap prior to injection into the Orbitrap. As ions enter the Orbitrap mass analyzer, the voltage on the central electrode increases, forcing the ion packets to circle around the electrode. The ions enter the Orbitrap slightly off axis and keep oscillating along the central electrode (left-right). The image current is recorded on the outer split electrodes. The signals are amplified and transformed into a frequency spectrum by fast Fourier Transformation and converted into a mass spectrum (Fig 8).[18]

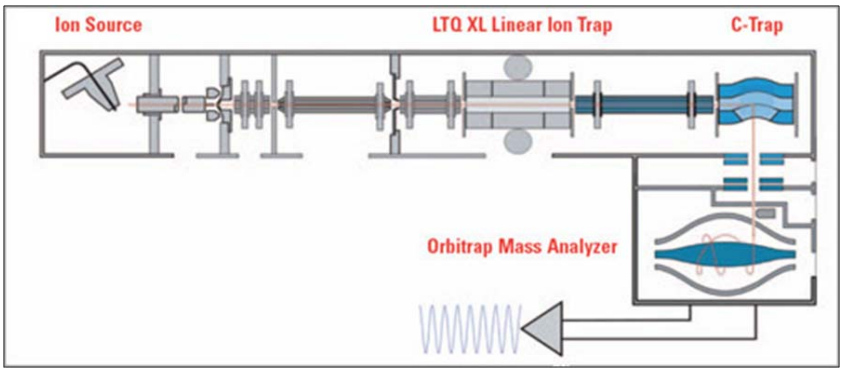


Fig 8. A schematic representation of the hybrid LTQ- Orbitrap analyzer.

FT-ICR (Fourier Transform Ion Cyclotron Resonance) mass spectrometers are useful, high-precision analyzers. Fourier transform mass spectrometry (FTMS) detects the image current produced by ions cyclotroning in the presence of a magnetic field. The ions are injected into a static electric/magnetic ion trap, where they form part of a circuit. Detectors in the system measure the electrical signal of ions that pass near them, producing a periodic signal. The frequency of the signal can be used determine the ion's mass/charge ratio. The FTMS is highly sensitive and boasts higher precision and resolution than other methods (Fig 9).[19]

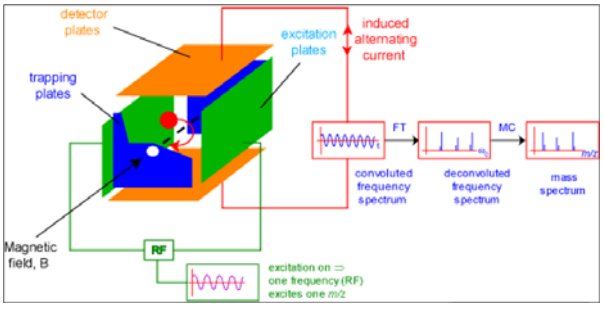


Fig 9. A Schematic of FT-ICR-MS showing the ion trapping, detection and signal generation.

In most cases, one or more forms of mass spectrometry, which utilize different methods of sample ionization, are used for protein identification. The first is MALDI-TOF mass spectrometry, used to perform peptide and protein mass fingerprinting; the

second is ESI tandem mass spectrometry (MS/MS), usually hyphenated to high performance liquid chromatography separation, used to perform peptide sequence elucidation and identification of the corresponding protein.[20]

1.3.2 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) adds another dimension to mass spectrometric measurement. Apart from improving the specificity of the technique, it can be used as a powerful structural elucidation tool.

Tandem MS instruments, such as the triple quadrupole (QqQ), TOF-TOF, and the hybrid quadrupole-time-of-flight (Q-TOF), are routinely applied in proteomics studies to generate peptide fragment ion spectra suitable for protein identification by sequence database searching or to determine additional structural features, as the site of attachment and type of post-translational modifications.

New configurations of ion sources and mass analyzers have found wide application for protein analysis. To allow the fragmentation of MALDI-generated precursor ions, MALDI ion sources have recently been coupled to quadrupole ion-trap mass spectrometers,[21] and to two types of TOF instruments. In the first, two TOF sections are separated by a collision cell ('TOF-TOF instrument'),[22] whereas in the second, the hybrid quadrupole-TOF instrument, the collision cell is placed between a quadrupole mass filter and a TOF analyzer.[23] Ions of a particular m/z are selected in a first mass analyzer (TOF or quadrupole), fragmented in a collision cell and the fragment ion masses are 'read out' by a TOF analyzer. These instruments have high sensitivity, resolution and mass accuracy, and the quadrupole-TOF instrument can be used interchangeably with an ESI ionization source. The resulting fragment ion spectra are often more extensive and informative than those generated in trapping instruments.

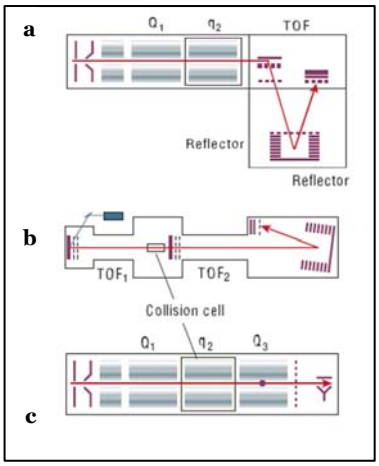


Fig 10. Examples of mass spectrometers used in proteome research. **a**, The quadrupole-TOF instrument combines the front part of a triple quadruple instrument with a reflector TOF section for measuring the mass of the ions; **b**, The TOF-TOF instrument incorporates a collision cell between two TOF sections. Ions of one mass-to-charge (m/z) ratio are selected in the first TOF section, fragmented in the collision cell, and the masses of the fragments are separated in the second TOF section. **c**, Quadrupole mass spectrometers select by time-varying electric fields between four rods, which permit a stable trajectory only for ions of a particular desired m/z. Again, ions of a particular m/z are selected in a first section (Q1), fragmented in a collision cell (q2), and the fragments separated in Q3. In the linear ion trap, ions are captured in a quadruple section, depicted by the red dot in Q3. They are then excited via resonant electric field and the fragments are scanned out, creating the tandem mass spectrum. Adapted from reference 24.

1.3.3 Bottom-up and top-down approaches

The most common approaches to identify proteins after their separation are divided in 'bottom-up proteomics' and 'top-down proteomics'.[25] The bottom-up proteomic strategy includes two different routes based on protein cleavage into smaller peptide fragments that allows protein identification. While one route to protein identification is based on intact mass measurements of a set of peptide-digestion products of the original protein, the other is based on MS fragmentation (MS^2) of one or more of these peptides.[26] The first approach leads to identification by peptide mass fingerprinting (PMF), since it creates a set of peptides that is unique for each protein and their masses are used as a fingerprint to identify the original protein.

Conversely, the second approach, based on the information obtained from dissociation of one or more peptides, is referred to as peptide-fragment fingerprint (PFF).[26] As in the bottom-up strategy the proteolysis of proteins occurs immediately after their isolation from the sample and a huge number of peptides are expected to be formed, it requires a high-throughput separation method (e.g., multi-dimensional chromatography). However, using the bottom-up strategy rarely achieves complete information on protein sequence.[26] The enhancement of methodologies, such as complex multi-stage instruments exploited in tandem mass spectrometry, has allowed the challenge of protein and proteome analysis to be tackled. For examples, tandem mass spectrometry is most used for peptide sequencing. Tandem mass spectrometry also provides detailed structural features of peptides that can be inferred from analysis of the resulting fragments and it is commonly used in product-ion mode to determine the amino-acid sequence of a specific peptide.

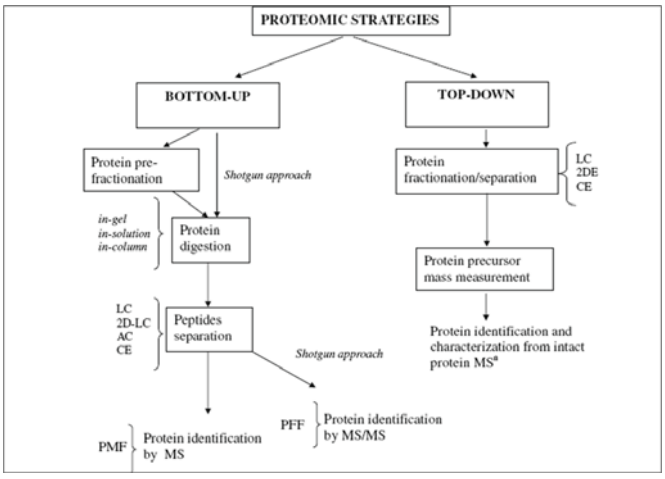


Fig 11. Strategies employed in proteomic studies for protein identification. Adapted from reference 27.

Top-down proteomics is an emerging MS strategy that involves gas-phase ionization of intact proteins and subsequent high-resolution mass measurement of the relevant ions obtained by direct fragmentation inside the mass spectrometer.[10] This approach directly analyzes intact proteins without prior digestion. Intact proteins are generally less effective for protein identification than peptide-level measurements, but they offer insights unattainable at the peptide level. MALDI with TOF analysis is currently employed in top-down proteomics for the mass measurement of intact proteins.

The accuracy of mass detection plays a fundamental role in protein identification and is more precise for small rather than large molecules. The limitation of mass detection

in protein identification is that any modification resulting in small changes of total protein mass can not be displayed, since it falls within the error of the method. As a consequence, modified and unmodified proteins are unlikely to be distinguished.[27]
The key objective in top-down analysis is efficient fragmentation of intact proteins to obtain sequence specific information for database searching and protein identification. This can also be achieved by using FT-ICR due to the high-resolution mass measurement and the ability to fragment intact proteins.[10] Besides, tandem mass spectrometry has also proved suitable for a top-down approach on instruments such as LTQ-Orbitrap that features characteristics, such as large ion capacity and large dynamic range, and high mass accuracy and high resolution.[10] Although top-down proteomics is a potential powerful tool for protein analysis, its analytical throughput and efficiency for large-scale proteome analysis are still limited.
A limit that embraces both the proteomics approaches mediated by mass spectrometry is that at least an homologue sequence should be already present in the database, otherwise the protein identification might be impossible.

1.3.4 Sequencing by Edman degradation

Sequence data of a protein molecule can be obtained by *de novo* sequencing using Edman degradation method. It allows to identify amino acids residues and their order starting from the N-terminal residue.[28] Usually the elucidation of the amino acid sequence of the N-terminal region is sufficient to perform a homology search in databases and identify a protein. Furthermore, this is an easy and fast procedure. Although the use of Edman sequencing is waning in the field of proteomics, it is a very useful tool for several reasons. Edman sequencing of relatively abundant proteins is a viable alternative to MS, if a mass spectrometer is in high demand for the identification of low-copy proteins or is not available. Besides, Edman sequencing is used to obtain the N-terminal sequence of a protein (if possible) to determine its true start. Indeed, one of the biggest limit in Edman sequencing is N-terminal modification of proteins. When the protein has a blocked N-terminal residue, this method is not useful and it becomes necessary to perform preliminary unblocking treatments,[29] or fragmentation of the molecule followed by the isolation of an internal fragment to be sequenced. Moreover, a quite consistent amount of protein is required for Edman sequencing, whereas MS can operate at picomolar level.
Actually, since both mass spectrometry and Edman degradation have advantages and disadvantages, the best strategy for the identification of a protein can be based on a clever combination of these methodologies. The unequivocal identification of a protein can be easily obtained by combining data concerning the amino acid sequence of a protein fragment, such as the sequence of the N-terminal region obtained by the Edman degradation method, and the exact molecular mass of the entire molecule estimated by mass spectrometry. Whenever the homology search indicates that neither the sequence obtained by Edman degradation or sequence information gained by mass spectrometry are present in databases, nor homologs of that sequence are detected, then a new protein has been discovered.

2. *Food safety and proteomics*

Through their food choice decisions and consumption behavior, consumers may be exposed to a number of potential food hazards.[30]

The risk of food borne disease is substantially heightened by biological and chemical contamination of areas where food is produced, processed and consumed. Potential undesirable residues in foods span a broad range, from natural (e.g. mycotoxins) and environmental contaminants (e.g. dioxins) to agro-chemicals (e.g. nitrates and pesticides), veterinary drugs, growth promoters, packaging components, and many more. Microbiological considerations are an even greater challenge to safety of food because potentially harmful microorganisms have the ability either to grow rapidly from very low numbers in food or to proliferate in the human body after ingestion.[31]

Being the fundamental components of foods and due to the advances in proteome analysis, growing interest is arising around proteins for the evaluation of food safety.

To date, the use concerns within food safety have concentrated on two main areas, the detection of micro-organisms, which may cause food spoilage or be hazardous to human health, and the safety evaluation of food components.

In the first case, controls are directed towards the detection of food borne micro-organisms. Traditional means of controlling microbial spoilage and safety hazards in foods include freezing, blanching, sterilization, curing and use of preservatives. However, the developing consumer trend for 'naturalness', as indicated by the strong growth in sales of organic and chilled food products, has resulted in a move towards milder food preservation techniques. This raises new challenges for the food industry.[1] Proteomic approaches have been directed to the development of methods for bacterial profiling through MS-based techniques, in order to distinguish among different species and, in some cases, among strains. Through these profiling methods, it was possible a fast and sensitive detection of pathogens or spoilage micro-organisms affecting food quality and safety during processing and storage.[32,33] The analysis of pathogenic micro-organism deserves particular caution, as the risks associated to their contamination are not limited to their living presence and capacity of infectivity, but they can generally release protein/peptide toxins able to survive for long time even in foods after bacterial cell contamination has been removed. These toxins, being heat-stable and resistant to proteases, can be a danger for the consumer health. For the exploration of virulence factors expressed in the secreted proteome fraction, in a very recent study different *Staphylococcus aureus* strains were analyzed using gel-based bottom–up proteomic approach.[34]

The second main issue, safety evaluation of food components is concerned with the presence of toxic compounds which may be originally present in the raw material (and therefore need to be eliminated by the manufacturing processes,) or, conversely, may be generated during the production process. Proteomic strategies based on MALDI-TOF-MS and LC-MS/MS methodologies have therefore carried out for detecting leguminous lectins,[35] and acrylamide as one of the products of the Maillard reaction in consequence of cooking processes,[36] respectively.

General public have become interested and often critical with regard to certain ways of producing food— both at the farm level and at the processing level. As a result, many concerns have risen on the use of genetically modified organisms (GMOs) in food production. The safety assessment of GM plants and derived food and feed

follows a “comparative” approach, i.e. the biomolecular expression pattern of food and feed is compared with their non-GM counterparts in order to predict intended and unintended effects. One of the pitfalls in the safety assessment of GM foods is the concept of “substantial equivalence” formulated by the Organisation for Economic Co-operation and Development in 1993, based on the idea that existing foods could serve as a baseline for comparing the properties of a GM food with its conventional counterpart. To identify possible unintended effects due to the use of GM crops, targeted analysis of specific compounds, which represent the key of important metabolic pathways in the plant like macro- and micro-nutrients, known allergens, anti-nutrients and toxins, has to be carried out in parallel with the comparative phenotypic analysis of the GM plant and of its near isogenic counterpart.

Proteome analysis was performed, for example, with virus-resistant tomatoes which were compared to wild-type lines through 2D gel electrophoresis and MALDI-TOF analysis of in-gel digested peptides.[37] In another study, the proteome of insect-resistant transgenic maize seeds, expressing an endotoxin of *Bacillus thuringiensis*, was compared to the proteome of an isogenic control grown under the same conditions, by using a nanoflow-HPLC system coupled to an high capacity ion trap.[38]

Proteome analysis can be also applied for a systematic search and evaluation of marker proteins, thus largely accelerating the development of assays to detect adulteration which can be a danger for the consumer health. For example, a very simple bottom-up approach identified marker proteins which are selectively expressed in tissues of the bovine brain and not in the bovine muscle, allowing the detection of proteins of the central nervous system in meats products used for human nutrition.[3]

Proteomic science has also started to provide an important contribution to the identification of protein involved in food allergy and to their mechanism of activation of toxicity.

2.1 Food allergy

Food allergies are within the more general concept of ‘adverse reaction to food’ which includes any clinically abnormal response induced by a food component, generally a protein, or a food additive.

According to Cianferoni, adverse food reactions are a broad term representing any abnormal clinical responses associated with ingestion of a food and they are further classified as food intolerance or food allergy.[39]

Food intolerance refers to an adverse physiologic response to a food and it may be due to inherent properties of the food (i.e. toxic contaminant, pharmacologic active component) or to the host (i.e. metabolic disorders, idiosyncratic responses, psychological disorder), it may not be reproducible, and it is often dose dependent.[39]

Food allergy refers to an abnormal immunologic response to a food that occurs in a susceptible host. This reaction is reproducible each time the food is ingested and it is often not dose dependent.

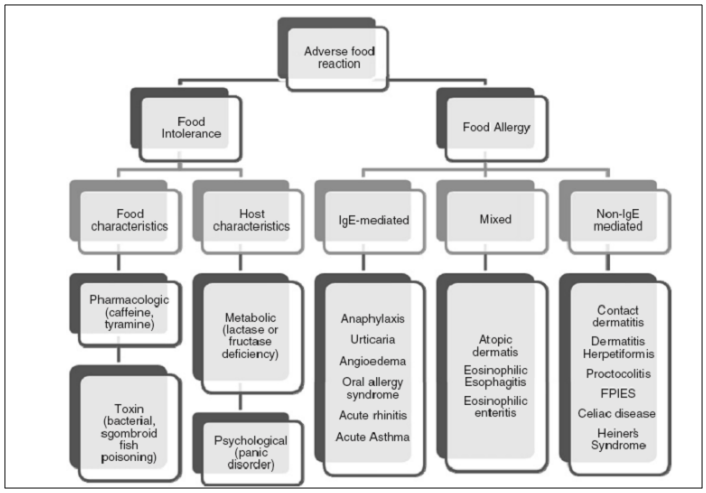


Fig 12. Classification and nomenclature of adverse reaction to food. Adapted from reference 39.

Based on the immunological mechanism involved, food allergies may be further classified in: IgE-mediated reactions (also called "type I"), which imply the production of specific immunoglobulines E (IgE or antibodies) formed by the patients, and are the best-characterized food allergy reactions; non IgE-mediated reactions, when the cell component of the immune system is responsible of the food allergy and mostly involve the gastrointestinal tract; mixed IgE mediated-cell mediated when both IgE and immune cells are involved in the reactions (Fig 12).[39]

The IgE-mediated reactions are characterized by symptoms which upraise on a time-scale ranging from seconds to 2-3 hours. They may involve one or more target organs, including the skin, the gastrointestinal and respiratory tracts, and the cardiovascular system. Cutaneous reactions are often involved in allergic reactions to food, like urticaria, pruritic erythema and dermatitis. Oral allergy syndrome (OAS) is a type of contact urticaria confined to the lips and oropharyngeal mucosa. Symptoms generally appear within 5 to 15 minutes following the food ingestion and consist of pruritus of the lips, tongue, palate, ears and throat. Due to the high prevalence of pollinosis in the adult population, and its frequent association with plant food allergies, OAS is the most frequent clinical presentation of food allergy seen in adult patients.[40] Food allergic reactions in the gastrointestinal tract may induce symptoms such as nausea, vomiting, abdominal pain and diarrhea and may also affect the respiratory tracts with rhinoconjunctivitis, bronchospasm, laryngeal oedema, rhinitis and asthma. Anaphylaxis is the severest manifestation of food allergy. It is a generalized allergic reaction caused by the massive release of mast cell mediators that may involve multiple organ systems and it could develop all the symptoms described above.[40]

Unlike IgE-mediated food allergy, non-IgE-mediated food allergy is rarely life-threatening. Symptoms of non-IgE-mediated food allergy are generally vague and, in most cases, are limited to gastrointestinal discomfort. Nevertheless, some studies reported non-IgE-mediated food allergy as cause of pathologies in infants.[41]

Concerning the epidemiology, recent estimations settled around 3.2–4% the prevalence of confirmed food allergies which arise during the first year of life,[42] although according to some authors prevalence could reach 6–7.5%.[43] By the way, an accurate evaluation is prevented by differences in study design, diversification of the diagnostic criteria and genetic and nutritional differences among population. The

epidemiologic assessment is further complicated by the self-perception of food reactions, which many studies are based on and which contributes to overestimate the prevalence of food allergies. Whatever the percentage of food allergy is, it has been showed that the prevalence of food allergy is higher in patients with atopic diseases.[44] Atopy is, in fact, an important risk factor in those food allergies arising in early infancy: food allergy seems to start the 'allergic march' from atopic dermatitis to allergic rhinitis and asthma.

The foods most frequently involved in allergic reactions are proteins from cow's milk, hen's egg, peanut and tree nuts, fish, shellfish, soya, fruits and legumes. The relative importance of these foods varies widely with the age of the patients and the geographical location.[40] Moreover, most food allergies with onset in the adult age are linked to inhalant allergies, and they develop as a consequence of an IgE sensitisation to the aeroallergen, which cross-reacts with the food in question (the so called "type II"), as a consequence of shared epitopes (the part of an antigen that is recognized by the cell of the immune system) in their primary and tertiary structures. The foods linked to pollen and latex allergies are of plant origin, mainly fresh fruits, tree nuts and vegetables; for this reason plant foods are the most prevalent food allergens in the adult population.[45]

2.1.1 Plant food allergens

The most natural classification system of plant food allergens might well be based on both structural and functional properties of proteins. Proteins are clustered together into families if they have residue identities of 30% or greater or if they have lower sequence identities but their functions and structures are very similar. Families whose members have low sequence identities but whose structures and functional features suggest a probable common evolutionary origin are placed together in superfamilies.[46] Most plant food allergens belong to only a few protein families and superfamilies, indicating that conserved structures and biological activities play a role in determining or promoting allergenic properties.[47]

Strikingly, only three dominating plant food allergen protein families/superfamilies were identified: the prolamin superfamily, the cupin superfamily and the pathogenesis-related proteins.[48]

The prolamin superfamily is named after the cereal prolamins, which are characterized by their high contents of proline and glutamine. In addition to the cereal prolamins, this group includes proteins rich in cysteine, with similar three-dimensional folds that are rich in α-helices, and are stable to thermal processing and proteolysis, like seed storage proteins, non-specific lipid transfer proteins (nsLTPs) and cereal seed inhibitors of α-amylase, trypsin.[47] The members of this protein family are capable to sensitize susceptible atopic individuals through ingestion or inhalation.

The cupins are a functionally diverse superfamily of proteins that share two short conserved consensus sequence motifs and a β-barrel structural core domain. The cupin superfamily comprises the major globulin storage proteins mainly from legumes and nuts. On the basis of their sedimentation coefficient, the globulins can be divided into the 7S vicilin-type globulins and the 11S legumin-type globulins.[48]

Pathogenesis-related proteins (PR) are not a protein superfamily, but represent a collection of unrelated protein families that function as part of the plant defense

system. Several studies have shown that many plant food allergens are homologous to proteins of the PR families.[49,50] The most interesting protein showing an homology to family 10 of the PR family is the major birch pollen allergen Bet v 1. The overall high levels of conserved surface residues between the members of the Bet v 1 family play an important role in conservation of IgE-binding epitopes and underlie the fruit–vegetable-pollen cross-reactive syndromes.[51]

The distinct biochemical properties associated with these allergenic proteins include: the abundance of the protein in the food; the sort of IgE-binding epitopes; the resistance of the protein to digestion and processing; the allergen structure. These features contribute mainly in enabling food proteins to become allergens or might affect their allergenicity.

One explanation for why abundant food proteins become allergens can be explained by the idea that the immune system is more likely to encounter these proteins rather than those representing only a small percentage of the total protein ingested; for example, seeds and nuts contain storage proteins which may account for 50% or more of the total proteins. However, there are exceptions to this rule: nsLTPs are potent food plant allergens but not very abundant. Although abundance might not be a universal characteristic of all food allergens, it seems to be a predisposing factor that, when coupled with other biochemical characteristics, could produce a food allergen.[52]

To be immunogenic, a protein must contain epitopes, that is short sequences which are recognized by cells of immune system. Two categories of IgE-binding epitopes, linear and conformational, are generally considered: conformational epitopes, when both the second and the tertiary structures are fundamental for IgE binding; linear epitopes, which require the primary amino acid sequence for IgE to bind. Whereas conformational IgE-binding epitopes are prevalent and important to the etiology of aeroallergen-mediated allergic reactions, linear epitopes are important to food allergens mainly because the immune system encounters them only after they have been partially denatured and digested by the human gastrointestinal (GI) tract. Therefore, the linear IgE-binding epitopes of food allergens have garnered more attention than the less prevalent conformational epitopes.[52]

Stability is thought to be a key attribute in determining protein immunogenicity via the GI tract and is directly related to the three-dimensional scaffold of a protein. Usually, a compact three-dimensional structure, ligand binding, disulphide bonds, and glycosylation contribute to protein stability.[52] In general, both intra- and inter-chain disulphide bonds constrain the three-dimensional scaffold such that perturbation of this structure by heat or chemical/enzymatic means is limited and frequently reversible. Important plant food allergens that have high numbers of disulphide bonds include members of the prolamin superfamily (nsLTPs, 2S albumins, cereal α-amylase/trypsin inhibitors) as well as of the pathogenesis-related proteins. *N*-glycosylation can have a significant stabilizing effect on protein structure, as shown for the 7S globulin of peas and its resistance to chemical denaturation.[53] The same behavior has been found in the presence of ligands, such as lipids binding to nsLTPs. In this case, the presence of fatty acids or phospholipid molecules in its lipid-binding pocket, increases the stability of the overall structure.[47] Bet v 1 and Bet v 1-homologous food allergens also have been shown to bind steroid ligands.[54]

Similarly, the ability of such proteins to bind lipids may flow in their propensity to aggregate as a result of food processing. So, the biologic activity of these proteins

could play a role in attenuating their allergenic potential, since such interactions may affect the uptake and the processing in the gastro-intestinal tract. On the other side, propensity of certain proteins to aggregate might affect their ability to sensitize by generally enhancing their immunogenicity. Both 7S and 11S globulins are highly thermostable. The cupin barrel seems to remain intact but the unfolding of other regions of the protein results in a loss of structure leading to formation of large aggregates as was examined in detail for soybean globulins.[55]

This description led to the conclusion that plant food proteins that are abundant and resistant to digestion and processing have the potential to become allergens. By the way, all the proteins with these features will not become allergens, since the development of an allergic reaction is a complex process involving a receptive immune system (that is, a genetic predisposition to propagation of an IgE response), the protein being presented with the correct inducers (that is, signals that elicit the immune system response), and a protein with the appropriate biochemical characteristics to survive the GI tract.[52]

3. *Analysis of food allergens*

The identification of the proteins that cause food allergies is important for clinical diagnosis and for the development of new food products. Furthermore, a detailed knowledge of the potential allergenicity is crucial when a food item is introduced into a new market. In order to protect sensitive consumers from undesirable allergic reactions, the European Commission (EC) introduced a directive which lists 14 common allergenic ingredients to be declared on the label of 'hazard food' (Directive 2007/68/EC). These are: cereals containing gluten and products thereof, eggs and products thereof, peanuts and products thereof, milk and products thereof (including lactose), nuts and products thereof, soybeans and products thereof, fish and products thereof, crustaceans and products thereof, mustard and products thereof, celery and products thereof, sesame seeds and products thereof, lupines and products thereof, mollusks and products thereof, sulfur dioxide and sulfites.[56]

Since legislation concerns only allergenic food ingredients introduced into food matrix intentionally, particular concern arises about the so called 'hidden allergens', allergenic contaminants occurring at any stage of food production accidentally, and thus not subjected to labelling. Being the hazard of hidden allergens a real risk for allergic consumers, the availability of sensitive methods is of supreme importance for the detection of traces of food allergens.

3.1 Methods for food allergen detection and quantification

The direct recognition of allergenic proteins generally relies on their antibody-binding properties and, thus, makes use of immunochemical techniques. Commercially available kits of this type are lateral flow devices or dipsticks, that are used for rapid screening, and enzyme-linked immunosorbent assays (ELISA), that also provide semi-quantitative determinations. In an ELISA assay the sample with an unknown amount of antigen is immobilized on a solid support. After the antigen is immobilized, serum contains immunoglobulins, which are able to bind to the allergen to be tested, is used. The principle of the quantification is based on the measurement of the

enzymatic activity of a second protein-specific antibody coupled to an enzyme. This second antibody binds to the allergen–primary antibody complex. A reaction with the enzyme substrate produces a coloured product, for which the absorption is proportional to the quantity of allergen in the food sample. The quantification can also be based on the measurement of the primary antibody bearing the enzyme label if any secondary antibody is used, as is the case of the direct ELISA. In the sandwich ELISA, immunocapture antibodies are coated and dried in microplate wells, and samples are then incubated in the wells. The target protein is thus bound in the well by the capture antibody, after which the unbound sample material is washed away. A second "detector" antibody is then added which binds to a different epitope on the immobilized target protein. This detector antibody may be also labeled with biotin, which provides an attachment point for some kinds of labels (Fig 13). The great advantages of ELISA techniques include the full automation of the test procedure by means of robots and the resulting extensive screening potential, the high sample throughput and the easy operation. A major drawback of immunological methods consists in the fact that the epitopes that are detected are usually not known and cross-reactivity with matrix components can result in false positives.

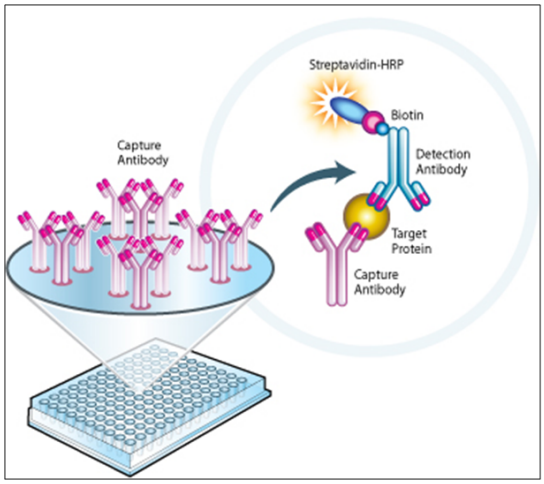


Fig 13. A schematic representation of Sandwich Elisa assay.

Other immunochemical assays of great importance are the radioimmunoassay (RIA), the immunofluorescence assay and the luminoimmunoassay (LIA), which can all be used in the determination of allergenic proteins. Some disadvantages however, such as the use of radioactive labels, have favored the use of the ELISA.

The standard method for the analysis of complex protein mixtures in laboratories dedicated to allergen determination is polyacrylamide gel electrophoresis (both one- or two- dimensional), followed by immunoblotting analysis (also known as Western blotting), in which separated proteins are adsorbed to a membrane that is then treated sequentially with an allergen-specific antibody solution and a secondary antibody conjugated to an enzyme or a radioactively labelled antibody. The enzyme–substrate reaction generates a colour product that precipitates on the membrane, indicating a positive reaction (Fig 14).[57]

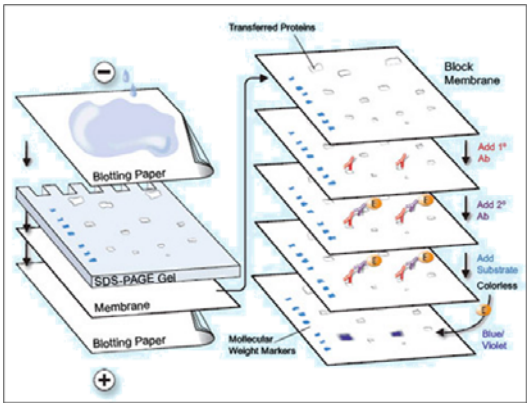


Fig 14. Illustration of Western blotting analysis step procedure.

An established strategy for relative allergenic protein quantification is based on 2D gel electrophoresis, which highlights changes in protein expression based on densitometric staining intensity, by comparison of gels from different sample states or varieties. However, gel-to-gel variability makes difficult the comparison and relative quantification of protein spots from different 2D gel electrophoresis experiments.

Although direct detection methods of allergenic proteins are usually the first choice for screening purposes, indirect methods, based on the detection of DNA sequences related to specific allergens, have also established their place in the determination of food allergens.

The PCR, a tool based on nucleic acids, has been developed for the indirect analysis of allergenic ingredients in food. It relies on the targeting of a segment of the gene coding for the allergenic protein of interest and amplifying only this DNA fragment to make the protein detectable. This tool is highly specific and sensitive, having a limit of detection (LOD) of less than 10 mg/kg for almond, hazelnut, soy, milk or peanut,[58] and it also can be coupled with ELISA or available as real-time PCR. Both approaches are gel-free since the amplified DNA fragments are hybridized to a protein probe and detected by ELISA, in the former, or are revealed as fluorescence proportional to the amount of the gene of interest, in the latter.

Recently, Peptide Nucleic Acids (PNAs), synthetic analogues of DNA, have been suited for hazelnut DNA detection, both in solution, in combination with HPLC,[59] and on platform, by using microarrays technology.[60]

The similarity between all the immunochemical methods is the use of biological sera and the fact that the detection is based on the antigen–antibody recognition. Thus, quantification depends on the quality of this recognition and might be distorted by several things, but mainly by the immunoglobulin specificity. As antibodies only recognize epitopes, rather than the whole molecule, the specificity of an antibody depends on the uniqueness of the epitope. A lack of specificity leads to false positives and negatives owing to cross-reactions between closely related proteins.[61]

PCR and real-time PCR methods are not based on the use of sera or cells, but the quantification remains indirect, like in the case of immunochemical methods. The presence of the target in food, a DNA fragment corresponding to the gene of a protein (the allergenic protein or a protein specific to the source species), does not necessarily prove the presence of the allergen itself, but indicates the source species in the case of

contamination. PCR methods, for example, are suitable for determining the origin (taxonomy) of the contaminating species.[61]
To overcome the limitations of the classically used methods, proteomic technologies have been increasingly used in the field of allergy.

3.1.1 Proteomics in food allergens characterization

The application of proteomics strategies to identify food allergens has been defined "allergenomics".[62]
As mentioned previously, since foods are complex matrices, protein pre-fractioning is required prior to MS analysis. (Fig 15, left panel). To identify allergens, electrophoresis-based separations, in either one- or two-dimensional version, coupled to MS analysis are usually integrated with antibody labelling in blotting setup, using sera from allergic patients as a source of specific IgE. Proteins, which are recognized by IgE of allergic patients, are usually identified using several MS platforms (MALDI-TOF, ESI-Q/TOF, ion-trap, LC–MS/MS). This approach has been successful in the identification of allergens from some of the most common injuring foods,[8,27,61,] and from less commonly offending foods such as, peach,[63] tomato,[64,65] and banana.[66]
Despite the wide application of 2D gel electrophoresis, time and labor requirements, low reproducibility, low sensitivity, low dynamic range for quantification, and poor detection of very hydrophobic and alkaline proteins still limit this technique. Moreover, denaturation that occurs during the electrophoresis may alter the allergenic properties of proteins, especially if they contain conformational epitopes, thus impairing the performance of the analysis. When it is the case, a native 2D gel electrophoresis may be applied for allergen discovery, like for the detection of allergens of hazelnut.[67]

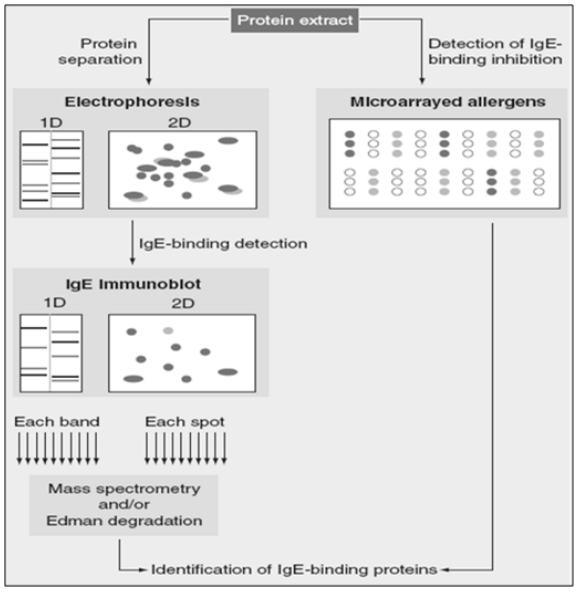


Fig 15. Identification of IgE-binding proteins. Classic (left side) and emerging (right side) strategies useful for protein/allergen identification in a protein extract. 1D and 2D indicate the protein separation by 1D and 2D electrophoresis, respectively. Adapted from reference 8.

An alternative method to electrophoresis is the liquid-phase separation coupled to MS analysis to identify proteins directly from complex mixtures, but the identification of intact protein allergens is affected by low sensitivity and is limited to low complex mixtures. The most effective LC/MS-based strategy is referred to as "shotgun" proteomics: in this case, the whole protein extract is digested by trypsin, thus generating a complex mixture of thousand peptides which is subsequently separated by LC prior to MS/MS sequencing. Although proteolysis increases the complexity of the mixture, the increasing of the peptides, derived from protein proteolysis, enhances the probability to detect and confidentially identify at least one of the constituent sequence. A similar strategy has been applied to monitor Ara h 1, Ara h 2, Ara h 3 as the major allergens in processed peanut. Some specific tryptic peptides were then selected as markers for monitoring selected peanut allergens in food products.[68]

In the shotgun approach, the information about the molecular weight of the intact proteins is lost. However, a careful analysis of MS/MS spectra can turn to be effective in discriminating isoallergen immunogenic proteins. Thus, using nano LC–ion-trap MS/MS, a screening method to specifically detect and differentiate celery allergens from IgE cross-reactive proteins of carrot and potato was developed.[69]

A decisive support to all of the methods applied in proteomic analysis is provided by bioinformatic tools for the interpretation of a huge amount of MS/MS spectra produced. The identification is performed by softwares such as MASCOT and SEQUEST in combination with constantly updated databases, such as SwissProt. With the aid of more specific software tools, the tissue localization of peach nsLTP (Pru p 3) was recently assessed by MALDI-MS imaging.[70]

Although MS is not a quantitative technique *per se*, it can be applied successfully for allergenic protein quantification, by using a quantification at protein level or at peptide level. In the direct quantification of intact proteins, according to a top-down approach, spraying directly intact proteins from solutions using electrospray yields MS spectra consisting of a series of peaks corresponding to charge state distributions of the protein. This technique, however, has strong limitations. The identification of targeted proteins in complex mixtures is hindered by two factors. The first one is the ion suppression that appears when different proteins are eluted at the same time. The second is the superposition of numerous peaks in the mass spectra, corresponding to different proteins that may not be resolved even using deconvolution algorithms.[61]

In the quantification at peptide level, according to a bottom-up approach, the final analyte is the peptide; therefore, a digestion step is necessarily in order to obtain the peptides to be analyzed. Quantification at the peptide level can be classified in methods involving stable isotopes, tagging by light (^{12}C) and heavy (^{13}C-labelled) tags (SILAC, ICAT, iTRAQ), and using isotopically labelled synthetic peptide to achieve, respectively, relative or absolute quantification. The first strategies incorporate isotopically labelled chemical moieties into the samples. They are useful to find biomarkers, to detect changes in protein abundances, for example, before and after the roasting process. The second one, used when the identity of the protein to be quantified is known in advance, uses a reference analyte, which is an isotopically labelled peptide. This reference peptide incorporates ^{13}C and ^{15}N stable isotopes on one of its amino acids, leading to a known mass difference with the endogenous peptide.[61]

At present, the 'classical' proteomic procedures for allergen detection may be made easier and faster by novel methodologies, such as the microarrays-based technologies (Fig 3, right panel). Microarray technology is directed to in vitro diagnosis. With an immunoassay format, this technology allows the simultaneous analysis of a large number of IgE antibodies with only a small amount of serum sample. Allergen microarrays are built by immobilizing multiple allergens onto a modified glass slide in an arrayed fashion. The incubation of microarrays with the samples gives rise to the binding of allergen-specific IgE from the samples with its corresponding allergen in the array. The assessment of reactivity is achieved with fluorescence, chemiluminescence or visible or UV absorbance.[71]

This methodologies may be also applied for the simultaneous detection of allergens in a complex mixture, such as a total protein extract from an allergenic source. It is possible to perform inhibition tests by pre-incubating the sera of allergic subjects with the protein extract. IgE antibodies of the sera will bind the allergenic proteins present in the extract, thus becoming unavailable for the interaction with the microarrayed molecules for which the reactivity with the sera IgE had already been established. Therefore, IgE binding to the proteins in solution is detected by measurement of decreased binding to the protein(s) on the solid phase.[8] The major disadvantage of this approach relies on the lack of some purified allergens to include on the platform, thus hiding the presence of proteins never described as allergens but present in offending foods.

Both the procedures shown in Fig 15 require the further biochemical and immunological characterizations, which imply the availability of purified molecules. This can be achieved by LC and the affinity chromatography is the best choice when specific antibodies are available. Then, the characterization of the molecules properties may be performed with the contribution of mass spectrometry, Edman degradation and molecular biology to gain information about post-translational modifications and the presence of isoforms, to elucidate the primary structure and to study the antigenic epitopes. When the structural, immunological and clinical characterization has been completed, the use of the allergenic molecule in diagnosis and, sometimes, in immunotherapy can be evaluated.[8]

4. References

[1] Mamone G, Picariello G, Caira S, Addeo F, Ferranti P. Analysis of food proteins and peptides by mass spectrometry-based techniques. *Journal of Chromatography A*, 2009, **1216**:7130–7142.

[2] Saravanan RS, Rose JK. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics*, 2004, **4**:2522–2532.

[3] Pischetsrieder M, Baeuerlein R. Proteome research in food science. *Chemical Society Reviews,* 2009, **38**:2600–2608.

[4] Laemmli U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 1970, **227**:680–685.

[5] Wittmann-Liebold B, Graack HR, Pohl T. Two-dimensional gel electrophoresis as tool for proteomics studies in combination with protein identification by mass spectrometry. *Proteomics,* 2006, **6**:4688–4703.

[6] Biron DG, Brun C, Lefevre T, Lebarbenchon C, Loxdale HD, Chevenet F, Brizard JP, Thomas F. The pitfalls of proteomics experiments without the correct use of bioinformatics tools. *Proteomics,* 2006, **6**:5577–5596.

[7] Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B. Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics,* 2005, **5**:2497–2507.

[8] Mari A, Ciardiello M, Tamburrini M, Rasi C, Palazzo P. Proteomic analysis in the identification of allergenic molecules. *Expert Review of Proteomics*, 2010, **7**:723–734.

[9] Ye M, Jiang X, Feng S, Tian R, Zou H. Advances in chromatographic techniques and methods in shotgun proteome analysis. *Trends in Analytical Chemistry*, 2007, **26**:80-84.

[10] Han X, Aslanian A, Yates III JR. Mass spectrometry for proteomics. *Current Opinion in Chemical Biology,* 2008, **12**:483-490.

[11] Carpentier S, Panis B, Vertommen A, Swennen R, Seargent K, Renaut J, Laukens K, Witters E, Samyn B, Devreese B. Proteome analysis of non-model plants: a challenging but powerful approach. *Mass Spectrometry Reviews*, 2008, **27**:354–377.

[12] Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 1989, **246**:64-71.

[13] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, 1988, **60**:2299–2301.

[14] Griffiths WJ, Jonsson AP, Liu S, Rai DK, Wang Y. Electrospray and tandem mass spectrometry in biochemistry. *Journal of Biochemistry*, 2001, **355**:545–561.

[15] Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews Molecular Cell Biology,* 2004, **5**:699–711.

[16] Newton RP, Brenton AG, Smith CJ, Dudley E. Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. P*hytochemistry*, 2004, **65**:1449–1485.

[17] March RE, Hughes RJ. Quadrupole Storage Mass Spectrometry. *Wiley*, 1989.

[18] http://planetorbitrap.com/technology.php

[19] http://www.ehow.com/list_6155669_types-mass-spectrometers.html

[20] Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chemical Reviews*, 2001, **101**:269–295.

[21] Krutchinsky AN, Kalkum M, Chait BT. Automatic identification of proteins with a MALDI quadrupole ion trap mass spectrometer. *Analytical Chemistry,* 2001, **73**:5066–5077.

[22] Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical Chemistry*, 2000, **72**:552–558.

[23] Loboda AV, Krutchinsky AN, Bromirski M, Ens W, Standing KG. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Communications in Mass Spectrometry,* 2000, **14**:1047–1057.

[24] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*, 2003, **422**:198-207.

[25] Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annual Review of Biomedical Engineering*, 2009, **11**:49–79.
[26] Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrometry Reviews,* 2005, **24**:168–200.
[27] Monaci L, Visconti A. Mass spectrometry-based proteomics methods for analysis of food allergens. *Trends in Analytical Chemistry*, 2009, **28**:581-591.
[28] Edman P. A method for the determination of the amino acid sequence of peptides. *Archives of Biochemistry and Biophysics,* 1949, **22**:475–483.
[29] Ciardiello MA, D'Avino R, Amoresano A. Tuppo L, Carpentieri A, Carratore V, Tamburrini M, Giovane A, Pucci P, Camardella L. The peculiar structural features of kiwi fruit pectin methylesterase: amino acid sequence, oligosaccharides structure, and modeling of the interaction with its natural proteinaceous inhibitor. *Proteins,* 2008, **71**:195–206.
[30] Miles S, Brennan M, Kuznesof S, Ness M, Ritson C, Frewer LJ. Public worry about specific food safety issues. *British Food Journal*, 2004, **106**:9-22.
[31] Wilcocky A, Pun M, Khanonax J, Aung M. Consumer attitudes, knowledge and behaviour: a review of food safety issues. *Trends in Food Science & Technology,* 2004, **15**:56–66.
[32] van Baar BLM. Characterization of bacteria by matrix-assisted laser desorption/ionization and electrospray mass spectrometry. *FEMS Microbiology Reviews,* 2000, **24**:193-219.
[33] Kawano Y, Ito Y, Yamakawa Y, Yamashimo T, Horii T, Hasegawa T, Ohta M. Rapid isolation and identification of staphylococcal exoproteins by reverse phase capillary high performance liquid chromatography–electrospray ionization mass spectrometry. *FEMS Microbiology Reviews*, 2000, **189**:103-108.
[34] Pocsfàlvi G, Cacace G, Cuccurullo M, Serluca G, Sorrentino A, Schlosser G, Blaiotta G, Malorni A. Proteomic analysis of exoproteins expressed by enterotoxigenic Staphylococcus aureus strains. *Proteomics*, 2008, **8**:2462-2476.
[35] Nasi A, Picariello G, Ferranti P. Proteomic approaches to study structure, functions and toxicity of legume seeds lectins. Perspectives for the assessment of food quality and safety. *Journal of Proteomics*, 2009, **72**:527-538.
[36] Rosén J, Hellenäs KE. Analysis of acrylamide in cooked foods by liquid chromatography tandem mass spectrometry. *Analyst*, 2002, **127**:880–882.
[37] Corpillo D, Gardini G, Vaira AM, Basso M, Aime S, Accotto GP, Fasano M. Proteomics as a tool to improve investigation of substantial equivalence in genetically modified organisms: the case of a virus-resistant tomato. *Proteomics*, 2004, **4**:193–200.
[38] Zolla L, Rinalducci S, Antonioli P, Righetti PG. Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. *Journal of Proteome Research*, 2008, **7**:1850–1861.
[39] Cianferoni A, Spergel JM. Food allergy: review, classification and diagnosis. *Allergology International*, 2009, **58**:457-466.
[40] Fernández-Rivas M, Ballmer-Weber B. Food allergy: current diagnosis and management. In: Managing allergens in food. *Woodhead Publishing Limited*, 2007.

[41] Caubet JC, Nowak-Wegrzyn A. Current understanding of the immune mechanisms of food protein-induced enterocolitis syndrome. *Expert Review of Clinical Immunology*, 2011, **7**:317-327.
[42] Venter C, Pareira B, Grundy J. Incidence of parentally reported and clinically diagnosed food hypersensitivity in the first year of life. *Journal of Allergy and Clinical Immunology*, 2006, **117**:1118-1124.
[43] Sicherer SH, Sampson HA. Food allergy. *Journal of Allergy and Clinical Immunology*, 2006, **117**:470-475.
[44] Sicherer SH, Sampson HA. Food hypersensitivity and atopic dermatitis: pathophysiology, epidemiology, diagnosis, and management. *Journal of Allergy and Clinical Immunology*, 1999, **104**:114–122.
[45] Osterballe M, Hansen TK, Mortz CG, Host A, Bindslev-Jensen C. The prevalence of food hypersensitivity in an unselected population of children and adults. *Pediatric Allergy and Immunology*, 2005, **16**:567–573.
[46] Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of Clinical Investigation*, 1995, **247**:536-540.
[47] Breiteneder H, Mills ENC. Plant food allergens—structural and functional aspects of allergenicity. *Biotechnology Advances*, 2005, **23**:395–399.
[48] Breiteneder H, Radauer C. A classification of plant food allergens. *Journal of Allergy and Clinical Immunology*, 2004, **113**:821-830.
[49] Mills ENC, Madsen C, Shewryx PR, Wichers HJ. Food allergens of plant origin—their molecular and evolutionary relationships. *Trends in Food Science & Technology*, 2003, **14**:145-156.
[50] Breiteneder H, Ebner C. Molecular and biochemical classification of plant-derived food allergens. *Journal of Allerg and Clinical Immunology*, 2000, **106**:27-36.
[51] Jenkins JA, Griffith-Jones S, Shewry PR, Breiteneder H, Mills ENC. Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. *Journal of Allergy and Clinical Immunology*, 2005, **115**:164–171.
[52] Bannon GA. What makes a food protein an allergen? *Current Allergy and Asthma Reports*, 2004, **4**:43–46.
[53] Pedrosa C, De Felice FG, Trisciuzzi C, Ferreira ST. Selective neoglycosylation increases the structural stability of vicilin, the 7S storage globulin from pea seeds. *Archives of Biochemistry and Biophysics*, 2000, **382**:203–210
[54] Marković-Housley Z, Degano M, Lamba D, von Roepenack-Lahaye E, Clemens S, Susani M, Ferreira F, Scheiner O, Breiteneder H. Crystal Structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. *Journal of Molecular Biology*, 2003, **325**:123-133.
[55] Wolf WJ, Nelsen TC. Partial purification and characterization of the 15S globulin of soybeans, a dimer of glycinin. *Journal of Agricultural and Food Chemistry*, 1996, **44**:785–791.
[56] COMMISSION DIRECTIVE 2007/68/EC of 27 November 2007 amending Annex IIIa to Directive 2000/13/EC of the European Parliament and of the Council as regards certain food ingredients (2007) OYcial Journal of the European Union. http://eur-lex.europa.eu

[57] Baumgartner S, Krska R, Welzig E. Detecting allergens in foods. In: Managing allergens in food. *Woodhead Publishing Limited*, 2007.

[58] Poms RE, Anklam E, Kuhn M. Polymerase chain reaction techniques for food allergen detection. *Journal of AOAC International*, 2004, **87**:1391–1397.

[59] Germini A, Scaravelli E, Lesignoli F, Sforza S, Corradini R, Marchelli R. Polymerase chain reaction coupled with peptide nucleic acid high-performance liquid chromatography for the sensitive detection of traces of potentially allergenic in foodstuffs. *European Food Research and Technology*, 2005, **220**:619–624.

[60] Rossi S, Scaravelli E, Germini A, Corradini R, Fogher C, Marchelli R. A PNA-array platform for the detection of hidden allergens in foodstuffs. *European Food Research and Technology*, 2006, **223**:1-6.

[61] Kirsch S, Fourdrilis S, Dobson R, Scippo ML, Maghuin-Rogister G, De Pauw E. Quantitative methods for food allergens: a review. *Analytical and Bioanalytical Chemistry*, 2009, **395**:57–67.

[62] Akagawa M, Handoyo T, Ishii T, Kumazawa S, Morita N, Suyama K. Proteomic analysis of wheat flour allergens. *Journal of Agricultural and Food Chemistry*, 2007, **55**:6863-6870

[63] Palacín A, Tordesillas L, Gamboa P, Sanchez-Monge R, Cuesta-Herranz J, Sanz ML, Barber D, Salcedo G, Díaz-Perales A. Characterization of peach thaumatin-like proteins and their identification as major peach allergens. *Clinical & Experimental Allergy*, 2010, **40**:1422-1430.

[64] Bässler OY, Weiss J, Wienkoop S, Lehmann K, Scheler C, Dölle S, Schwarz D, Franken P, George E, Worm M, Weckwerth W. Evidence for novel tomato seed allergens: IgE-reactive legumin and vicilin proteins identified by multidimensional protein fractionation–mass spectrometry and in silico epitope modeling. *Journal of Proteome Research*, 2009, **8**:1111-1122.

[65] López-Matas MA, Larramendi CH, Ferrer A, Huertas AJ, Pagán JA, García-Abujeta JL, Bartra J, Andreu C, Lavín JR, Carnés J. Identification and quantification of tomato allergens: in vitro characterization of six different varieties. *Annals of Allergy, Asthma* and *Immunology,* 2011, **106**:230–238.

[66] Grob M, Reindl J, Vieths S, Wüthrich B, Ballmer-Weber BK. Heterogeneity of banana allergy: characterization of allergens in banana-allergic patients. *Annals of Allergy, Asthma & Immunology*, 2002, **89**:513-516.

[67] Hird H, Pumphrey R, Wilson P, Sunderland J, Reece P. Identification of peanut and hazelnut allergens by native two-dimensional gel electrophoresis. *Electrophoresis*, 2000, **21**:2678-2683.

[68] Chassaigne H, Nørgaard JV, Hengel AJ. Proteomics-based approach to detect and identify major allergens in processed peanuts by capillary LC-Q-TOF (MS/MS). *Journal of Agricultural and Food Chemistry*, 2007, **55**:4461-4473.

[69] Fæste CK, Jonscher KR, Sit L, Klawitter J, Løvberg KE, Moen LH. Differentiating cross-reacting allergens in the immunological analysis of celery (Apium graveolens) by mass spectrometry. *Journal of AOAC International*, 2010, **93**:451-461.

[70] Cavatorta V, Sforza S, Mastrobuoni G, Pieraccini G, Francese S, Moneti G, Dossena A, Pastorello EA, Marchelli R. Unambiguous characterization and tissue localization of Pru P 3 peach allergen by electrospray mass spectrometry and MALDI imaging. *Journal of Mass Spectrometry*, 2009, **44**:891-897.

[71] González-Buitrago JM, Ferreira L, Isidoro-García M, Sanz C, Lorente F, Dávila I. Proteomic approaches for identifying new allergens and diagnosing allergic diseases. *Clinica Chimica Acta*, 2007, **385**:21–27.

Chapter III

Assessing allergenicity of different tomato ecotypes by using pooled sera of allergic subjects: identification of the main allergens

1. Introduction

Food allergies have become a serious concern in Western countries, in which they affect about 6% of young children and 3%–4% of adults, and their prevalence appears to be on the rise.[1] On November 2007 a directive of the Commission of the European Communities established a list of food ingredients that have to be indicated on the labels, since they may cause adverse reactions in susceptible individuals.[2] Fruits, vegetables and their derivatives have generally not been included in this list as a possible cause of allergic reactions, with the exception of cereals, soybean, nuts, celery, mustard, sesame and lupines. Anyway, other allergenic foods of vegetal origin may also represent a serious threat for consumers' safety, as plant food allergies are spreading quickly, probably due to the cross reaction with other pollen allergens.[3,4,5]

Among cultivated crop plants, tomato (*Solanum lycopersicum* L.) is increasingly consumed due both to its beneficial effects on human health and to the spreading of fresh or processed cooked products. In spite of this increasing consumption (or maybe because of that) tomato allergy is also increasing and this food can be considered as an emerging allergen, which actually affects 1.5%-16% of food-allergic population,[6] and it has been shown to be related to other allergies, like allergy against grass pollen,[7] and latex[8] by cross-reactivity with homologous protein sequences. So far the International Union of Immunological Society (I.U.I.S) recognized four tomato allergens: *Lyc e 1* (tomato profilin),[6] *Lyc e 2* (tomato β-fructofuranosidase),[9] *Lyc e 3* (tomato nonspecific lipid-transfer protein),[10] and *Lyc e 4* (tomato intracellular pathogenesis-related protein) (http://www.allergen.org) but many other reactive proteins to tomato fruit extracts have been reported in the literature. Kondo and co-workers identified a polygalactunorase 2A (PG2A, 46 kDa), a pectinesterase (PE, 14 kDa), a β-fructofuranosidase (22 kDa) and a superoxide dismutase (18 kDa) as IgE-binding allergenic proteins in tomato fruit, by N-terminal amino acid sequencing.[11] During studies devoted to elucidate the correlation existing between cooked and fresh tomato assumption and symptoms, an allergen of about 9 kDa, heat-labile and pepsin-resistant was partially characterized by immunoblotting, pepsin digestion and heating.[12] Several studies were also carried out in order to analyze allergenic properties of the different parts of tomato fruits. Using a multidimensional protein fraction strategy and LC-MS/MS, a legumin (47 kDa) and vicilin (65 kDa) proteins were purified from tomato seeds, showing strong IgE reactivity in immunoblots.[13] Pravettoni and co-workers found different LTP allergenic isoforms in fresh tomato peel, pulp and seeds. Although conventional heat treatments employed during the production of tomato-based products usually degrade proteins, thus strongly reducing their IgE binding capacity,[14] industrial tomato derivatives have been demonstrated to still contain LTP.[15]

Rather than exclude tomato from the diet of allergic subjects, genetic engineering may be applied in order to obtain hypoallergenic tomato fruits. Gene-silencing approaches were successfully used to target *Lyc e 1* and *Lyc e 3* genes and low-allergenic tomato plants were obtained.[10,16] Although these strategies have proved to be very efficient in eliminating allergenic proteins, hypoallergenic genetically modified plants are difficult to produce and maintain as stable lines, since, as mentioned above, allergic subjects' IgE cross-react toward different tomato proteins, so several genes

corresponding to main tomato allergens should be silenced in transgenic lines. This could damage cell homeostasis because some tomato allergenic proteins play important physiological and structural roles.[17] Last but not least, genetically modified plants are still not accepted among European public opinion. Alternatively, the selection of genetic resources with low expression of allergenic proteins, represents a valuable tool both to understand the genetic base of the accumulation of allergenic proteins and to develop new hypoallergenic varieties.[18] The comprehensive study of the protein expression in a given species, often referred as proteomics, can also be applied to the study of the allergenic proteins. Although trends in food allergy research are increasingly focusing on mass spectrometry-based proteomics (sometimes referred as "allergenomics"),[19] examples of its application to an in-depth study of tomato allergens,[20] are still quite scarce in the literature.

2. *Aim of the work*

In this study using two pools of sera of allergic people coming from different Italian regions (Campania and Emilia Romagna), twelve tomato ecotypes were screened by a bottom-up proteomic approach, in order to identify the major allergens involved and to evaluate differences in IgE binding properties of these cultivars.

3. *Experimental part*

3.1 Plant material

Twelve tomatoes accessions were used for the analysis and they are summarized below.

Accession Name	Code	Final destination
Ventura D	VD (1)	Fresh markets/canning
Tondino D	TD (2)	Fresh markets/canning
Principe Borghese I	PB (3)	Fresh markets/canning
Sorrento Globoso Rosato	SGRI (4)	Fresh markets
San Marzano Murano	SMMU (5)	Fresh markets/canning
Pisanello	PS (6)	Fresh markets
Tondo Liscio	TLI (7)	Fresh markets
Tondino I	TI (8)	Fresh markets/canning
San Marzano	SMC (9)	Fresh markets/canning
Sorrento Rosato	STLR (10)	Fresh markets
Principe Borghese D	PBD (11)	Fresh markets/canning
San Marzano Morini	SMMO (12)	Fresh markets/canning

They were grown in a breeding farm in Sarno, Salerno (Southern Italy) over a spring–summer growing cycle. They were harvested during summer 2008 and frozen at –20 °C until the analysis time.

3.2 Protein Extraction

3.2.1 Chemicals

- Acetone (CH_3COCH_3) (Carlo Erba Reagents, Italy)
- Diethyl ether ($(C_2H_5)_2O$) (Carlo Erba Reagents, Italy)
- Sodium Hydroxide pellets (NaOH) (Carlo Erba, Italy)
- Chloridric acid (HCl) (Sigma-Aldrich, USA)
- Sodium chloride (NaCl) (Carlo Erba Reagents, Italy)
- Disodium hydrogen phosphate dodecahydrate ($Na_2HPO_4 \cdot 12\ H_2O$) (Carlo Erba Reagents, Italy)
- Potassium dihydrogen phosphate (KH_2PO_4) (Sigma-Aldrich, USA)
- Potassium chloride (KCl) (Sigma-Aldrich, USA)
- Milli Q H_2O obtained with Millipore Alpha Q system
- PBS (Phosphate Buffered Saline), preparation of 1 L solution
  - 5.5 g NaCl (94 mM)
  - 3.58 g $Na_2HPO_4 \cdot 12\ H_2O$ (10 mM)
  - 0.2 g KH_2PO_4 (1.5 mM)
  - 0.2 g KCl (2.7 mM)
  - Adjust the pH to 7.4 with HCl
  - Add Milli Q H_2O to 1 liter

3.2.2 Instrumentation

- Digital Scale BCE 62 PT (Orma, Italy)
- Homogenizer Ultraturrax T50 basic (IKA Werke, Germany)
- Rotary evaporator 111 (Buchi, Switzerland)
- pH meter 212 (Hanna Instruments, Italy)
- Reciprocating shaker SSL2 (Stuart Scientific, UK)
- Centrifuge Universal 320 R (Hettich, Germany)
- Qubit® fluorometer 1.0 (Invitrogen, UK)

3.2.3 Procedure

A total of 50 g of fresh tomato (including peels, pulps and seeds) was homogenized by means of Ultraturrax in cold acetone and the proteins were allowed to precipitate at -20 °C overnight. Pellets were washed twice with cold acetone and once with cold acetone/diethyl ether (1:1). Then the dry powder was extracted in a phosphate-buffered saline solution for 1 hour at 4 °C under continuous stirring, after the neutralization of the pH by a NaOH solution. After centrifugation (8334 g three times for 45'), the supernatant liquid was recovered and protein concentrations were determined using the Qubit® fluorometer according to the instructions by the manufacturer. The extracts were frozen at –20 °C until the analysis time.

3.3 Proteins analysis by electrophoresis

3.3.1 Chemicals

- 2-Iodoacetamide IAA (CH_2CONH_2) (Sigma-Aldrich, USA)
- 50% glycerol (Bio-Rad, Germany)

- Acetic acid (CH_3COOH), (solution 10% in Milli Q H_2O) (Carlo Erba, Italy)
- Agarose (Bio-Rad, Germany)
- Bromophenol blue (Bio-Rad, Germany)
- Carrier ampholytes (Bio-Rad, Germany)
- CHAPS (Bio-Rad, Germany)
- Coomassie Brilliant Blue R-250 (Bio-Rad, Germany)
- Criterion XT Precast Gel, 12% Bis-Tris, 12 wells (Bio-Rad, Germany)
- Criterion XT Precast Gel, 12% Bis-Tris, IPG+1 comb (Bio-Rad, Germany)
- Dithiothreitol DTT ($HSCH_2(CHOH)_2CH_2SH$) (Sigma-Aldrich, USA)
- Methanol (CH_3OH), HPLC grade, (Sigma-Aldrich, USA)
- Milli Q H_2O obtained with Millipore Alpha Q system
- Mineral oil (Bio-Rad, Germany)
- Ready Strip IPG strips, 11 cm, pH: 4-11 (Bio-Rad, Germany)
- Sodium Dodecyl Sulphate SDS, in powder and in solution 20% (Sigma-Aldrich, USA)
- SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad, Germany)
- Tris/HCl (50 mM solution in Milli Q H_2O) (Sigma Aldrich, USA)
- Urea (NH_2CONH_2) (Sigma-Aldrich, USA)
- XT MES Running buffer 20X (Bio-Rad, Germany)
- XT sample buffer 4X (Bio-Rad, Germany)
- XT Reducing Agent 20X (Bio-Rad, Germany)

3.3.2 Instrumentation

- Amicon® Ultra-4 Centrifugal Filter Units Amicon Ultra 5 KDa device — 5,000 NMWL (Nominal Molecular Weight Limit) (Millipore, Italy)
- Centrifuge Universal 320 R (Hettich, Germany)
- Rotor 1620A (97 mm radius) (Hettich, Germany)
- Adapters 1451 (Hettich, Germany)
- Centrifuge 1-13 (Sigma-Aldrich, USA)
- Criterion Cell (Bio-Rad, Germany)
- GS-800 calibrated densitometer (Bio-Rad, Germany)
- Glass Syringe 50 µL (Hamilton, USA)
- Gyro-Rocker SSL3 (Stuart Scientific, UK)
- PDquest software (Bio-Rad, Germany)
- Power supply: Power Pac Universal (Bio-Rad, Germany)
- PROTEAN IEF Cell (Bio-Rad, Germany)
- PCR Express Thermal Cycler (Thermo Hybaid, UK)

3.3.3 Procedures for 1D gel electrophoresis

i. Protein standard preparation

- 2 µL SDS-PAGE Molecular Weight Standards, Broad Range
- 28 µL of reducing sample buffer (prepared with 9.5 µL XT buffer 4X, 0.5 µL XT reduction buffer 20X and Milli Q H_2O to final 30 µL volume)
- 5’ incubation at 95 °C

ii. Sample preparation

- XT sample buffer 4X, diluted 4 times
- XT Reduction buffer 20X, diluted 20 times
- 30 µg of proteins (or 1 mg, in case of semipreparative analysis), following protein sample quantification described in paragraph 3.2.3
- Milli Q H_2O to final 25 µL volume (or 500 µL, in case of semi-preparative analysis)
- 5' incubation at 95 °C

iii. Electrophoretic run

For each wells of the Criterion XT Precast gel, 25 µL of sample prepared (or 450 µL, in case of semi-preparative analysis) were loaded or 5 µL of protein standard. The running buffer used was XT MES Running buffer diluted 20 times, the voltage applied to the Criterion Cell was 150 V. The run lasted 60'. In the sample buffer there was a 1% of Bromophenol blue, which ran as indicator.

3.3.4 Procedures for 2D gel electrophoresis

i. Desalting

In this case the procedure required a desalting step.

Pre-rinsing: the ultra-filtration membranes in Amicon Ultra devices contained trace amounts of Polyethylene *glycol* (PEG). The device must be pre-rinsed with 5 washes of a CH_3OH:H_2O (1:1) solution.

Preserving: the devices were stocked in a CH_3OH:H_2O (5:95) solution, at 4 °C until use.

Use: the ultra-filtration device was loaded with 4 mL of the solution to be desalted and centrifuged. The sample was loaded and centrifuged at 6500 RCF, 4 °C for 45', and then 3 washes with 4 mL of Milli Q H_2O were performed. The filtered solution was discarded each time and the retentate (upper part of the device) was recovered after washes using an Eppendorf pipette P100. A major recovery was obtained adding 250 µL of Milli Q H_2O, the membranes were washed using the pipette Pasteur and the recoveries were added to the suspension in an Eppendorf.

ii. First dimension: isoelectrofocusing

150 µg of proteins were re-dissolved in 185 µL of rehydration buffer, whose composition was: 8 M urea, 50 mM DTT, 4% CHAPS, 0.2% Carrier ampholytes, 0.0002% Bromophenol blue, Milli Q H_2O to 5 mL. Re-dissolved samples were placed in the IEF tray, and the strips were settled with the active part in contact with the sample. Passive rehydration of the strips was allowed for 1 hour, after that 3 mL of mineral oil were added to cover completely the strips. The rehydration was performed then overnight.

The strips were then placed in the focusing tray: +/- strip direction follow +/- sense of the tray and 2.5 mL of mineral oil were added to the strips, in order to prevent evaporation.

The PROTEAN IEF CELL was programmed as follow:

Volt: 8000

Volt/hours: 29000

Ramp: rapid
Max μA/strip: 50
At the end of the focusing, strips were taken and exceeding mineral oil was discarder on filter paper.

iii. *Strip equilibration*

Strips were placed, with gel sided up, in a tray and proteins were reduced and alkylated by adding 4 mL of DTT Equilibration buffer for 10' and, then, 4 mL of IAA Equilibration buffer for 10', respectively to each strip. Equilibration base buffer was composed of: 6 M Urea, 2% SDS, 0.05 M Tris/HCl buffer pH 8.8, 20% glycerol, Milli Q H_2O to 10 mL. 2% DTT Equilibration buffer was prepared immediately prior to use adding 200 mg of DTT. 2.5% IAA Equilibration buffer was prepared immediately prior to use by adding 250 mg of IAA.

iv. Second dimension: gel electrophoresis

Equilibrated strips were applied to the top of the Precast gels Criterion XT (12% Bis-Tris, IPG+1 comb). The strips were aligned so that the plastic back of the strip was against the back plate and the IPG strip was touching the top of the gel. "+" part of the strips was towards the well for the MW standards. MW standards were placed in the well (as described in paragraph 3.3.3). Molten agarose was added to fix the strips and, as it was set, the gels were covered with the XT MES Running buffer diluted 20 times. The Criterion Cell was set for the run as described previously.

3.3.5 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 Milli Q H_2O to 1000 mL. Gels were placed in glass containers and were covered with the staining solution where they were allowed to soak for at least 1 hour. Any dye that was not bound to protein diffused out of the gel during the de-staining steps, when the gels were rinsed with a solution of 10% CH_3COOH, 40% CH_3OH, 50% Milli Q H_2O, changed at least twice, to achieve the desired contrast.

3.4 Analysis of allergenic proteins by Western Blot

3.4.1 Patients sera

Human sera were provided by the Department of Medical Clinics, Nephrology, and Prevention Sciences of the University of Parma and by the Department of Clinical Medicine and Cardiovascular and Immunological Sciences of the University of Naples "Federico II". They were collected from patients with a clinical history of allergic reactions towards fresh tomato and/or tomato products. All subjects had positive skin prick test for tomato and specific IgE (>0.7 kU/L) detected at ImmunoCAP dosage.
After being tested individually, sera from each group were pooled together, in order to screen IgE-binding pattern of major tomato allergens.

3.4.2 Chemicals

- Glycine (Bio-Rad, Germany)
- Sodium Dodecyl Sulphate SDS, (Sigma-Aldrich, USA)

- Ovalbumin, from hen eggs white (Sigma-Aldrich, USA)
- Anti-Human IgE from rabbit (Bethyl laboratories Inc, USA);
- Goat anti rabbit-HRP (Bio-Rad, Germany)
- Opti-4-CN kit (Bio-Rad, Germany):
  - REAGENT Opti-4-CN
  - Opti-4-CN DILUENT
- Acetic acid (CH_3COOH), (Carlo Erba, Italy)
- PBS buffer (prepared as decribed in paragraph 3.2.1)
- Tween 20 (Sigma-Aldrich, USA)
- PBS-T (PBS buffer + 0.1% Tween 20)
- Blocking buffer (3% ovalbumin in PBS-T)
- Antibody dilution buffer (0.1% ovalbumin solution in PBS-T)
- Tris (25 mM solution in Milli Q H_2O) (Sigma-Aldrich, USA)
- Methanol (CH_3OH), HPLC grade, (Sigma-Aldrich, USA)
- Milli Q H_2O obtained with Millipore Alpha Q system

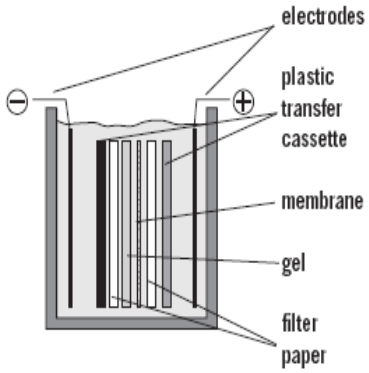
3.4.3 Instrumentation

- PVDF membrane 0.2 μm (Bio-Rad, Germany)
- Fiber pad (Bio-Rad, Germany)
- Filter paper (Bio-Rad, Germany)
- Gyro-Rocker SSL3 (Stuart Scientific, UK)
- Transblot electrophoretic transfer cell (Bio-Rad, Germany)
- Power supply: Power Pac Universal (Bio-Rad, Germany)
- GS-800 calibrated densitometer (Bio-Rad, Germany)

3.4.4 Procedure

After separation, for immunoblot analysis proteins were transferred onto 0.2-μm polyvinylidene fluoride (PVDF) membranes according with the manufacturer's instructions. Transfer buffer (25 mM Tris, 195 mM glycine, 10% CH_3OH, 0.01% SDS, pH 8.3) was prepared. The not-stained gel was placed into transfer buffer for 15', after cutting fringes of the wells of the gel and the residual Bromophenol blue run indicator. Cut pieces of the filter paper and of the PVDF membrane were wet for about 30 sec in CH_3OH on a rocker at room temperature, then were wetted 10' in Milli Q H_2O and finally 15' in the transfer buffer, together with fiber pad and filter papers. "Sandwich" for Bio-Rad's Transblot was assembled as follows:

Cathod (-) side of transblot
Fiber pad
Filter paper
Gel
PVDF membrane
Filter paper
Fiber pad
Anod (+) side of transblot



The closed sandwich was placed in the Transblot tank, which was filled with transfer buffer. Transfer occured for 1 h at 100 V in a refrigerated system (cooling coil).
After soaking in blocking buffer for 1 hour, the membranes were incubated overnight with individual serum sample or with pooled sera, diluted to 0.071 kU/L in antibody dilution buffer. The membranes were then incubated with rabbit anti-human antibodies, diluted 1:3000 in antibody dilution buffer, for 1 hour. After soaking twice in PBS-T in order to remove the excess of antibodies, they were incubated with goat anti-rabbit antibodies, diluted 1:3000 in antibody dilution buffer, linked to horseradish peroxidase enzyme (HRP) for 1 hour. The unbound antibodies were removed by rinsing membranes twice in PBS-T. Allergenic protein detection was achieved by incubation of the blotted membranes with the Opti-4-CN substrate, prepared by using 0.2 mL of reagent Opti-4-CN every 10 mL of Opti-4-CN diluents, for 30' till sufficient coloration was achieved. The membranes were washed with Milli Q H_2O for 15' and the images were acquired at the densitometer.

3.5 Trypsin in-gel digestion

3.5.1 Chemicals

- Ammonium bicarbonate (NH_4HCO_3) (25 mM and 100 mM solutions in Milli Q H_2O) (Sigma-Aldrich, USA)
- 2-Iodoacetamide IAA (CH_2CONH_2) (55 mM solution in 100 mM NH_4HCO_3) (Sigma-Aldrich, USA)
- Dithiothreitol DTT ($HSCH_2(CHOH)_2CH_2SH$) (Sigma-Aldrich, USA) (10mM solution in 100mM NH_4HCO_3)
- Methanol (CH_3OH), HPLC grade, (Sigma-Aldrich, USA)
- Acetic acid (CH_3COOH) (solution 1% in Milli Q H_2O) (Carlo Erba, Italy)
- Acetonitrile (CH_3CN) (Sigma-Aldrich, USA)
- Trypsin Proteomics Grade from Porcine Pancreas, Dimethylated (100 ng/µL solution in 1% acetic acid) (Sigma-Aldrich, USA)
- Milli Q H_2O obtained with Millipore Alpha Q system
- N_2-flux

3.5.2 Instrumentation

- Cutter
- Reciprocating shaker SSL2 (Stuart Scientific, UK)

- Hotplate Stirrer F 80 (Falc)

3.5.3 Procedure

With a cutter, protein bands or spots were excised from a stained polyacrylamide gel and placed into a 0.5 mL tube. Coomassie stained spots were destained as follow: 50 µL of 25 mM NH_4HCO_3/50% CH_3CN solution were added (all the volumes are intended to be enough to immerse the gel slides) and the tubes were placed on a reciprocating shaker overnight (16 h) on a low setting. Resulting gel particles were pretty clear, so it was possible to dehydrate gel pieces by adding CH_3CN until the gel slides were shrunk and opaque-white colored. CH_3CN was removed with Eppendorf pipette P100 by N_2-flux for 5-10'.
Proteins contained in gel slides were then reduced by addition of 10 mM DDT in 100 mM NH_4HCO_3 solution for 1 hour at 56 °C and alkylated by addition of 55 mM IAA in 100 mM NH_4HCO_3 solution for 45' at room temperature and in the dark. Then gel slides were swelled and shrunk twice by addition of 100 mM NH_4HCO_3 and 100% ACN, respectively, and then were dried under a nitrogen stream.
Trypsin stock solution (100 ng/µL in 1% acetic acid) was diluted 1:10 in 25 mM NH_4HCO_3 and added to the gel slides. After incubation at 37 °C for 16 hours the supernatants were recovered and the gel pieces were washed twice with 25 mM NH_4HCO_3/50% CH_3CN solution for 15' and once with acetonitrile and again the supernatants were recovered and dried under N_2-flux.

3.6 LTQ-Orbitrap analysis

3.6.1 Chemicals

- Acetonitrile (CH_3CN) (Sigma-Aldrich, USA)
- Methanol (CH_3OH) (Carlo Erba, Italy)
- Formic acid (HCOOH) 99% (Acros, Belgium)
- *Trifluoroacetic acid* 99% TFA (Acros, Belgium)
- Milli Q H_2O obtained with Millipore Alpha Q system
- Wetting solution: 100% CH_3CN
- Equilibration solution: 0.1% TFA in Milli Q H_2O
- Washing solution: 0.1% TFA in 5% CH_3OH
- Eluting solution: 1% formic acid in 50% CH_3OH for LTQ-Orbitrap analysis or 0.1% TFA in 50% CH_3CN for MALDI-TOF analysis

3.6.2 Instrumentation

- ZIP TIPs C_{18} pipette tips (Millipore, USA)
- pH meter 212 (Hanna Instruments, Italy)
- ESI-LTQ-Orbitrap mass spectrometer (Thermo Electron Corporation)
- Ultimate 3000 micro HPLC (Dionex)
- Column C_{18} Jupiter 4U Proteo (90 Å, 300 µm x 15 cm) (Phenomenex, Italy)
- Xcalibur® software (Thermo Fisher Scientific Inc.)

3.6.3 Procedure

i. Desalting

The dried samples were re-dissolved with 10 µL of 0.1% TFA solution in Milli Q H_2O, ensuring that the final sample solution had a pH<4.
The ZIP TIPs C_{18} were suitable for an Eppendorf pipette P10. Using the maximum volume setting of 10 µL, the necessary solution was taken into the tip and dispensed to waste. This procedure was repeated 3 times with the wetting solution and 3 times with the equilibration solution.
The binding of peptides to the ZipTip was achieved by fully depressing the pipette plunger to a dead stop, aspirating and dispensing the sample 7–10 times from and into the tube for maximum binding of complex mixtures. The binding capacity is 3 – 5 µg of peptides.
Desalting was achieved by using the maximum volume setting of 10 µL, aspirating into the tip and dispensing to waste 2 times the washing solution. The elution was performed aspirating and dispensing 6-10 times 10 µL of the eluting solution and recovering the eluted fraction into a new tube.
Samples were then dried again under N_2-flux.

ii. Instrumental conditions of analysis

Peptide analysis was performed with a Dionex Ultimate 3000 micro HPLC coupled with the LTQ-Orbitrap mass spectrometer equipped with a conventional ESI source. The source parameter was configured as follow: spray voltage 3.5 kV, capillary voltage 49 V and tube lens 75 V. For the chromatography separation, a Jupiter 4U Proteo (90 Å, 300 m £ 15 cm) column was used, and the column oven temperature was set to 25 °C; the separation was run for 82' using a gradient of 99.8/0.2 H_2O/HCOOH (eluent A) and 99.8/0.2 ACN/HCOOH (eluent B) and a flow/rate rate of 5 µL/min. The gradient was run as follows: 0–4' 95% A and 5% B, then to 50% A at 60', and 10% A and 90% B at 62, 62–72' 10% A, followed by the re-equilibration of the column. For MS1 scans, the Orbitrap resolution was 60.000 and the ion population 5×10^5, with an m/z window from 200 to 1800. For MS/MS in the LTQ, the population ion was 3×10^4 (isolation width of 3 m/z unit). 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) in all cases (measures by CIM-Parma, Italy) and collected by Xcalibur® software.

3.7 Maldi-TOF analysis

3.7.1 Chemicals

- Acetonitrile (CH_3CN) (Sigma-Aldrich, USA)
- Methanol (CH_3OH) (Carlo Erba, Italy)
- Trifluoroacetic acid 99% TFA (Acros, Belgium)
- MALDI Calibrants (Sigma-Aldrich, USA)
- α-Cyano-4-hydroxycinnamic acid (CHCA), (Sigma-Aldrich, USA)

3.7.2 Instrumentation

- MALDI-TOF 4700 Proteomic Analyzer (Applied Biosystem, USA)
- Software 4000 Series Explorer 3.5 v (Applied Biosystem, USA)
- Software Data Explorer v 4.9 (Applied Biosystem, USA)

3.7.3 Procedure

MALDI MS and MS/MS experiments were carried out on a 4700 Applied Biosystem Proteomics Analyzer. Each spectrum was taken by the following procedure: 1 μL of the desalted sample was spotted on the target plate and, immediately after, 1 μL of matrix solution (α-cyano-4-hydroxycynnamic acid, 10 mg/mL in 50% CH_3CN 50% TFA 0.1%) was spotted over it and 1 μL of sample was spotted again over it. After evaporation has occurred, the target was ready for the analysis. Mass spectrum acquisition was performed in positive ion reflectron mode by accumulating 625 shots/spectrum. The accelerating voltage was 20 kV. External mass calibration was performed with mass peptide standards. Data were acquired by using 4000 Series Explorer 3.5 v and analyzed by Data Explorer v 4.9.

4. Results and Discussion

4.1 Tomato genetic resources

A group of twelve tomato local varieties, both indeterminate and determinate in growth habits, were selected, differing in fruit shape and size and characterized by different final destination (fresh markets or canning) as a representative pool of tomato samples presently spread in local markets. Moreover, among the different cultivated forms of 'San Marzano' grown in the 'San Marzano' PDO area,[21] other three accessions were selected that are referred to in this study as 'SMMU', 'SMC' and 'SMMO' as they are presently in evaluation for the inclusion in the PDO disciplinary.

The protein extraction protocol used here allowed to obtain quite different amount of dry matter among ecotypes: in particular, VD, TD and SGRI samples were characterized by a lower dry matter amount, while PB and TI showed an higher amount of dry matter, compared to the other ecotypes (Table 1).

This variability may result from quantitative and qualitative differences in dry matter among the analyzed ecotypes, due to both genetic characteristics, typical of each genotype, and to the different degree of ripeness at harvest (the amount of fruit dry matter varies during ripening), despite the fact that tomatoes here analyzed were all ready for consumption. Concerning the amounts of protein extracted from the dry matters, they were even higher than the expected (1%), which might also due to eventual contaminants in the solution revealed by Qubit® fluorometer during the quantification process.

Table 1. Evaluation of protein extraction with the method described here and performed by Qubit® fluorometer.

Sample	A_s (g)	A_{dm}(g)	P (mg)	Prot %
SMMO	50,03	1,08	24,75	2,29%
PBD	50,20	1,17	24,00	2,05%
SMMU	50,03	1,12	26,65	2,38%
PS	50,07	1,32	33,24	2,52%
VD	50,07	0,73	23,43	3,21%
TD	50,07	0,36	13,04	3,62%
SGRI	50,04	0,79	26,48	3,35%
PB	50,05	1,47	34,80	2,37%
STLR	50,14	1,16	22,88	1,97%
TI	50,30	1,76	42,7	2,43%
SMC	50,85	1,03	26,56	2,58%
TLI	50,03	1,21	34,05	2,81%

A_s= Amount of sample; A_{dm}= Amount of dry matter; P= Amount of extracted proteins, obtained by Qubit® fluorometer; Prot %= Percentage of proteins related to A_{dm}.

4.2 Sera of allergic subjects

Human sera used in this work were collected from patients with a clinical history of allergic reactions towards fresh tomato and/or tomato products. Allergic subjects were recruited from two Italian regions, Campania and Emilia Romagna, in order to establish if environmental conditions could affect sensitization towards different allergenic proteins.

In Parma (North Italy) a total of five human sera were provided by Allergy and Clinical Immunology Center of the University Hospital from allergic subjects with Oral Syndrome after introduction of tomato and/or tomato products. All subjects had positive skin prick test for tomato and specific IgE ranging from 0.70 to 17.49 kU/L at ImmunoCAP dosage. All patients were also sensitized to other food allergens.

In Naples (South Italy) a total of six human serum samples were provided by the Division of Allergy and Clinical Immunology of the University Hospital 'Federico II', from allergic patients with food allergy to tomato. All subjects had positive skin prick test to tomato. Serum concentrations of IgE to tomato antigen ranged from 0.5 to 2 kU/L. Five out of six of these patients were also sensitized to other food allergens (mostly nuts, peanuts, peach, soybean or sesame seeds).

After some of them had been tested individually, sera belonging to each geographical group were blended together before immunoblotting experiments, thus using them as technological means to detect IgE-binding pattern of major tomato allergens.

4.3 Immunoblotting of tomato fruit extract with human sera

In order to define which proteins in tomato fruit extracts were recognized by the sera of allergic subjects and if differences were present among the different tomato ecotypes, protein extracts, obtained and quantified as reported paragraph 4.1, were separated by mono dimensional SDS-PAGE, by allowing two lane for every single sample (Fig 1a).

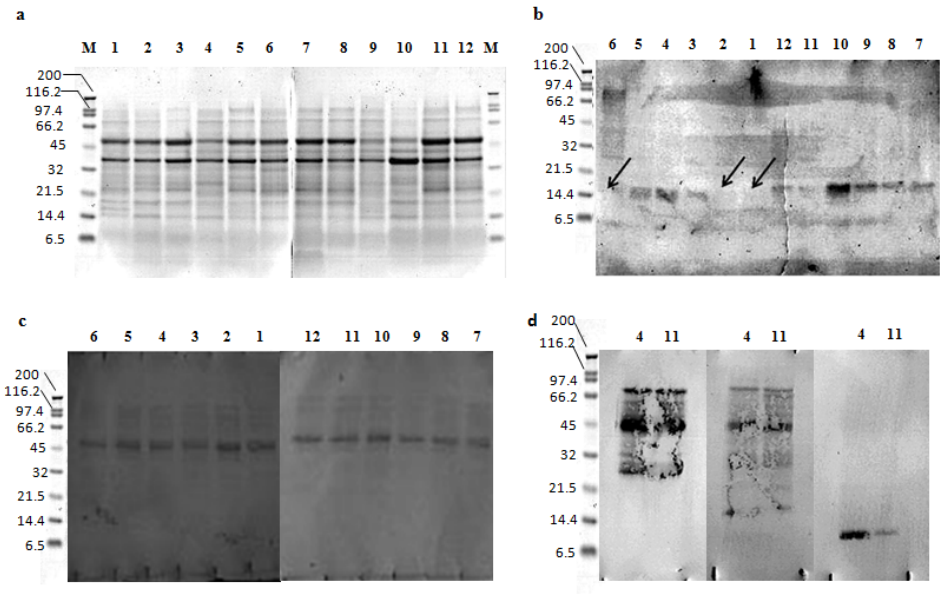


Fig 1. SDS-PAGE of twelve tomato ecotypes' fruit extracts. Numbers indicate ecotypes as described in "Materials and methods". The lane marked as 'M' represents the molecular weight standards, and their relative molecular masses are reported on the right side. **a**, staining with Coomassie brilliant blue; **b**, immunoblotting using pooled sera of patients from Campania region; arrows show the absence of reactive bands; **c**, immunoblotting using pooled sera of patients from Emilia Romagna region; **d**, immunoblotting using three sera of patients from Emilia Romagna region, incubated individually with two ecotypes. See text for details.

After separation, for every single ecotype, one lane in the gel was stained with Coomassie brilliant blue, whereas the other one was used in order to transfer the separated proteins onto a PVDF membrane. After incubation with pooled sera of patients from Campania region, a protein with a molecular mass of about 14 kDa was detected as the main allergen. No other nonspecific band was detected, even in correspondence to the most intense bands on the gel indicating a greater expression of those proteins. The immunoblot profiles of tomato ecotypes are shown in Fig 1b.

Quite interestingly, the IgE binding activity of the ecotypes 'Tondino D' (1), 'Ventura D' (2) and 'Pisanello' (6) was less intense, showing a potential reduced allergenicity for these individuals toward this tomato 14-kDa allergenic protein (see paragraph 4.4).

On the other hand, incubation of PVDF membrane with pooled serum sample collected from patients of the Emilia Romagna region did not confirm the potential hypoallergenicity of those ecotypes, since a totally different immunological response was obtained. In this case, a protein with molecular mass of about 45 kDa reacted toward the pooled sera of the allergic subjects, and scarce differences in the IgE binding properties of tomato ecotypes were observed (Fig 1c). The same results were

also obtained using a different pool of sera of allergic patients, selected in the same way as the previous one (data not shown).
The incubation of the same protein extracts with single serum of patients from Emilia Romagna region confirmed partially this result. As shown in Fig 1d, two of these sera, used individually and incubated with the protein extract of two ecotypes, revealed the same 45-kDa reactive protein, as the pool did, while another allergenic protein, whose apparent molecular weight was slightly lower than 14 kDa, was detected using the serum of another patient.
These results showed that the IgE binding patterns of these tomatoes ecotypes were highly serum-specific: the allergenic profiles may differ when immunoblottings with single serum are carried out. Since pools of sera contain all the involved IgE, the antigens which are recognized frequently by the single serum are mostly evidenced, quenching the detection of those less recurrent. Anyway, exactly for those reasons, when the most important allergens are to be evidenced, the use of pools of sera might be a very useful technological tool to assess the general allergenicity of a food commodity.
Hence, it could be assumed that the potential hypoallergenicity of tomatoes cannot be generalized for all the allergic subjects, but it should be assessed according to the specific allergen to which every single allergic subject is sensitized.

4.4 Identification of the 14-kDa allergenic protein

In order to identify the protein involved in specific IgE binding recognition, and also to possibly gain more evidences on the reason of the low IgE reactivity of some ecotypes, the bands on the gel corresponding to the reactive protein were in-gel-digested by trypsin and the peptide mixture was separated by HPLC (Fig 2) and analyzed by LTQ-Orbitrap mass spectrometer.

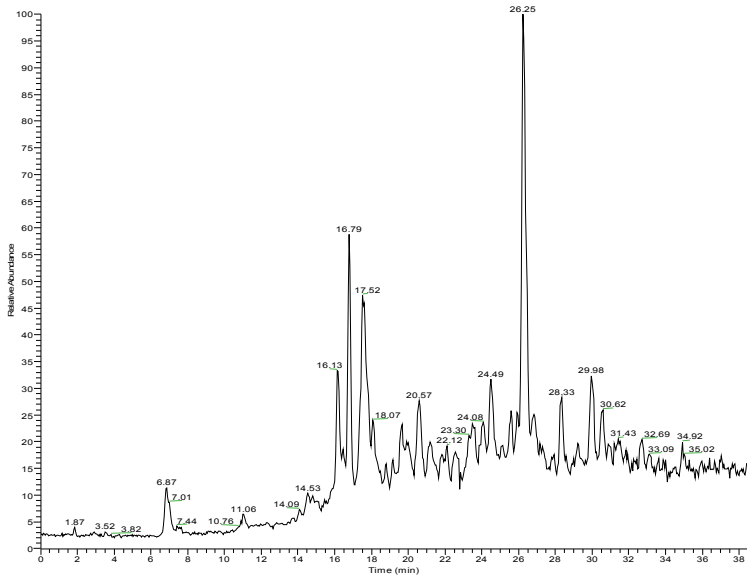


Fig 2. Chromatogram of the peptide mixture obtained, after the in-gel digestion of 14-kDa band of Tondino D ecotype and separated by HPLC.

Analysis were carried out on ten ecotypes, including those which showed a lower immunological response. Molecular masses and sequences of the recovered peptides (listed in Table 2) were used to query UniProt database, restricted to *Solanum lycopersicum* L. proteome, and the corresponding protein was identified as profilin (*Lyc e 1*). Among these, diagnostic peptides were also searched to elucidate which of the three profilin isoforms reported in Uniprot database was involved in allergic reactions.

Table 2. Identified peptides originated from 14-kDa allergen digestion and LTQ-Orbitrap analysis.

SWISS Protein Database Accession Number	Sequence	Position	Calculated MW (Da)[c]	Observed MW (Da)	MS spectra characteristic ions (m/z)
Q8VWR0	HMW peptide	1-71	7671.6	(discussed in text)	-
	YM*VIQGE**A**G**G**AVIR[a]	72-84	1421.7	1421.7	1112.9 (y_{11}), 1013.3 (y_{10}), 900.4 (y_9), 772.4 (y_8), 586.4 (y_6), 515.3 (y_5)
	G**A**GGIT**V**K	88-95	701.4	701.4	574.4 (y_6), 517.2 (y_5), 460.2 (y_4), 246.2 (y_2), 147.3 (y_1)
	G**A**GGIT**V**KK	88-96	829.5	829.5	773.5 (y_8), 588.4 (y_5), 475.3 (y_4), 275 (y_2)
	KGAGGITVK	87-95	829.5	829.5	574.4 (y_6), 517.3 (y_5), 460.3 (y_4), 246.2 (y_2)
	TNQALIIGIYDEPMTPGQC*NMIVER[b]	97-121	2862.4	2862.4	1660.5 (b_{15}), 1559.5 (b_{14}), 1331.3 (b_{12}), 924.2 (b_9), 641.2 (b_6)
	LGDYIIEQ**G**L	122-131	1119.6	1119.6	989.0 (b_9), 932.1 (b_8), 804.4 (b_7), 675.3 (b_6), 561.9 (b_5)
Q93YG7	HMW peptide	1-71	7678.7	(discussed in text)	-
	YMVIQGE**PE**AVIR	72-84	1503.8	1503.8	1210.5 (y_{11}), 1111.5 (y_{10}), 998.4 (y_9), 870.4 (y_8), 684.3 (y_6)
	G**P**GGIT**I**K	88-95	741.4	741.4	588.3 (y_6), 530.8 (y_5), 474.4 (y_4)
	G**P**GGIT**I**KK	88-96	869.5	869.5	724.4 (b_8), 382.2 (b_5)
	TNQALIIGIYDEPMTPGQC*NMIVER[b]	97-121	2862.4	2862.4	1660.5 (b_{15}), 1559.6 (b_{14}), 1331.1 (b_{12}), 924.3 (b_9), 754.2 (b_7), 415.2 (b_4)
	LGDYLIEQ**S**L	122-131	1149.6	1149.6	1019.2 (b_9), 932.3 (b_8), 804.3 (b_7), 675.1 (b_6)

a: M* indicates an oxidized methionine

b: C* indicates a carboxamidomethylcysteine

c: Monoisotopic MW

Bold letters represent amino acids residues which are different in homologous peptides of Profilin isoforms

As expected, marker peptides belonging to profilin-1 (SWISS Protein database accession no. Q41344) (Fig. 7) corresponding to the sequences AEEITNIMK (46-54), DFDEPGHLAPTGLFLAGTK (55-73), YMVIQGEPGAVIR (74-86), TAQALIFGVYEEPVTPGQCNMVVEK (99-123) and IGDYLVDQGY (124-133) were not found, since this isoform is expressed in a pollen specific manner in tomato.[22] Profilin and profilin-2 isoforms (SWISS Protein database accession no. Q8VWR0 and Q93YG7, respectively) were both found in all ecotypes but one, 'Pisanello' (PS) (quite interestingly, one of those showing low IgE reactivity) which showed only peptides belonging to Q8VWR0 isoform (Fig 3).

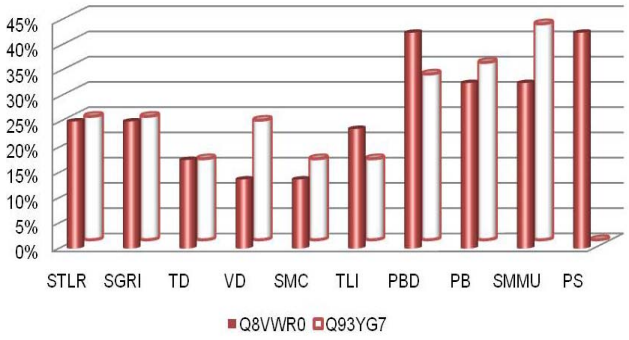


Fig 3. Percentage of sequence coverage among ecotypes for profilin Q8VWR0 and Q93YG7 isoforms calculated for the analyzed ecotypes, determined by the abundance of peptides identified for each of them.

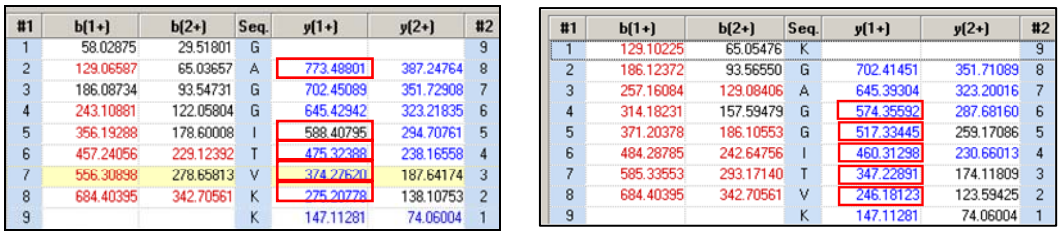
The low percentage of sequence coverage (average among ecotype was estimated 27% for Q8VWR0 isoform and 23% for Q93YG7 isoform) was likely due to protein's primary structure, poor in positively charged amino acids and thus not easily digested by trypsin. As shown in Fig 4. the first suitable cleavage site for trypsin on Lys_{71} generates high molecular weight peptides which are difficult to elute from gel and detected by mass spectrometry. Moreover, in the same peptide, Lys_{43} was not accessible by the enzyme since it is followed by a proline residue, except in the Q41344 isoform in which Ala_{46} and Lys_{54} would allow to generate two additional diagnostic peptides for this isoform discrimination. So, although these two HMW peptides could be useful to discriminate profilin isoforms since they differ for some amino acids and, subsequently, for their molecular masses, they were not found in any ecotypes.

```
Q93YG7[a] MSWQTYVDEHLLCENE--GNHLTSAAIIGQDGTVWAQSANFPQFKPEEITGIMNDFAVPG  58
Q8VWR0 MSWQTYVDDHLMCDIE--GNHLTSAAIIGQDGSVWAQSANFPQFKPEEITAIMNDFAEPG  58
Q41344[a] MSWQTYVDDHLMCDIEGTGHHLSSAAILGFDGSVWAQSPNFPKFKAEEITNIMKDFDEPG  60
          ********:**:*: *  *:**:****:* **:*****.***:**.**** **:**  **
Q93YG7[a] TLAPTGLYLGGIKYMVIQGEPEAVIRGKKGPGGITIKKTNQALIIGIYDEPMTPGQCNMI  118
Q8VWR0 TLAPTGLHLGGTKYMVIQGEAGAVIRGKKGAGGITVKKTNQALIIGIYDEPMTPGQCNMI  118
Q41344[a] HLAPTGLFLAGTKYMVIQGEPGAVIRGKKGPGGITIKKTAQALIFGVYEEPVTPGQCNMV  120
           ******.*.**********. ********.****:*** ****:*:*:**:*******:
Q93YG7[a] VERLGDYLIEQSL  131
Q8VWR0 VERLGDYIIEQGL  131
Q41344[a] VEKIGDYLVDQGY  133
          **::***:::*.
```

a Initiator methionine is cleaved from the mature protein

Fig 4. Sequence similarity of tomato profilin isoforms: Q8VWR0, Q93YG7, Q41344. Cutting sites for trypsin are highlighted.

High resolution tandem mass spectrometry also allowed to identify the sequences of two isobaric peptides of Q8VWR0 isoform, GAGGITVKK and KGAGGITVK (88-96 and 87-95, respectively), which were generated by the cleavage of trypsin on two consecutive lysine at both N-terminal and C-terminal amino acid sequences, by detecting characteristic ions for each sequence, therefore allowing the unequivocal discrimination of these two peptides (Fig 5).



#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	58.02875	29.51801	G			9
2	129.06587	65.03657	A	773.48801	387.24764	8
3	186.08734	93.54731	G	702.45089	351.72908	7
4	243.10881	122.05804	G	645.42942	323.21835	6
5	356.19288	178.60008	I	588.40795	294.70761	5
6	457.24056	229.12392	T	475.32388	238.16558	4
7	556.30898	278.65813	V	374.27620	187.64174	3
8	684.40395	342.70561	K	275.20778	138.10753	2
9			K	147.11281	74.06004	1

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	129.10225	65.05476	K			9
2	186.12372	93.56550	G	702.41451	351.71089	8
3	257.16084	129.08406	A	645.39304	323.20016	7
4	314.18231	157.59479	G	574.35592	287.68160	6
5	371.20378	186.10553	G	517.33445	259.17086	5
6	484.28785	242.64756	I	460.31298	230.66013	4
7	585.33553	293.17140	T	347.22891	174.11809	3
8	684.40395	342.70561	V	246.18123	123.59425	2
9			K	147.11281	74.06004	1

Fig 5. Summary of ions obtained after the fragmentation of GAGGITVKK and KGAGGITVK isobaric peptides. Red bolds indicate discriminating y(+1) ions for each peptide.

Although the fact that one of the low IgE-responding ecotype did not show the presence of one isoform, this feature was not present in the other two non-IgE binding isoforms, and thus no clear correlation was found between the distribution of the detected Profilin isoforms and the weaker immunological response of 'Tondino D', 'Ventura D' and 'Pisanello' ecotypes. Probably, the reduced immunological reactivity can depend not only on the isoform distribution, but also on a natural downregulation in the expression of profilin.

4.5 Identification of the 45-kDa allergenic protein

The same procedure presented above was applied for the identification of the 45-kDa allergen. Anyway, in this case, after HPLC separation (Fig 6), LTQ-Orbitrap analysis of the tryptic digest of the band with molecular mass of 45 kDa recognized peptides belonging to three different tomato proteins, whose calculated molecular masses were all consistent with the observed one (Table 3).

Table 3. Identified peptides originated from 45-kDa allergen digestion and LTQ-Orbitrap analysis.

Sequence	Protein Accession	Calculated MW (Da)[b]	Observed MW (Da)	MS spectra characteristic ions (m/z)
AVVDSAIDAETR	P15003	1245.6	1245.6	1076.5 (y_{10}), 977.4 (y_9), 862.4 (y_8), 775.4 (y_7), 704.4 (y_6), 591.3 (y_5), 476.3 (y_4), 405.3 (y_3), 276.2 (y_2)
LGGQTYSVALGR	P15003	1220.7	1220.7	1108.6 (y_{11}), 1051.6 (y_{10}), 866.5 (y_8), 765.5 (y_7), 602.5 (y_6), 515.4 (y_5), 416.3 (y_4), 345.2 (y_3), 232.2 (y_2)
VGADMSVINR	Q96577	1060.5	1060.6	962.5 (y_9), 905.5 (y_8), 834.4 (y_7), 719.4 (y_6), 588.4 (y_5), 501.4 (y_4), 402.3 (y_3), 289.1 (y_2)
LTSDDDFFTNPM*VK[a]	Q8L5J1	1644.7	1644.8	1431.5 (y_{12}), 1344.4 (y_{11}), 1229.4 (y_{10}), 1114.5 (y_9), 999.4 (y_8), 852.4 (y_7), 705.4 (y_6), 604.4 (y_5), 490.3 (y_4)

a : M* indicates an oxidized methionine

b: Monoisotopic MW

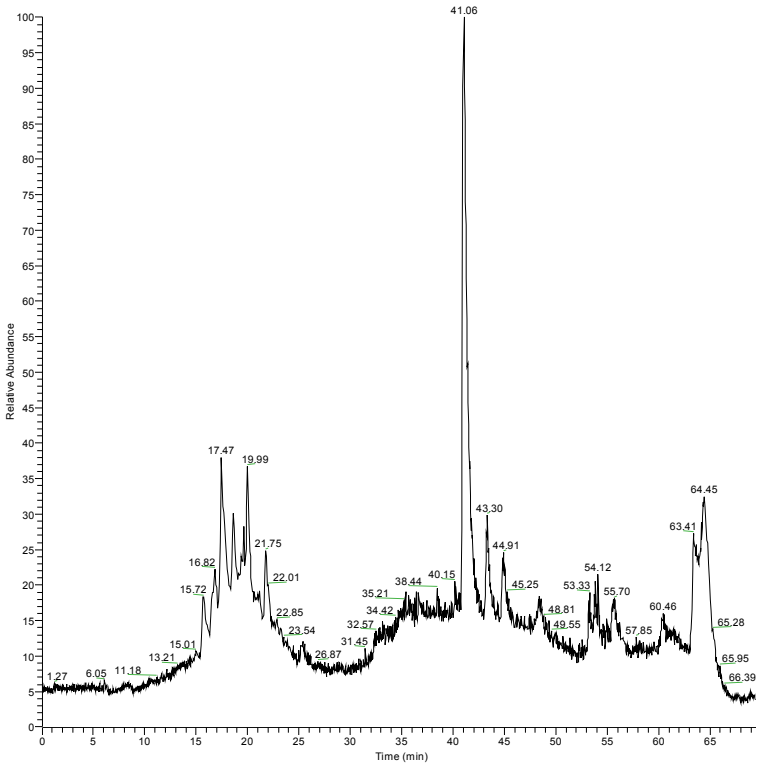


Fig 6. Chromatogram of the peptide mixture, obtained after the in-gel digestion of 45-kDa band of Sorrento Rosato ecotype and separated by HPLC.

Peptides AVVDSAIDAETR and LGGQTYSVALGR (Figg 7 and 8) were assigned to tomato suberization-associated anionic peroxidase (SWISS Protein database accession no. P15003). These sequences represent two diagnostic fragments in the discrimination of two known tomato peroxidase isoforms, whose molecular masses are very similar (Protein database accession no. P15003 and P15004): the first differed from the analogous peptide in two alanine residues (Ala_1 and Ala_9, Gly_1 and Asn_9, respectively), while the latter showed a Ser_7 in place of Thr_7.

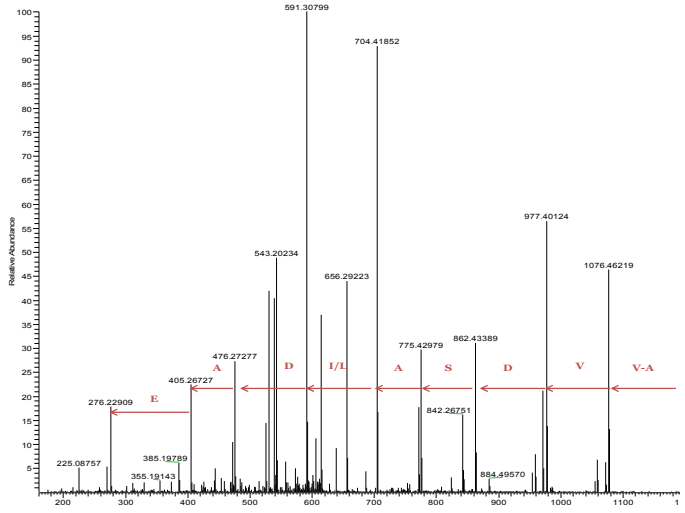


Fig 7. MS/MS spectrum of AVVDSAIDAETR peptide.

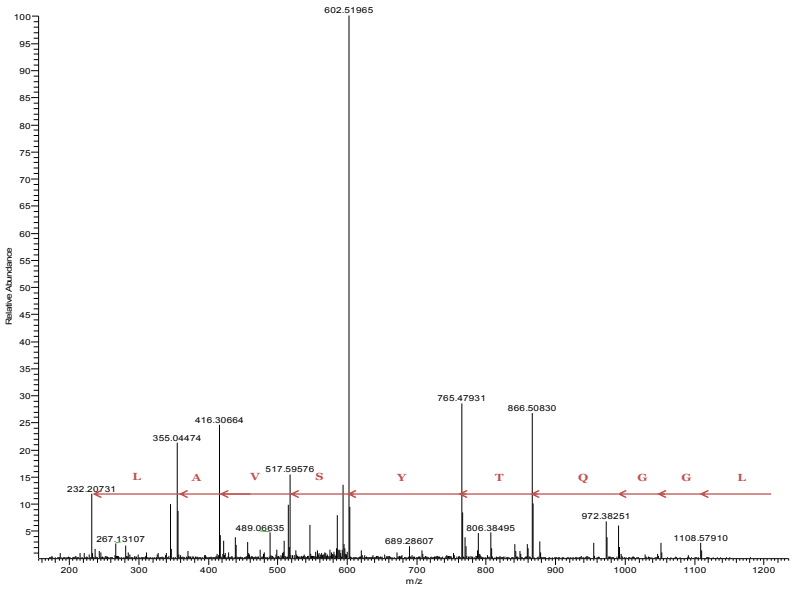


Fig 8. MS/MS spectrum of LGGQTYSVALGR peptide.

Peptide VGADMSVINR (Fig 9) was found to be shared by three tomato pectinesterase isoforms (SWISS Protein database accession no. P14280, Q96576 and Q96577); according to their calculated molecular masses, the Q96577 protein seemed to fit better the molecular weight observed on gel.

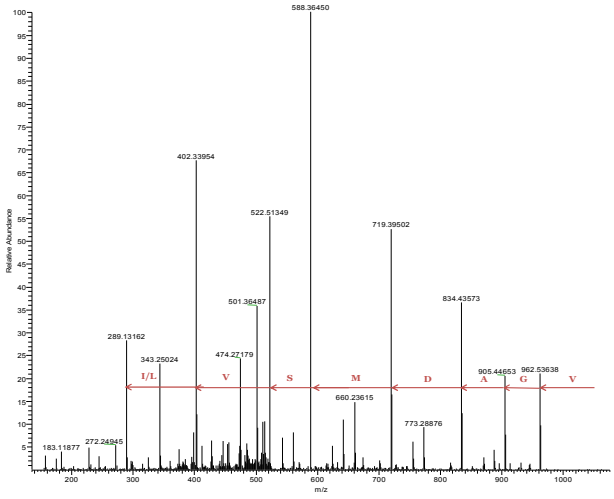


Fig 9. MS/MS spectrum of VGADMSVINR peptide.

Finally, for peptide LTSDDDFFTNPMVK (Fig 10), the Protein BLAST software only found a perfect match with tomato mannan endo-1,4-beta-mannosidase (SWISS Protein database accession no. Q8L5J1).

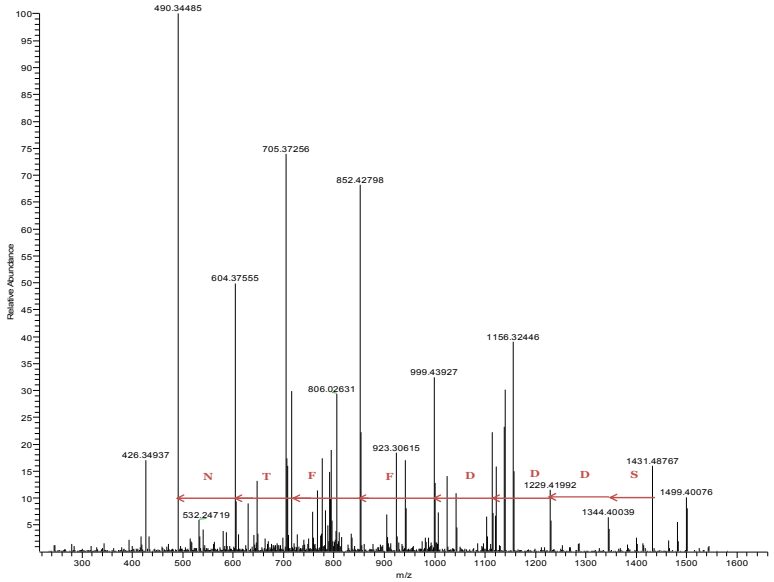


Fig 10. MS/MS spectrum of LTSDDDFFTNPMVK peptide.

Weangsripanaval et al. have already found peroxidase to be the major tomato allergen recognized by the sera of patients suspected to suffer from food allergies and diagnosed to be atopic dermatitis.[23] A different pectinesterase isoform has already been observed as tomato allergens by Kondo et al. (SWISS Protein database accession no. P14280) in patients with OAS after ingestion of fresh tomato fruits.[11] As regards mannan endo-1,4-beta-mannosidase, no evidence of induction of any allergic reaction toward tomato has so far been reported. So, in order to clarify which of the identified proteins was involved in triggering the immunological response, we exploited their different pI for a separation of tomato protein total extracts by 2D PAGE (Fig 11a). Immunoblotting on the bidimensional gel electrophoresis, performed on 'Sorrento Rosato' ecotype, indicated a reactive spot at 45 kDa with a pI of about 4 (Fig 11b), thus excluding the possibility that pectinesterase (whose isoeletric points range from 8.2 to 8.5, depending on the isoforms),[24] and mannan endo-1,4-beta-mannosidase (with an observed isoeletrical point of 9),[25] were involved in the allergic reaction.

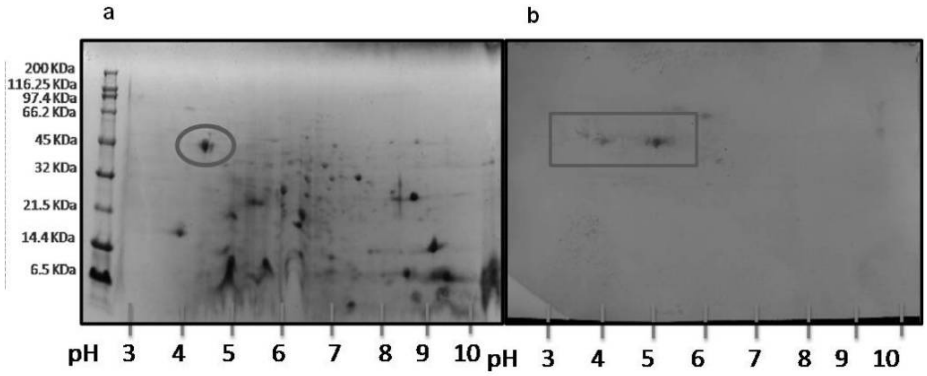


Fig 11. a, 2D SDS-PAGE of protein total extract of Sorrento Rosato ecotype; **b**, corresponding immunoblotting on PVDF membrane using patients' sera from Emilia Romagna region. The square and the circle indicate the reactive spot and the related protein after electrophoresis, respectively.

In order to definitely confirm the identity of the allergenic protein, the corresponding spot was subjected to in-gel trypsin digestion and the tryptic mixture, analyzed by MALDI-TOF spectrometer, revealed peptides belonging to tomato suberization-associated anionic peroxidase (SWISS Protein database accession no. P15003). MALDI-TOF spectrum is shown in Fig 12.

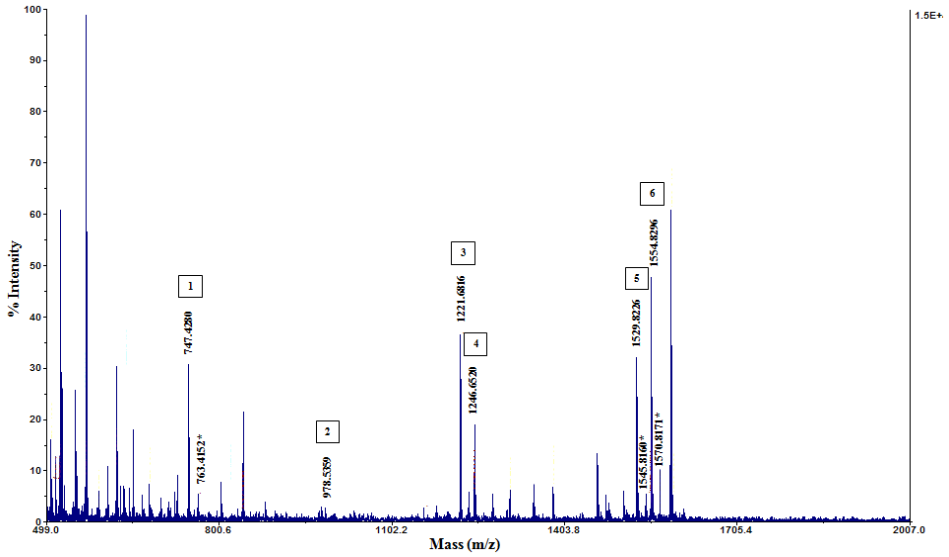


Fig 12. Maldi-TOF spectrum of 2D reactive spot at 45 kDa, after in-gel trypsin digestion. Numbered peaks were further characterized by MS/MS analysis, in order to confirm the expected sequence suggested by in silico analysis. Ions marked with a star correspond to the previously identified peptide plus an oxidized methionine.

Amino acid sequences of these peptides, derived by in silico digestion of this protein, were further confirmed by MS/MS analysis (Table 4, Fig 13).

Table 4. Peptides originated from 45-kDa spot digestion identified by MALDI TOF-TOF analysis.

Maldi TOF peak[a]	Sequence	Position	Calculated MW (Da)[b]	Observed MW (Da)	MS spectra characteristic ions (m/z)
1	MGASLIR	100-106	746.4	746.4	432.2 (a_5), 460.3 (b_5), 347.2 (b_4), 260.2 (b_3), 271.2 (z_2), 175.1 (y_1)
2	GYEVIAQAK	145-153	977.5	977.5	978.5 (MH^+)
3	LGGQTYSVALGR	182-193	1220.7	1220.7	765.9 (y_7), 602.4 (y_6), 515.4 (y_5), 416.3 (y_4), 232.2 (y_2), 175.1 (y_1)
4	AVVDSAIDAETR	88-99	1245.6	1245.7	862.4 (y_8), 476.3 (y_4), 276.2 (y_2), 175.1 (y_1)
5	EMVALAGAHTVGFAR	231-245	1528.8	1528.8	858.5 (y_8), 549.4 (y_5), 450.4 (y_4), 175.1 (y_1)
6	MGDLPPSAGAQLEIR	336-350	1553.8	1553.8	1251.6 (y_{12}), 1138.6 (y_{11}), 1041.5 (y_{10}), 786.5 (y_7), 658.4 (y_5), 288.2 (y_2), 175.1 (y_1)

a: numbers indicate peaks of Maldi TOF spectrum in Fig. 15

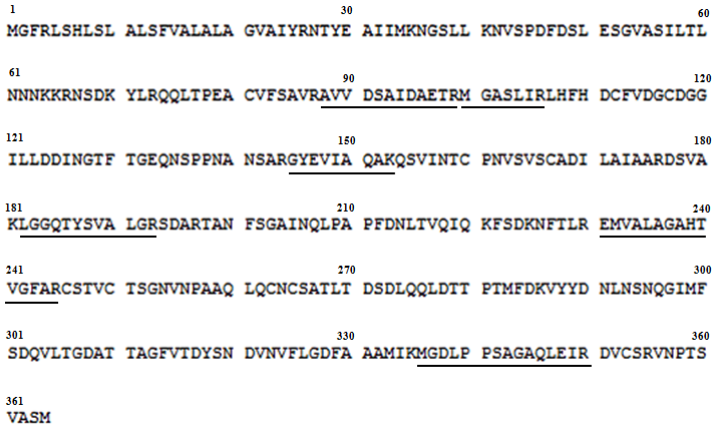


Fig 13. Amino acid sequence of suberization-associated anionic peroxidase protein, including signal peptide (1–25), identified by MALDI-TOF spectrometer.
The lines mark peptides identified by MALDI MS/MS analysis.

5. *Conclusions*

Pooled sera of allergic subjects from two different parts of Italy, Campania and Emilia Romagna regions, were used in order to assess the allergenicity potential of 12 different tomato ecotypes. The ecotypes were assessed by a proteomic approach, performing immunoblotting experiments and identifying the reactive proteins by high-resolution mass spectrometric techniques. Quite interestingly, the two pools showed a totally different immunological response: for the first group, the main allergen was identified as profilin, whereas for the second one the main allergen was identified as suberization-associated anionic peroxidase.
Moreover, experiments by using single sera were performed, further outlining the individual response to different allergenic proteins.
Three different ecotypes showed a low IgE response by using pooled sera of subjects from Campania region, whereas no hypoallergenicity was observed by using pooled sera of subjects from Emilia Romagna region. The hypoallergenicity observed might be due to several reasons: lower expression of the proteins in the different ecotypes, or the presence of protein isoforms having lower allergenicity (although specific experiments aimed at detecting different profilin isoforms outlined that all varieties seemed to have two different profilin isoforms). In any case, it is quite evident that the property of hypoallergenicity for a defined variety, given this scenario, is strictly dependent on which protein is responsible for the allergen response in first instance. For people sensitized to profilin, according to our experiments, some varieties showed

a diminished IgE response (and thus might be hypoallergenic), but for people sensitized to suberization-associated anionic peroxidase, the very same varieties were not hypoallergenic at all. These results underline the fact that allergies to a defined food in different subjects can often rely on a completely different immunological response at the molecular level. Although for some subjects hypoallergenic tomato varieties might eventually be found, generally hypoallergenic tomatoes, whose hypoallergenicity is valid for all consumers, seem to be very difficult to obtain, since the characteristic molecular responses of the single allergic subjects are always to be taken into account.

6. References

[1] Sicherer SH, Sampson HA. Food allergy: recent advances in pathophysiology and treatment. *Annual Review of Medicine*, 2009, **60**:261–277.

[2] COMMISSION DIRECTIVE 2007/68/EC of 27 November 2007 amending Annex IIIa to Directive 2000/13/EC of the European Parliament and of the Council as regards certain food ingredients (2007) Official Journal of the European Union. http://eur-lex.europa.eu.

[3] Osterballe M, Hansen TK, Mortz CG, Bindslev-Jensen C. The clinical relevance of sensitization to pollen-related fruits and vegetables in unselected pollen-sensitized adults. *Allergy,* 2005, **60**:218–225.

[4] Vieths S, Scheurer S, Ballmer-Weber B. Current understanding of cross-reactivity of food allergens and pollen. *Annals of the New York Academy of Sciences,* 2002, **964**:47–68.

[5] Ebner C, Hirschwehr R, Bauer L, Breiteneder H, Valenta R, Ebner H, Kraft D, Scheiner O. Identification of allergens in fruits and vegetables: IgE cross-reactivities with the important birch pollen allergens Bet v 1 and Bet v 2 (birch profilin). *Journal of Allergy and Clinical Immunology*, 1995, **95**:962–969.

[6] Westphal S, Kempf W, Foetisch K, Retzek M, Vieths S, Scheurer S. Tomato profilin Lyc e 1: IgE cross-reactivity and allergenic potency. *Allergy*, 2004, **59**:526–532.

[7] De Martino M, Novembre E, Cozza G, De Marco A, Bonazza P, Vierucci A. Sensitivity to tomato and peanut allergens in children monosensitized to grass pollen. *Allergy* 1988, **43**:206–213.

[8] Beezhold DH, Sussman GL, Liss GM, Chang NS. Latex allergy can induce clinical reactions to specific foods. *Clinical & Experimental Allergy*, 1996, **26**:416–422.

[9] Westphal S, Kolarich D, Foetisch K, Lauer I, Altmann F, Conti A, Crespo JF, RodrÂguez J, Enrique E, Vieths S, Scheurer S. Molecular characterization and allergenic activity of Lyc e 2 (β-fructofuranosidase), a glycosylated allergen of tomato. *European Journal of Biochemistry*, 2003, **270**:1327–1337.

[10] Le LQ, Lorenz Y, Scheurer S, Fötisch K, Enrique E, Bartra J, Biemelt S, Vieths S, Sonnewald U. Design of tomato fruits with reduced allergenicity by dsRNAi-mediated inhibition of ns-LTP (Lyc e 3) expression. *Plant Biotechnology Journal*, 2006, **4**:231–242.

[11] Kondo Y, Urisu A, Tokuda R. Identification and characterization of the allergens in the tomato fruit. *International Archives of Allergy and Immunology*, 2001, **126**:294–299.

[12] Asero R, Mistrello G, Roncarolo D, Amato S, Arcidiacono R, Fortunato D. Detection of a novel allergen in raw tomato. *Journal of Investigational Allergology and Clinical Immunology,* 2008, **18**:397–400.
[13] Bassler OY, Weiss J, Wienkoop S, Lehmann K, Scheler C, Dolle S, Schwarz D, Franken P, George E, Worm M, Weckwerth W. Evidence for novel tomato seed allergens: IgE-reactive legumin and vicilin proteins identified by multidimensional protein fractionation-mass spectrometry and in silico epitope modeling. *Journal of Proteome Research*, 2009, **8**:1111–1122.
[14] Germini A, Paschke A, Marchelli R. Preliminary studies on the effect of processing on the IgE reactivity of tomato products. *Journal of the Science of Food and Agriculture*, 2007, **87**:660–667.
[15] Pravettoni V, Primavesi L, Farioli L, Brenna OV, Pompei C, Conti A, Scibilia J, Piantanida M, Mascheri A, Pastorello EA. Tomato allergy: detection of IgE-binding lipid transfer proteins in tomato derivatives and in fresh tomato peel, pulp, and seeds. *Journal of Agricultural and Food Chemistry*, 2009, **57**:10749–10754.
[16] Le LQ, Mahler V, Lorenz Y, Scheurer S, Biemelt S, Vieths S, Sonnewald U. Reduced allergenicity of tomato fruits harvested from Lyc e 1–silenced transgenic tomato plants. *Journal of Allergy and Clinical Immunology*, 2006, **118**:1176–1183.
[17] Breiteneder H, Radauer C. A classification of plant food allergens. *Journal of Allergy and Clinical Immunology*, 2004, **113**:821–830.
[18] Kitagawa M, Moriyama T, Ito H, Ozasa S, Adachi A, Yasuda J, Ookura T, Inakuma T, Kasumi T, Ishiguro Y, Ito Y. Reduction of allergenic proteins by the effect of the ripening inhibitor (rin) mutant gene in an F1 hybrid of the rin mutant tomato. *Bioscience Biotechnology & Biochemistry*, 2006, **70**:1227–1233.
[19] Monaci L, Visconti A. Mass spectrometry-based proteomics methods for analysis of food allergens. *Trends in Analytical Chemistry*, 2009, **28**:581–591.
[20] Yagami T, Haishima Y, Tsuchiya T, Tomitaka-Yagami A, Kano H, Matsunaga K. Proteomic analysis of putative latex allergens. *International Archives of Allergy and Immunology*, 2004, **135**:3–11.
[21] Rao R, Corrado G, Bianchi M, Di Mauro A. (GATA)4 DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. *Plant breeding,* 2006, **125**:173–176.
[22] Yu LX, Nasrallah J, Valenta R, Parthasarathy MV. Molecular cloning and mRNA localization of tomato pollen profilin. *Plant Molecular Biology,* 1998, **36**:699–707.
[23] Weangsripanaval T, Nomura N, Moriyama T, Ohta N, Ogawa T. Identification of suberization-associated anionic peroxidise as a possible allergenic protein from tomato. *Bioscience Biotechnology & Biochemistry*, 2003, **67**:1299–1304.
[24] Gaffe J, Tieman DM, Handa A. Pectin methylesterase isoforms in tomato. *Plant Physiology*, 1994, **105**:199–203.
[25] Carrington CMS, Vendrell M, Dominguez-Puigjaner E. Characterisation of an endo-(1, 4)-β-mannanase (LeMAN4) expressed in ripening tomato fruit. *Plant Science,* 2002, **163**:599–606.

Chapter IV

Proteomic approach for the characterization of tomato allergen nsLTP (nonspecific Lipid Transfer Protein)

1. Introduction

Among the allergens found in plants, Lipid Transfer proteins (LTPs) undoubtedly represents a very important class. LTPs are cationic polypeptides, subdivided into two families.[1] Proteins that form the first family, called LTP1, have molecular masses of approximately 10 kDa and are basic with isoeletric points (pI) ranging between 9 and 10, since their primary structure is abundant in charged lysine residues. These LTPs typically have 90-95 amino acid residues, of which eight are cysteines conserved in similar positions along the primary structure. These eight cysteines, bound to each other, form four disulphide bridges that help the stabilization of the protein tertiary structure.[2] The LTP2 family is formed by peptides that have molecular masses of approximately 7 kDa, having on average 70 amino acids; they share all the other characteristics, such as a high pI, lipid transfer activity and a conserved disulphide skeleton, with the LTP1 family.[3,4,5] Both protein types are synthesized with a signal peptide at the amino terminal region, which in general varies between 21 and 27 amino acids, for the LTP1 family,[6,7,8] and from 25 to 35 amino acids, for the LTP2 family.[9,10] This signal peptide is then removed, yielding mature proteins ready to be delivered to the cell secretory pathway, where they are exported to the apoplast.

The secondary structure of proteins belonging to the LTP1 family, found in rice and maize, is composed of four α-helices (helices H_1 from Cys_3 to Ala_{17}, H_2 from Ala_{25} to Ala_{37}, H_3 from Thr_{41} to Ala_{56} and H_4 from Ala_{63} to Cys_{73}) and a long carboxy terminal tail that lacks defined secondary structure.[11,12] The LTP2 family, isolated from wheat and rice, follows a similar secondary structural pattern as the LTP1 family, but has three α- helices (H_1 from Cys_3 to Ala_{16}, H_2 from Thr_{22} to Ala_{31} and H_3 from Gln_{33} to Ala_{40}) and a region containing two single-turn helices (Tyr_{45} to Tyr_{48} and Ala_{54} to Val_{58}).[4,13]

The three dimensional structure of several plant LTP1 proteins, determined either by X-ray crystallography or NMR, reveals a compact and globular structure that is stabilized by four disulphide bridges among the eight cysteines.[11,12] The most striking structural feature of the LTP1 family is the presence of a flexible hydrophobic cavity in a form of a tunnel that runs through the molecule's axes. The cavity has two entrances, one smaller and one larger,[11,12] and has two charged amino acids, an Arg_{44} and a Lys_{35}, which are strategically localized on the larger entrance of the hydrophobic cavity, indicating a possible role in the interaction with lipids. Lipid molecules may interact with the protein at the larger entrance and their hydrophobic portions stay buried inside the cavity, while the carboxylate portion remains turned toward or exposed to the solvent. Additionally, a Tyr at the carboxyl terminal region at approximately position 79, is of particular note. In the three-dimensional structure, this residue is positioned at the larger entrance of the hydrophobic cavity and it has been shown that it interacts with fatty acids and stabilizes the binding between the peptide and the hydrophobic molecule by a hydrogen bond that is formed between the hydroxyl of the Tyr and the carboxyl group of the polar head of the lipid.[14]

This structure in the LTP2 family is a triangular hollow box, instead of a tunnel.[13] The volume of the cavity of both groups can increase or contract in order to better accommodate the hydrophobic molecule and this plasticity is presumably responsible

for the lack of specificity in the transport ability; hence the name 'nonspecific' lipid transfer protein.[13,14]

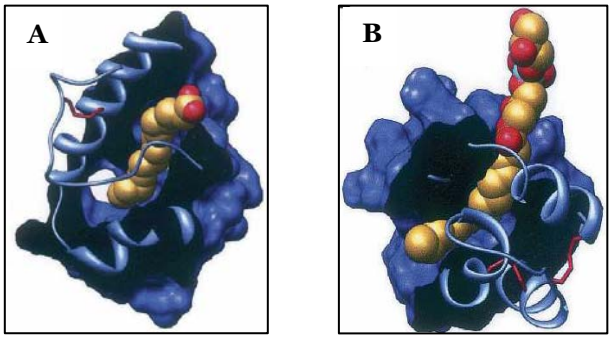


Fig 1. **A**: Cutaway view showing the lipid binding pocket of rice LTP1 occupied by myristic acid; **B**: Cutaway image showing wheat LTP2 bound to L-a-palmitoyl-phosphatidyl glycerol. Adapted from reference 15.

Different possible functions have been proposed for plant LTPs, but their true physiological roles have yet to be determined. It is possible that LTPs fulfil more than one role *in planta*.

The genetic structure of LTP1 indicates the presence of several genes coding for LTP proteins in a single genome. This multigene nature yields many protein isoforms which show differential tissue and developmental expression patterns and possibly different functions, as well.[8,16,17,18,19] Hence, the analysis of where, when and how the LTP genes are expressed may help the understanding of their functions in vivo.

On the basis of their observed in vitro capacity for transferring lipids and binding acyl-chains, the suggestion that LTPs could be involved in many aspects of cell function in which movement of lipids is thought to be important was a logical one.[2] However, the fact that LTPs are also located extracellularly might imply other roles.[20]

The finding of LTP in young leaves, petal and sepal abscission zone in *Brassica oleracea* and *Arabidopsis thaliana* may suggest a role of the LTP in the transport of monomers of cutin and in deposition of lipophilic cuticular material.[21,22]

LTP expression was also showed to be enhanced by some environmental changes, such as drought, cold and salt stress. For example, the expression of LTP genes could be regulated through the same signalling pathway observed in the activation of stress-induced genes by signal molecules such as abscisic acid, salicylic acid, ethylene and methyl jasmonate.[9,23] The preferred location of LTPs in outer cell layers and their expression induction by abscissic acid may also be indicative of a role for LTPs in repulsing or suppressing pathogenic attack from outside. Because of their basic properties, LTPs could act as membrane permeabilizing agents, thus inhibiting the growth of bacterial pathogens and fungi.[24] For these reasons, LTPs have been included into the group of the so-called "pathogenesis-related proteins-14", a large family of inducible proteins that are produced by plants upon stimuli associated with defence and related stress.[25]

Several reports have unambiguously suggested that the major allergens of different plant species are proteins belonging to LTP1 family. Indeed these proteins have been identified in fruits of *Rosaceae*, such as peach and apple,[26,27] plum,[28] and pear,[29] in fruits of *Vitaceae*,[30] as well as in other plant species such as *Zea mays*,[31] *Triticum*

aestivum,[32] *Punica granatum*,[33] *Corylus avellana*,[34] *and Actinidia chinensis*.[35] Due to their compact three dimensional structure, these LTPs are relatively stable, resisting thermal and chemical denaturation and enzymatic digestion.[36] Considering that the stability to digestion and to heat treatments has been claimed as characteristic of a 'true' food allergen,[37,38] LTPs, showing stability to both proteolytic and heat treatments could maintain their immunogenic and allergenic motifs and thus could interact with the immune system associated with the gastrointestinal epithelia, inducing both sensitization and systemic symptoms.[37] This stability also explains why LTPs have been identified as relevant allergens in processed foodstuffs and beverages, as beer,[39] and wine,[30] peach-derived products,[40] polenta,[41] roasted hazelnuts,[42] baked or boiled apples.[43]

LTPs also differ from the other food allergens in the route of sensitization. Whereas allergic reactions to plant-derived foods in patients with Oral Allergy Syndrome (OAS) are usually due to sensitization to pollens which contain allergens cross-reacting with homologous molecules in food (like the birch-apple and latex-fruit syndrome),[44,45] allergy to LTP in food does not depend on sensitization to LTP-containing pollens: due to the resistance to gastric digestion, LTPs are capable of sensitizing through the gastrointestinal tract, giving severe systemic reactions.[46]

As far as tomato is concerned, the presence of LTP isoforms has been detected both in tomato peel, pulp, seeds and in the commercial tomato derivatives.[47] In particular, it has been shown that chemical peeling or thermal treatment, such as technological processes to prepare tomato products, did not reduce the IgE reactivity due to LTP, since this protein is present not only in peel, but also in pulp and seeds.

The lack of positive scores on Skin Prick Test (SPT) with tomato extracts in patients monosensitised to LTP may due to the use of commercial extracts which did not contain the LTP protein, thus underestimating the real hazard. Conversely, in the literature very few cases of systemic symptoms upon ingestion of tomato have been reported, raising the question of the prevalence of LTP as allergen in tomato allergy.[48]

The availability of the purified LTP proteins, from natural sources or obtained by recombinant technologies, may allow a faster and reliable screening of LTP-allergic patients, when used as unique component of different allergenic tests, such as Skin Prick Test or Immuno Sorbent Allergo Chip (ISAC) or to produce monoclonal specific antibodies.

2. *Aim of the work*

Despite its growing importance at clinical level as tomato allergen, very few attempts to better characterize tomato LTP have been carried out so far.[47,49] These studies mainly focus on N-terminal sequencing of blotted reactive proteins and searching public databases for alignments. Moreover, these studies have been performed using tomato cultivars which are not very common and spread across local markets, which might not contain representative widespread LTP isoforms, which are commonly recognized by IgE of tomato allergic patients.

This chapter describes preliminary results to the characterization of tomato LTP isoforms, isolated from tomatoes purchased in the market, which have been achieved by two main approaches:

Part I: purification of LTP in tomato peel extracts by using chromatographic systems working at high pressure (fast-protein LC) and ion exchange chromatography, followed by characterization of the obtained protein by LC/MS using a bottom-up approach and analysis of its resistance to gastrointestinal digestion;
Part II: isolation of LTP in peel, pulp and seed extracts by using ultracentrifuge devices with molecular cut off able to retain proteins with 10 kDa, followed by characterization of retained proteins by LC/MS, to investigate the occurrence and the localization of allergenic tomato nsLTP isoforms.

3. Experimental part – Part I

3.1 Protein extraction from tomato peel

3.1.1 Chemicals

- 2-(*N*-morpholino)ethanesulfonic acid (MES) (Sigma Aldrich, Usa)
- Ethylenediaminetetraacetic acid ($C_{10}H_{16}N_2O_8$) (EDTA) (Sigma Aldrich, Usa)
- Diethyldithiocarbamic acid (DIECA) (Sigma Aldrich, Usa)
- Sodium azide (NaN_3) (Sigma Aldrich, Usa)
- Polyvinylpolypyrrolidone (PVPP) (Sigma Aldrich, Usa)
- Polyvinylpyrrolidone (PVP) (Sigma Aldrich, Usa)
- Tween 20 (Sigma Aldrich, Usa)
- Sodium Hydroxide (NaOH) (Sigma, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Liquid nitrogen
- MES buffer, preparation of 1 L solution:
  - 3.8 g MES (20 mM)
  - 0.7 g EDTA (2 mM)
  - 2.3 g DIECA (10 mM)
  - Adjust pH to 7 with NaOH
  - 0.2 g NaN_3 (3 mM)
  - Add Milli Q H_2O to 1 liter

3.1.2 Instrumentation

- S2896 B & T Flatspin 12V magnetic stirrer
- Digital scale U4600P Universal (Sartorius, Gemany)
- Stirrer on stand RZR 2050 electronic (Heidolph, Germany)
- J6-MI Centrifuge (Beckman Coulter, UK)
- CD 720 pH meter (WPA ltd, UK)
- 3410 Electrochemistry Analyzer (Jenway, UK)

3.1.3 Procedure

3 kg of common grapevine tomatoes were peeled accurately by a cutter and frozen in liquid nitrogen, immediately. After 420 g of peels have been poured in a blender, liquid nitrogen was added till covering the materials entirely and, after plugging the

blender with paper (to prevent powder outflow) and with its lid, peels were ground for 3' or until peel became a fine powder. The content was then moved in a steel container and it was allowed to cool down.
Separately, PVPP (4%), PVP (2%) and Tween 20 (0.2%) were weighed in a 2 L beaker and 500 mL of MES buffer were added.
The whole MES buffer was poured to the container containing the peel powder and it was stirred with a trowel, gently. The mixture was poured in the blender again and it was mixed for 3'. The content was transferred to the steel container and was allowed to stir for 15' in cold room (r.p.m. 28, 4 °C). Afterwards, the mixture was split into two 500 mL centrifuge tubes and the content was centrifuged for 45' at 5 °C (4200 r.p.m). The supernatant was poured into a 2 L beaker and its pH was brought up to 6 by adding 5 M NaOH. After that the conductivity of both sample and MES buffer were checked, the sample was diluted with Milli Q H_2O in order to lower its conductivity closer to the one of MES Buffer.

3.2 Flash ion exchange preparative column

3.2.1 Chemicals

- Streamline™ SP XL in 0.2 M sodium acetate and 20% ethanol (Amersham Pharmacia Biotech, Sweden)
- 2-(*N*-morpholino)ethanesulfonic acid (MES) (Sigma Aldrich, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Sodium Hydroxide (NaOH) (1 M solution in Milli Q H_2O) (Sigma Aldrich, USA)
- Sodium Chloride (NaCl) (Sigma Aldrich, USA)
- MES buffer, 2 L solution preparation:
  - 7.8 g MES (20 mM) in 1900 mL Milli Q H_2O
  - Adjust pH to 6 with NaOH 1 M
  - Add Milli Q H_2O to 2 L
- Elution buffer, 500 mL solution preparation:
  - 29.2 g NaCl (1 M) in 400 mL of 20 mM MES buffer
  - Add MES buffer (20 mM) to 500 mL

3.2.2 Instrumentation

- 501 U peristaltic pump (Watson-Marlow, UK)
- S2896 B & T Flatspin 12V magnetic stirrer
- CD 720 pH meter (WPA ltd, UK)
- Sonifier 2510 bath (Branson Ultrasonics Corporation, USA)

3.2.3 Procedure

100 mL of Streamline™ SP XL resin were placed in a 500 mL bottle and washed several times with Milli Q H_2O, removing the stocking solution after the resin was settled down. Milli Q H_2O was replaced by 1 M NaOH solution and the resin was sonicated in order to break air bubbles and to remove bacteria and lipids. NaOH solution was removed and several washes with Milli Q H_2O were repeated. Then the resin was filtered on a Buchner and rinsed with Milli Q H_2O until its pH was about 7

(checked roughly with a litmus paper). The resin was packed into a funnel of a vacuum flask. MES buffer was allowed to pass through the resin applying vacuum and, when pH of the eluate was about 6, the sample was filtered. Afterwards, the resin was washed with MES Buffer for three times and then it was re-packed into a proper chromatographic column. After the resin settled down, the elution buffer was pumped into the column, at first slowly (speed 50), in order to avoid to break the surface of the resin, then its flow was increased (speed 110). Fractions of the eluate of 10 mL each were collected in 15 mL Falcon tubes and stored at 4 °C until the analysis time.

3.3 Protein fractions reading by UV spectrometer

3.3.1 Chemicals

- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- MES buffer (20 mM), prepared as described in paragraph 3.2.1
- Ethanol (CH_3CH_2OH) (Sigma-Aldrich, USA)

3.3.2 Instrumentation

- Lambda 35 UV/VIS spectrometer (Perkin Elmer, USA)
- UV winlab software (Perkin Elmer, USA)
- Glass fluorometric cuvettes

3.3.3 Procedure

Samples to be tested were allowed to cool down at room temperature.
The sides of the cuvettes which were going to be exposed to the laser beam were wiped with paper and ethanol. Two cuvettes were filled with 20 mM MES buffer in order to read the blank sample. One of the cuvette was then emptied from the MES buffer and filled with the first protein fraction, until the liquid reached cuvette's shoulder. The reading was performed at λ = 280 nm and the OD were recorded. This procedure was repeated for all the protein fractions.
In the end, all the OD measures were exported to an Excel sheet to display the fractions containing proteins directly, using 'XY scatter' graphic option. All fractions with OD ≥ 0.5 were kept and pooled together for further analysis.

3.4 Protein extract concentration by diafiltration system

3.4.1 Chemicals

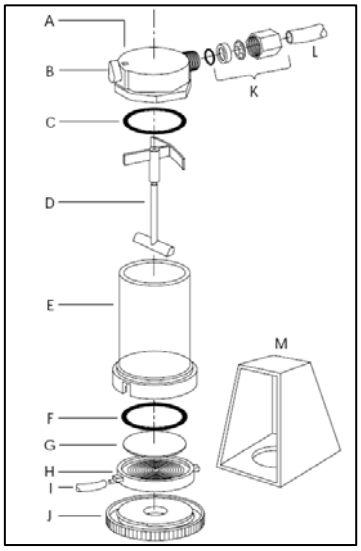
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- N_2 Flux

3.4.2 Instrumentation

- Stirred Cell Model 8200 (Millipore, USA)
- Membrane Ultracel Ultrafiltration Disc, 1 kDa NMWL (Millipore, USA)
- S2896 B & T Flatspin 12V magnetic stirrer
- Millex Filter Units 0.22 µm (Millipore USA)
- 10 mL disposable syringes

3.4.3 Procedure

The membrane was soaked in Milli Q H_2O until the crystals came to the surface of the membrane. The stirred cell was assembled as follows:



- Fit membrane holder into cell body, aligning tabs on sides of holder with slots in base of cell body.
- Invert cell body and membrane holder; screw base firmly into bottom of cell body.
- Push filtrate exit tubing onto exit spout of membrane holder
- Place stirrer assembly into cell body. When properly installed, arms of stirrer assembly will rest on small ridge inside top of cell body.
- Pour sample into cell.
- Rinse the cover with distilled water. Push cap assembly down onto cell body using a twisting motion, orienting gas inlet port on cap opposite filtrate exit port on holder.
- Once assembled, slide cell into retaining stand and starting stir.
- Link the line from the N_2 container to pressure relief valve and elastomeric tubing into a measuring cylinder.
- Open of a quarter of round the safety valve of the N_2 container.
- Open the N_2 cylinder valve until the pressure reaches 35 psi.
- Close the safety valve of the N_2 container.
- Solutes above the membrane molecular weight cut-off are retained in cell, while water and solutes below the cut-off pass into the filtrate and out of cell.
- Bring volume down in order to concentrate the sample at least 5 times.
- Close the N_2 cylinder valve.
- Open of a quarter of round the safety valve of the N_2 container.
- Stirring for 10' to let the gas dissolve in the liquid.
- Stop the stirring.
- Push cap down, then slide cell out from retaining stand.
- Recover the retentate and store it at 4 °C for further analysis.

After concentration, if samples were not clear enough, they could be further filtered by 0.22-µm filter with a syringe.

3.5 Size Exclusion Chromatography analysis

3.5.1 Chemicals

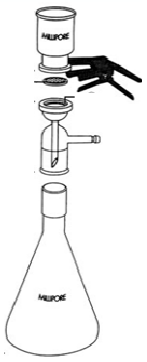
- Sodium phosphate monobasic monohydrate ($NaH_2PO_4 \cdot H_2O$) (Sigma Aldrich, USA)
- Sodium Chloride (NaCl) (Sigma Aldrich, USA)
- Sodium Hydroxide (NaOH) (Sigma Aldrich, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)

- Ethanol (CH_3CH_2OH) (20% solution in Milli Q H_2O) (Sigma Aldrich, USA)
- Chromatographic buffer, 2 L solution preparation:
  - 6.89 g $NaH_2PO_4 \cdot H_2O$ (25 mM)
  - 17.52 g NaCl (150 mM)
  - Adjust the pH to 7.0 with NaOH solution
  - Add Milli Q H_2O to 2 liter
- Gel Filtration Standards (Bio-Rad, Germany)
- Apropotin (Sigma-Aldrich, USA)

3.5.2 Instrumentation

- CD 720 pHmeter (WPA ltd, UK)
- S2896 B & T Flatspin 12V magnetic stirrer
- Vacuum filter (Millipore, USA)
- Durapore® Membrane Filters 0.22 μm (Millipore, USA)
- Akta Chromatographic System (Amersham Pharmacia Biotech, Sweden)
  - Pump P-920
  - Valve INV-907
  - Monitor UPC-900
  - Mixer M-925
  - Monitor *UV-900*
  - Fraction Collector Frac-950
- HiLoad™ Superdex™ 75 prep grade column, 16 mmD x 600 mmL (GE Healthcare, UK)
- Unicorn v 3.20 Software (GE Healthcare, UK)

3.5.3 Procedure



Before being used, the chromatographic buffer was filtered using a vacuum filter, assembled as showed in the picture on the side.
Both the pumps of the AKTA system were washed with 20% ethanol, so that the liquid behind and in front the pistons were replaced by ethanol. All the system, including the column, was cleaned by flowing NaOH 1 M. After washing with Milli Q H_2O, the column was equilibrated with the phosphate chromatography buffer for 2 hours (flow rate: 1 mL/min, pressure: 0.7 MPa).
The first run was performed using a mixture of molecular weight standard diluted 1:1 in Milli Q H_2O. It contained: bovine thyroglobulin (M_r 670 kDa), bovine gamma globulin (M_r 158 kDa), chicken ovalbumin (M_r 44 kDa), horse myoglobin (M_r 17 kDa), Aprotinin (M_r 6.5 kDa), vitamin B12 (M_r 1.35 kDa).
Before loading the standards, 1 mL of Milli Q H_2O was injected in order to wash the loop. Then 400 μL of standard mixture were injected, at 1 mL/min flow rate, with an isocratic gradient of 2 column volume. After the separation of standard mixture was achieved, a 'super-loop' of 50 mL capacity was linked to the instrument and 46 mL of the concentrated sample were loaded to be separated in multiple runs, at the same conditions used for the standard. The volume injected for each run was 4 mL and 1 mL of sample was collected for each fraction. The eluent was monitored for LTP by

following the absorbance at 220 nm. Data were collected and analyzed by Unicorn v 3.20 Software.

3.6 Protein fractions analysis by SDS-PAGE

3.6.1 Chemicals

- NuPAGE® Novex 12% Bis-Tris Gel 1.0 mm, 12 well (Invitrogen, USA)
- NuPAGE® LDS Sample Buffer 4x (Invitrogen, USA)
- SeeBlue® Plus2 Pre-Stained Standard (Invitrogen, USA)
- Mark12™ Unstained Standard (Invitrogen, USA)
- SimplyBlue™ SafeStain (Invitrogen, USA)
- NuPAGE® MES SDS Running Buffer 20x (Invitrogen, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Deionised H_2O
- Dithiothreitol DTT ($HSCH_2(CHOH)_2CH_2SH$) (0.5 M solution in Milli Q H_2O) (Sigma Aldrich, USA)
- Methanol (CH_3OH) (Sigma Aldrich, USA)
- Acetic Acid (CH_3COOH) (Sigma Aldrich, USA)

3.6.2 Instrumentation

- Digital scale 1602 MP Analytical (Sartorius, Gemany)
- Zoom Dual Power, 100-120 VAC 47 – 60 Hz (Invitrogen, USA)
- XCell SureLock™ Mini-Cell Electrophoresis System (Invitrogen, USA)
- S2896 B & T Flatspin 12V magnetic stirrer
- Universal Shaker Swip KS10 (Edmund Buhler, Germany)
- GS-800 calibrated densitometer (Bio-Rad)
- Heating Block QBT1 (Grant, UK)
- Autovortex

3.6.3 Procedure

i. Protein sample preparation

- LDS Sample Buffer 4x diluted 4 times (25 µL)
- DTT 0.5 M solution (10 µL) or, alternatively, 10 µL of Milli Q H_2O when the electrophoresis was performed under native conditions
- 65 µL of sample
- 10' incubation at 70 °C (not performed in case of native conditions analysis)

ii. Electrophoretic run

After removing the gel from its packaging by scissors, it was washed in deionised water and the comb was removed. Wells were rinsed with 1 mL of MES SDS Running Buffer diluted 20 times for three times. Then the gel was placed in the tank, which was locked with the appropriate device and half-filled with diluted MES SDS Running Buffer. 10 µL of each sample, prepared as described above, were loaded in each well and 7 µL of SeeBlue® Plus2 Pre-Stained Standard were used. Afterwards, the tank was

filled completely with the running buffer, without overflowing the upper part of the gel. The voltage applied to the XCell was 200 V. The run lasted 35'.

iii. Gel staining

When the run was completed, the gel was pull out of the tank and its supports were separated by a spatula. The gel was placed in a tray with 100 mL of fix solution (50% CH_3OH, 10% CH_3COOH in Milli Q H_2O) for 30' on shaker. After the fixing step, gel was rinsed with deionised H_2O for four times, 5' each. A measure of SimplyBlue™ SafeStain was applied to the gel and it was allowed to incubate in agitation on the shaker until achieving the desired contrast. Finally, the gel was rinsed with deionised H_2O for 15' and then it was acquired with GS-800™ Calibrated Densitometer.

3.7 Dialysis

3.7.1 Chemicals

- MES buffer (20 mM), prepared as described in paragraph 3.2.1
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)

3.7.2 Instrumentation

- Spectra/Por molecular porous membrane tubing (diameter: 29 mm, Vol/lenght: 6.4 mL/cm) (Spectrum Laboratories, Inc)
- S2896 B & T Flatspin 12V magnetic stirrer
- CD 720 pH meter (WPA ltd, UK)
- 3410 Electrochemistry Analyzer (Jenway, UK)

3.7.3 Procedure

Dialysis buffer volume to be prepared depended on the total volume of the sample: if it is under 100 mL, 2 L of dialysis buffer were enough. In order to calculate the length of the dialysis tube, the volume of the sample to be dialyzed must be known. The total volume of the tube should be 2.5 times the one of the sample. After that, the vol/lenght has to be calculated as follows:

6.4 mL: 1 cm= total volume: x (length of tube)

After the tube has been cut by scissors, it was soaked in Milli Q H_2O for a couple of minutes, then tied an end and filled with Milli Q H_2O (just to check if any hole was present). Then the sample was poured into the tube by a funnel. Air bubbles were eliminated from the sample and, after that, also the other end was tied. Finally, the tube was allowed to stirrer in a beaker containing 20 mM MES buffer for 2 hours. After the dialysis was achieved the sample was poured into a Duran bottle and its conductivity was checked.

3.8 Ion Exchange Chromatography

3.8.1 Chemicals

- Sodium Chloride (NaCl) (Sigma Aldrich, USA)
- Solvent A: MES buffer (20 mM), prepared as described in paragraph 3.2.1

- Solvent B: MES buffer (20 mM solution in Milli Q H_2O, pH 6.0) + Sodium Chloride (1.0 M solution in Milli Q H_2O)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)

3.8.2 Instrumentation

- CD 720 pH meter (WPA ltd, UK)
- S2896 B & T Flatspin 12V magnetic stirrer
- Vacuum filter (Millipore, USA)
- Durapore® Membrane Filters 0.22 µm (Millipore, USA)
- AKTA Chromatographic System (Amersham Pharmacia Biotech, Sweden)
  - Pump P-900
  - Valve INV-907, PV-908
  - Monitor UPC-900
  - Mixer M-925
  - Monitor *UV-900*
  - Monitor pH/C-900
  - Fraction Collector Frac-950
- POROS® 20 HS Perfusion Chromatography column, 4.6 mmD x 100 mmL (GE Healthcare, UK)
- Unicorn v 5.01 Software (GE Healthcare, UK)

3.8.3 Procedure

Before being used, the Solvent A and the Solvent B were filtered using a vacuum filter, assembled as described in paragraph 3.5.3. 5 mL of dialyzed sample were applied to the cation exchange column POROS® 20 HS Perfusion Chromatography attached to an AKTA Chromatographic System, which has been previously equilibrated with solvent A. Injections were repeated until the depletion of the dialyzed sample. After washing with the Solvent A (1 column volume) to remove unbound proteins, the bound proteins were eluted using 2 volume column, 0-0.4 M NaCl gradient in equilibration buffer. The flow rate was set at 3 mL/min and the pressure limit applied to the column was 12.50 MPa. The eluent was monitored for protein following the absorbance at 220 and 280 nm. Fractions containing proteins of interest were detected by SDS-PAGE analysis, as previously described, and afterwards they were pooled together to be further analyzed.

3.9 BCA (Bicinchoninic Acid) assay for protein quantification

3.9.1 Chemicals

- PBS buffer tablet in 200 mL of Milli Q H_2O (Sigma Aldrich, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Bicinchoninic Acid Kit (Sigma Aldrich, USA):
  - Bicinchoninic Acid Solution

- 4% (w/v) $CuSO_4$ • 5 H_2O Solution
- BSA Standard Solution

3.9.2 Instrumentation

- Microtiter plates (Bio-Rad, Germany)
- Excella™ E24 benchtop incubator shaker (New Brunswick Scientific, USA)
- Benchmark Plus Microplate Spectrophotometer (Bio-Rad, Germany)
- Microplate Manager Software 6.0 (Bio-Rad, Germany)
- GraphPad Software v 5.0

3.9.3 Procedure

i. Working standard preparation

BSA protein standard (1.0 mg/mL) was allowed to warm up at room temperature and it was used to prepare the working standards as follows:

Working protein standard	Volume of 1.0 mg/mL BSA standard	Volume of PBS solution
0.0 µg/mL	0 µL	250 µL
200 µg/mL	50 µL	200 µL
400 µg/mL	100 µL	150 µL
600 µg/mL	150 µL	100 µL
800 µg/mL	200 µL	50 µL
1000 µg/mL	250 µL	0 µL

ii. BCA assay working solution

The BCA assay working solution was prepared by mixing a 1:50 solution of 4% $CuSO_4$ • 5 H_2O in bicinchoninic acid solution.

iii. BCA assay performance

25 µL of standards and samples were placed in triplicates into wells of the micro-titre plate and the position of each of them was recorded. 200 µL of the BCA working solution was added into the wells containing standards and samples and the plate was wrapped in cling-film to prevent evaporation. The micro-titre plate was then placed in incubator at 37 °C, 120 r.p.m. for 30' to allow colour development to occur. The absorbance of the samples was measured at 562 nm using the spectrophotometer. Data were recorded and analyzed by GraphPad Prism 5 which allowed to obtain an XY plotter, describing the OD of absorbance as function of sample concentration.

3.10 Reversed Phase Chromatography

3.10.1 Chemicals

- Trifluoroacetic acid (TFA) (Sigma-Aldrich, USA)
- Acetronitrile (ACN) (Sigma-Aldrich, USA)
- Eluent A: Milli Q H_2O + 0.1% TFA
- Eluent B: 90% ACN + 10% Milli Q H_2O + 0.085% TFA

- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- β-lactoglobulin 1 mg/mL (containing A and B β-lactoglobulin) (Sigma-Aldrich, USA)

3.10.2 Instrumentation

- Vacuum filter (Millipore, USA)
- Durapore® Membrane Filters 0.22 μm (Millipore, USA)
- Dionex HPLC system (Dionex Corporation, UK):
  - Pump P-580
  - Automated Sample Injector ASI-100
  - Variable Wavelength Detector Ultimate 3000
- Jupiter 5 μm, C4, 300Å, 4.6 mmD x 250 mmL column (Phenomenex, UK)
- Chromeleon 6.80 Chromatography Data System Software (Dionex Corporation, UK)

3.10.3 Procedure

The solutions were filtered as described in paragraph 3.5.3. The reversed phase HPLC of pooled fractions containing protein of interest was performed by injecting 25 μL of protein sample onto an HPLC column at a 1 mL/min flow rate. Separation was carried out at 25 °C. Proteins were eluted by a linear gradient of solvent B from 0 to 100% in 75' and the elution was monitored following the absorbance at 220 nm. The efficiency of these separation conditions were previously tested on a β-lactoglobulin mixture, containing both A and B β-lactoglobulin genetic variants, which differ for two substitutions: 64th [Asp (A)→ Gly (b)] and 118th [Val(A) → Ala(B)]. The high resolution of the separation of the mixture gave a good evidence on the suitability of the chromatographic method.

3.11 Analytical Size Exclusion Chromatography

3.11.1 Chemicals

- Sodium phosphate monobasic monohydrate ($NaH_2PO_4 \cdot H_2O$) (Sigma Aldrich, USA)
- Sodium Chloride (NaCl) (Sigma Aldrich, USA)
- Sodium Hydroxide (NaOH) (Sigma Aldrich, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Chromatographic buffer, 2 L solution prepared as described in paragraph 3.5.1
- Gel Filtration Standards (Bio-Rad, Germany)
- Apropotin (Sigma-Aldrich, USA)

3.11.2 Instrumentation

- CD 720 pH meter (WPA ltd, UK)
- S2896 B & T Flatspin 12V magnetic stirrer
- Vacuum filter (Millipore, USA)
- Durapore® Membrane Filters 0.22 μm (Millipore, USA)
- AKTA Chromatographic System (Amersham Pharmacia Biotech, Sweden)
  - Pump P-900
  - Valve INV-907, PV-908
  - Monitor UPC-900
  - Mixer M-925
  - Monitor *UV-900*
  - Monitor pH/C-900
  - Fraction Collector Frac-950
- Superdex™ 75 HR analytical column, 10 mmD x 300 mmL (GE Healthcare, UK)
- Unicorn v 5.01 Software (GE Healthcare, UK)

3.11.3 Procedure

The chromatographic buffer was prepared as described in paragraph 3.5.3. 400 μL of purified protein were analyzed on Superdex™ 75 HR analytical column, equilibrated and eluted with 25 mM NaH2PO4 buffer, 150 mM NaCl pH 7.0 with an isocratic gradient of 1.2 column volume, at a flow rate of 0.4 mL/min. The column was calibrated with a set of gel filtration molecular weight standards and, additionally, with apropotin. The absorbance was monitored at 280 and 220 nm and data collected by Unicorn Software.

3.12 Freeze-drying of purified protein

3.12.1 Chemicals

- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- N_2 Flux

3.12.2 Instrumentation

- Miracloth
- Stirred Cell Model 8200 (Millipore, USA)
- Membrane Ultracel Ultrafiltration Disc, 1 kDa NMWL (Millipore, USA)
- S2896 B & T Flatspin 12V magnetic stirrer
- Millex Filter Units 0.22 μm (Millipore USA)
- 10 mL disposable syringes
- Freeze-Dryer 3.5 (Birchover Instruments Ltd, UK)

3.12.3 Procedure

Solution containing purified protein was concentrated using the Stirred Cell Model 8200, as described in the paragraph 3.4.3. As the starting solution was concentrated 5

times, a dialysis was carried out, switching the flux from N_2 to water, in order to replace the MES buffer, in which the sample has been since the extraction, with Milli Q H_2O, to allow further analysis.
After the dialysis, the sample was concentrated again, using the same stirred cell system, this time switching the flux from water to N_2 again. As soon as the desired volume was achieved, the sample was pour in a 20 mL Sterilin tube (previously weighed) and the membrane disc was rinsed with 2 mL of Milli Q H_2O to recover any protein aggregate, which were then added to the recovered sample. After concentration, if samples were not clear enough, they could be further filtered by 0.22-µm filter with a syringe. Afterwards, the tube was covered with Miracloth paper and placed in the freeze-dryer for 24 hours. The net weight of purified protein was calculated by weighing the lyophilized material and subtracting the tare.

3.13 Circular dichroism

3.13.1 Chemicals

- Sodium phosphate monobasic monohydrate ($NaH_2PO_4 \cdot H_2O$) (10 mM solution in Milli Q H_2O) (Sigma Aldrich, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Ethanol EtOH (CH_3CH_2OH) (Sigma-Aldrich, USA)
- N_2 flux

3.13.2 Instrumentation

- Jasco J-170 Spectropolarimeter (Jasco, Tokyo, Japan)
- Jasco J-170 Spectropolarimeter Power Supply (Jasco, Tokyo, Japan)
- Jasco CD software (Jasco, Tokyo, Japan)
- 0.5 mm-path length quartz cell (Hellma, UK)

3.13.3 Procedure

Purified protein structure was determined using circular dichroism. The cell was rinsed with EtOH, first; then it was rinsed in Milli Q H_2O, EtOH again, dried outside with paper and inside with N_2 flux. Before analyzing protein solution, a run with air and a run with CD standards were performed, in order to set up the instrument. A blank run (with 10 mM NaH_2PO_4 phosphate buffer) was also performed, in order to subtract the background noise. Two different concentrations of protein solution in 10 mM NaH_2PO_4 phosphate buffer were tested (0.74 mg/mL and 0.25 mg/mL). Far-ultraviolet (UV) CD spectra (260-180 nm) were collected. Spectra represented the average of three accumulations collected at 50 nm/min with a 2-s time constant, 0.5 nm resolution and sensitivity of ± 100 mdeg. Molar CD was calculated using CD Pro. Secondary structure was predicted using CDSSTR and SELCON3 and the prediction were averaged, as the prediction were similar.

3.14 In-gel tryptic digestion

3.14.1 Chemicals

- Ammonium bicarbonate (NH_4HCO_3) (Sigma-Aldrich, USA)

- Acetonitrile (CH_3CN) (Fisher Scientific, UK)
- Dithiothreitol ($HSCH_2(CHOH)_2CH_2SH$) DTT (Sigma-Aldrich, USA)
- 2-Iodoacetamide IAA (CH_2CONH_2) (Sigma-Aldrich, USA)
- Trypsin Gold (Promega, UK)
- Acetic Acid (CH_3COOH) (50 mM solution in Milli Q H_2O) (Fisher Scientific, UK)
- Formic Acid (HCOOH) (Fisher Scientific, UK)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Solution A: 400 mM NH_4HCO_3 in Milli Q H_2O
- Solution B: 10 mL of solution A in 25 mL of CH_3CN
- Solution C: 50 mM NH_4HCO_3 in Milli Q H_2O
- Solution D: 10 mM DTT in solution C
- Solution E: 100 mM 2-Iodoacetamide in solution C
- Solution F: 10 mM NH_4HCO_3 in Milli Q H_2O
- Solution G: 5 µg of trypsin in 50 mM acetic acid
- Solution H: 1% formic acid in Milli Q H_2O

3.14.2 Instrumentation

- PCR thermo cycler

3.14.3 Procedure

Gel spots of interest were excised using a 5000 µL tip. The tip was placed into a tube for proteomic analysis and 80 µL of Milli Q H_2O were added. The gel plugs were released off the tip by squirting a teat for Pasteur pipette.

Gel plugs were washed in 200 µL of solution B for 15' twice, to equilibrate to about pH 8 and to remove the staining. Solution B was removed with Pasteur pipette and a brief wash with 100 µL of CH_3CN was performed to remove aqueous solutions. CH_3CN was removed with a Pasteur pipette and gel plugs were shrunk by washing in 100 µL of CH_3CN for 10'. The solution was removed and gel plugs were air dried for 10' to remove all the CH_3CN. 100 µL of solution D were added to each tube and they were allowed to incubate for 30' at 60 °C. The solution D was removed, without shrinking the gel plugs, and 100 µL of solution E were then added. Tubes were allowed to incubate for 30' at room temperature in dark. The solution E was removed and it was replaced with 200 µL of solution B for 15' for three times. After the solution B was removed, gel plugs were briefly rinsed with 100 µL of CH_3CN, at first; then they were shrunk again by washing in 100 µL of CH_3CN for 10'. The CH_3CN was eliminated and the gel plugs were air dried for 10'. 5 µL of solution G were added to gel pieces in solution F and samples were incubated at 37 °C for 3 hours with tube lids on to limit condensation. After the digestion step, 10 µL of solution H were added, the tubes were agitated and allowed to stand for 10'. The samples were flash freezed in dry ice and then stored at -70 °C until needed.

3.15 MALDI -TOF analysis

3.15.1 Chemicals

- a-cyano-4-hydroxycinnamic acid
- Acetone (C_3H_6O) (Fisher Scientific, UK)
- Isopropanol ($CH_3CH(OH)CH_3$) (Fisher Scientific, UK)
- Calibration Standards (Sigma-Aldrich, USA)

3.15.2 Instrumentation

- Reflex III MALDI-ToF mass spectrometer (Bruker UK Ltd., Coventry)
- MASCOT v 1.9 (IFR and JIC Joint Proteomic Facilities)
- mMass software v 3.9

3.15.3 Procedure

The acidified digests were spotted directly onto a thin layer of matrix on a stainless steel target plate. The matrix consisted of the following: four parts of a saturated solution of a-cyano-4-hydroxycinnamic acid (CCA) in acetone was mixed with one part of a 1:1 mixture of acetone:isopropanol containing 10 mg/mL nitrocellulose. Digests were externally calibrated against a calibration curve of the following peptides to yield data with mass accuracies of better than 50 ppm. Standards and their monosoitopic masses were: Angiotensin II (1046.5423 Da), Angiotensin I (1296.6900 Da), Substance P (1347.7359 Da), Bombesin (1619.8229 Da), Adrenocorticotropic Hormone Clip 1-17 (2093.0900 Da), Adrenocorticotropic Hormone Clip 18-39 (2465.2027 Da), Somatostatin (3147.4700 Da). Analysis of peptide digests was carried out on a Reflex III MALDI-ToF mass spectrometer. A nitrogen laser was used to desorb/ionise the matrix/analyte material, and ions were detected in positive ion reflectron mode. The calibrated spectra for each sample were searched against a weekly updated copy of the SPtrEMBL database, using an in-house version (v1.9) of the MASCOT search tool.

3.16 LTQ-Orbitrap analysis

LC-MS/MS analysis was performed using a LTQ-Orbitrap mass spectrometer and a nanoflow-HPLC system (nanoAcquity, Waters Corp.). Peptides were applied to a precolumn (Symmetry C18 5μm beads, 180 μm x 20 mm column, Waters Corp) connected to a 25 cm analytical column (BEH 130 C18 1.7 μm beads, 75 μm x 250 mm column, Waters Corp.). Peptides were eluted by a gradient of 5 to 40% acetonitrile in 0.1% formic acid from 1 to 40 min at a flow rate of 250 nL min^{-1}. Mass spectra were obtained in positive ion electrospray mode. The mass range for the survey scans was m/z 300 – 2000, resolution 60,000, with m/z values determined by the Orbitrap FTMS stage. The FTMS fill target was 200,000 ions with a maximum fill time of 1000 ms The resultant monoisotopic masses were accurate to better than 10 ppm. MS/MS spectra were obtained using collision induced dissociation with collision voltage 35 V with m/z values determined by the Linear Ion Trap stage. The MS/MS was triggered by a minimal signal of 5000 ions with a fill target of 10,000 ions and 150 ms maximum fill time with exclusion of 4+ charge states. A maximum of 4 MS/MS spectra per survey scan were obtained by defaulting to the most abundant ions, with

m/z values determined to better than ~ 0.4 Da. Charge state selection was not enabled. Dynamic exclusion was set to 1 count and 60 s exclusion with an exclusion mass window of -0.5 to +1.5 Da.

3.17 In vitro gastrointestinal digestion of purified protein

3.17.1 Chemicals

- Egg lecithin (10 mg, Grade 1) (Lipid Products, Surrey England)
- Sodium chloride (NaCl) (Sigma-Aldrich, USA)
- Chloridric acid (HCl) (1.0 M solution in Milli Q H_2O) (Sigma-Aldrich, USA)
- Sodium hydroxide (NaOH) (1 M, 0.1 M and 0.01 M solutions in Milli Q H_2O) (Sigma-Aldrich, USA)
- Bis-Tris base (Sigma-Aldrich, USA)
- Calcium Chloride ($CaCl_2$) (1 M solution in Milli Q H_2O) (Sigma-Aldrich, USA)
- Dichloromethane (CH_2Cl_2) (Sigma-Aldrich, USA)
- Ammonium bicarbonate (NH_4HCO_3) (Sigma-Aldrich, USA)
- Ethanol (CH_3CH_2OH) (Sigma-Aldrich, USA)
- Methanol (CH_3OH) (Sigma-Aldrich, USA)
- Sodium taurocholate NaTC ($C_{26}H_{44}NO_7S{\bullet}Na$) (2.386 mg/mL in CH_3OH) (Sigma-Aldrich, USA)
- Sodium glycodeoxycholate NaGDOC ($C_{26}H_{42}NNaO_5$) (2.094 mg/mL in CH_3OH) (Sigma-Aldrich, USA)
- Pepsin from porcine gastric mucosa, 3300 u/mg (Sigma-Aldrich, USA)
- Trypsin from porcine pancreas, 13500 u/mg (Sigma-Aldrich, USA)
- α-Chymotrypsin from bovine pancreas, 40 u/mg (Sigma-Aldrich, USA)
- Phenylmethanesulfonyl fluoride (PMSF) (Sigma-Aldrich, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- N_2 flux
- Ar flux

3.17.2 Instrumentation

- Digital scale U4600P Universal (Sartorius, Gemany)
- Digital scale 1602 MP Analytical (Sartorius, Gemany)
- CD 720 pH meter (WPA ltd, UK)
- S2896 B & T Flatspin 12V magnetic stirrer
- Millex Filter Units 0.22 μm (Millipore, USA)
- 10 mL disposable syringes
- Vacuum oven OVL-570-010J (Weiss Gallenkamp, UK)
- Digital Sonifier 250 (Branson Ultrasonics Corporation, USA)
- Excella™ E24 benchtop incubator shaker (New Brunswick Scientific, USA)
- Analytical Balance ME-30 (Mettler Electronics Corp, USA)
- Rotary Evaporator R201B-III (Sunwain Co., Ltd, China)
- TotalLab TL120 - 1D gel image analysis software (Shimadzu Ltd, UK)

3.17.3 Procedure

i. Digestion solutions preparation

- Egg lecithin stock solution

A vial containing 10 mg of egg lecithin was poured in 50 mL round-bottom flask, previously cleaned with ethanol and dried by N_2 flux. The content was made up to 10 mL with CH_2Cl_2. Once prepared, it was flushed with Ar flux and stored at -20 °C.

- Simulated Gastric Fluid (SGF)

8.77g of NaCl were dissolved in 900 mL of Milli Q H_2O. The pH was adjusted to 2.5 using 1.0 M HCl and the volume was brought to 1000 mL. The solution was filtered using a 0.22-µm filter.

- Simulated Duodenal Fluid (SDF)

8.77g of NaCl were dissolved in 900 mL of Milli Q H_2O. The pH was adjusted to 6.5 using 0.01 M NaOH and the volume was brought to 1000 mL. The solution was filtered using a 0.22-µm filter.

- Simulated Duodenal Fluid (SDF) 15x

0.785 g of Bis-Tris base were dissolved in 15 mL of SDF. 187.5 µL of 1 M $CaCl_2$ were added to give 7.5 mM in 25 mL. The pH was brought to 6.5 using 1.0 M HCl and more SDF was added up to 25 mL. The solution was filtered using a 0.22-µm filter and stored at 4 °C.

- 0.5 M ammonium bicarbonate

3.95 g of NH_4HCO_3 were dissolved in 100 mL of Milli Q H_2O and filtered through a 0.22-µm filter.

- Egg lecithin for gastric digestion phase

91.4 µL of egg lecithin stock solution were placed in a 50 mL round-bottom flask. 1 mL of dichloromethane was added and the solution was allowed to dry at the rotary evaporator. After being flushed 4 times with air and three times with Ar, the lecithin was allowed to dry in vacuum oven overnight.

- Egg lecithin for duodenal digestion phase

1 mL of each bile salts solution (NaTC and NaGDOC) were placed into a 20 mL round-bottom flask and 30.8 µL of egg lecithin stock solution were added. The solution was evaporated at the rotary evaporator and, after being flushed 4 times with air and three times with Ar flux, it was dried in vacuum oven overnight.

- Pepsin stock solution preparation

Pepsin solution was prepared in order to be 6628 u/mL, that is 2.008 mg/mL when using 3300 u/mg stock bottle. At least 10-15 mg of pepsin were weighed and made up to 2.008 mg/mL with cold SGF pH 2.5, dissolved by stirring and added to digestion as soon as possible.

- α-Chymotrypsin stock solution preparation

Chymotrypsin solution was prepared in order to be 38.61 u/mL, that is 0.97 mg/mL when using 40 u/mg stock bottle. At least 10-15 mg of chymotrypsin were weighed and made up to 0.978 mg/mL with cold SGF pH 2.5, dissolved by stirring and added to digestion as soon as possible.

- Trypsin stock solution preparation

Trypsin solution was prepared in order to be 3330 u/mL, that is 0.247 mg/mL when using 13500 u/mg stock bottle. At least 10-15 mg of trypsin were weighed and made

up to 0.247 mg/mL with cold SGF pH 2.5, dissolved by stirring and added to digestion as soon as possible.

ii. Gastric digestion phase

The egg lecithin prepared for gastric digestion phase was re-suspended in 10 mL of pre-warmed SGF pH 2.5 for 20' at 37 °C in incubator with 3-4 glass beads as dispersing vehicle. Then the suspension was sonicated until it looked pretty clear, using the following settings:
0.9" on – 0.1" off, for 2', 40% power, repeated 3 times.
2 mg of purified protein were weighed in a 7 mL Sterilin tube with a magnetic flea; 1 mL of SGF pH 2.5 and 600 µL of egg lecithin for gastric digestion were added. After a brief stir, the pH of the solution was adjusted to 2.5 with 1.0 M HCl and its volume was noted. In order to make up 2 mL, a suitable volume of SGF pH 2.5 was added. 50 µL of sample were taken and added to a 0.5 mL eppendorf tube containing 10 µL of 0.5 M NH_4HCO_3 (time 0 of gastric digestion).
50 µL of pepsin stock solution were added to digestion mix and the sample was placed in incubator at 37 °C for 1 h in agitation (120 r.p.m.).
At the end of gastric phase, 50 µL of digestum were collected and transferred to a 0.5 mL eppendorf tube containing 10 µL of 0.5 M NH_4HCO_3 (time 60'). The pH of the remaining solution was adjust to 7.5 with 1 M and 0.1 M NaOH in order to stop the digestion by arresting pepsin activity.

iii. Duodenal digestion phase

1.08 mL of gastric digesta and 80 µL of SDF 15x were added to the lecithin/bile salt mix in the 20 mL round-bottomed flask, flushing with Ar and mix in incubator for 5'. 1.06 mL of the emulsion were taken and added to 7 mL Sterilin tube with small magnetic flea; its pH was adjusted to 6.5 with 1.0 M HCl or NaOH as necessary by adding 1 µL at time using P20 pipette. In order to make up 2 mL, a suitable volume of SDF pH 6.5 was added. After the stock solution of enzymes were prepared, 10 µL of both chymotrypsin and trypsin stock solution were added to the digestion mix. 50 µL of digesta were collected immediately after the addition of the enzymes (time 0 of duodenal digestion) and placed into 0.5 mL eppendorf tube containing 4 µL of 0.1 M PMSF, a proteinases inhibitor to halt the activity of the enzymes. The duodenal digestion mix was placed in incubator at 37 °C in agitation (120 r.p.m.) and several sampling (50 µL each) were taken after 1, 2, 5, 10, 15, 30, 60 and 120' of incubation and transferred to 0.5 mL eppendorf tube containing 4 µL of 0.1 M PMSF. Remaining digestum was placed into a 1.5 mL eppendorf tube containing 40 µL of 0.1 M of PMSF solution.
All the samples obtained after the digestion process were analyzed by SDS PAGE under both denaturing and native conditions, as described in paragraph 3.6, and the profile and the relative quantification of protein and peptides were outlined by TotalLab TL120 software.

3.18 Reduction and alkylation of digested samples for MALDI-TOF analysis

3.18.1 Chemicals

- Ammonium bicarbonate (NH_4HCO_3) (Sigma-Aldrich, USA)
- Dithiothreitol DTT ($HSCH_2(CHOH)_2CH_2SH$) (Sigma-Aldrich, USA)
- 2-Iodoacetamide IAA (CH_2CONH_2) (Sigma-Aldrich, USA)
- Milli Q H_2O obtained with Barnstead Nanopure Diamond Uv Toc D11951 (Thermo Scientific)
- Digestion buffer: 50mM NH_4HCO_3 in Milli Q H_2O
- Reducing buffer: 50mM DTT in digestion buffer
- Alkylation buffer: 100mM IAA in digestion buffer

3.18.2 Instrumentation

- Digital scale U4600P Universal (Sartorius, Gemany)
- S2896 B & T Flatspin 12V magnetic stirrer
- Heating Block QBT1 (Grant, UK)

3.18.3 Procedure

Digested samples were further analyzed by MALDI-TOF. 65 µL of digestion buffer and 15 µL of reducing buffer were added to 15 µL of each digested sample. They were incubated at 65 °C for 10'. Afterwards, 15 µL of alkylation buffer were added to each reduced sample and were incubated at room temperature in a dark environment.
MALDI-TOF analyses were performed as already describe in paragraph 3.15.

4. Experimental part – Part II

4.1 Protein extraction from tomato peel, pulp and seeds

Protein extractions from tomato peel, pulp and seeds were performed using Piccadilly variety and following the protocol already described in paragraph 3.2, Chapter III - Section I, slightly modified as follows:
Peel: tissues were removed from frozen berries and they were scraped out by a spatula from the inside. A total of 10 g of peel were used for the homogenization in 100 mL of cold acetone, rather than 50 mL, in order to allow the blades of the Ultraturrax to homogenize properly the cellulose tissues.
Seeds: after being removed from the locular cavities of the tomato berries, seeds were placed into a colander and rinsed with Milli Q H_2O to eliminate any residue of pulp tissue. 7 g of whole seeds were reduced to a fine powder by milling in liquid nitrogen before being homogenized in cold acetone.

4.2 Protein fractionation in the range 30-3kDa

4.2.1 Chemicals

- Milli Q H_2O obtained with Millipore Alpha Q system
- Methanol (CH_3OH) (Sigma-Aldrich, USA)

4.2.2 Instrumentation

- Digital Scale BCE 62 PT (Orma, Italy)
- Amicon® Ultra-4 Centrifugal Filter Units Amicon Ultra 3 KDa device — 3.000 NMWL (Nominal Molecular Weight Limit) (Millipore, Italy)
- Amicon® Ultra-4 Centrifugal Filter Units Amicon Ultra 30 KDa device — 30.000 NMWL (Nominal Molecular Weight Limit) (Millipore, Italy)
- Centrifuge Universal 320 R (Hettich, Germany)
- Rotor 1620A (97 mm radius) (Hettich, Germany)
- Adapters 1451 (Hettich, Germany)

4.2.3 Procedure

Pre-rinsing: the ultra-filtration membranes in Amicon Ultra devices contained trace amounts of Polyethylene glycol (PEG) which could interfere with further analyses. For this reason, the devices were pre-rinsed 5 times. Each wash was performed with 4 mL of CH_3OH:H_2O (1:1) solution, at 6500 RCF and at room temperature, for 45' for 3 kDa devices, or for 15' for 30 kDa devices. In order to avoid the breaking of the membranes due to the drying, devices were stocked in CH_3OH:H_2O (5:95) solution, at 4 °C until use.

Fractionation below 30 kDa: 3.5 mL of each tissue protein extract were loaded onto a 30 kDa device and centrifuged at 6500 RCF, 4 °C for 30'. The analyses were performed in duplicate for each tomato tissue, in order to allow the characterization of proteins both in the native and in the denatured conformation. The filtered solutions (about 3.2 mL) were recovered and stored in 15 mL Falcon tubes. In order to obtain a major recovery, membranes were washed twice with 3.5 mL of Milli Q H_2O, under the same conditions, and the filtered solutions were recovered again and pooled to the previous ones.

Fractionation above 3 kDa: each solution containing proteins below 30 kDa was loaded onto a 3 kDa device and centrifuged at 6500 RCF, 4 °C for 45', until the depletion of the whole recovered solution. Two washes with 3.5 mL each of Milli Q H_2O of the retentates (the upper part of the device) were performed using the same centrifugation settings. In this case, the retentate (about 250 µL) was recovered using 250 µL of Milli Q H_2O, placed into a 1.5 mL eppendorf tube and stored at -20 °C.

4.3 Reduction and alkylation of native proteins

4.3.1 Chemicals

- Milli Q H_2O obtained with Millipore Alpha Q system
- Urea (NH_2CONH_2) (Sigma-Aldrich, USA)
- Ammonium bicarbonate (NH_4HCO_3) (Sigma-Aldrich, USA)
- Dithiothreitol DTT ($HSCH_2(CHOH)_2CH_2SH$) (Sigma-Aldrich, USA)
- 2-Iodoacetamide IAA (CH_2CONH_2) (Sigma-Aldrich, USA)
- Solution A: 25 mM NH_4HCO_3 in Milli Q H_2O
- Solution B: 8 M NH_2CONH_2 in solution A
- Solution C: 50 mM DTT in solution A
- Solution D: 50 mM IAA in solution A
- N_2 flux

4.3.2 Instrumentation

- Digital Scale BCE 62 PT (Orma, Italy)
- Hotplate Stirrer F 80 (Falc)
- Digital scale E42 S Analytical (Gibertini, Italy)
- Autovortex SA6 (Stuart Scientific, UK)

4.3.3 Procedure

The volume of all recovered retentates was measured, the same amount of solution B was added and the mixture was briefly vortex. A volume of the solution C was added so that the final concentration of DTT in the sample was 5 M. The mix was stirred and it was incubate at 60 °C for 30'. After cooling down at room temperature, a volume of solution D was added so that the final concentration of IAA in the sample was 15 mM. The mixture was stirred again and allowed to incubate at room temperature for 45' in the dark. In the end, a volume of the solution C was added so that DTT concentration was about 3.33 mM in order to block the activity of IAA. After a brief shake, samples were transferred to 1.5 mL vials and were dried under N_2 stream.

4.4 UPLC/MS analysis of fractionated proteins

4.4.1 Chemicals

- Milli Q H_2O obtained with Millipore Alpha Q system
- Formic Acid FA (HCOOH) (0.1% solution in Milli Q H_2O) (Acros Organics, Belgium)
- Acetonitrile ACN (CH_3CN) (Sigma-Aldrich, USA)
- Eluent A: Milli Q H_2O + 0.1% FA + 0.2% ACN
- Eluent B: ACN + 0.1% FA

4.4.2 Instrumentation

- Autovortex SA6 (Stuart Scientific, UK)
- Acquity Ultra Performance LC (Waters, Italy)
- Acquity UPLC BEH 300, C18 1.7 µm, 2.1 mmD x 150 mmL (Waters, Italy)
- Acquity UPLC BEH 300, C8 1.7 µm, 2.1 mmD x 150 mmL (Waters, Italy)
- Acquity SQ Detector with ESI interface (Waters, Italy)
- Software Mass Lynx 4.1 (Waters, Italy)

4.4.3 Procedure

Protein extracts from tomato peel, pulp and seeds, further fractionated in the range of 30-3 kDa, were characterized by UPLC/MS analysis, which was performed both with the samples containing native proteins and with the samples whose proteins were reduced and alkylated.

Dried samples were re-suspended in 200 µL of 0.1% FA in Milli Q H_2O and briefly mixed by autovortex.

The chromatographic separation of protein extracts and fractions was performed using both a C18 and a C8 column as stationary phase, in order to develop a method which was more effective in performing the separation of small-medium size proteins.

As concerns the chromatographic conditions, they were set as follows:
5 µL of each sample were injected each time, at a 0.20 mL/min flow rate, for C18 column, and at a 0.15 mL/min flow rate for C8 column, because of the higher back pressure. The column oven was kept at 35 °C. Gradient:
7' isocratic elution with 100% A, 43' linear gradient from 0% to 50% of B, 2.60' isocratic elution with 50% A and 50% B, 0.4' linear gradient to reach 100% B, 5.20' isocratic elution with 100% B, 0.8' linear gradient from 100% to 0% of B and 13' of reconditioning with 100% of eluent A.
ESI-MS spectrometer 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.

5. Results and Discussion – Part I

5.1 LTP purification from tomato peel

i. Extraction step

Different protocols were tested to set up the best conditions for the extraction of LTP from tomato peel. The most suitable one was found to be an extraction with a buffer containing 4-Morpholineethanesulfonic acid (MES), a compound having a pK_a around 6, maximum water solubility and minimum solubility in all other solvents, minimal salt effects, minimal change in pK_a with temperature, chemically and enzymatically stable. The buffer was added of polyvinylpirrolidone (PVP) and polyvinylpolypirrolidone (PVPP), polymers which are, respectively, partly soluble in water and Tween 20, a surfactant. The combination of these chemicals allowed soluble proteins to be easily extracted in the buffer, by removing protein-polyphenolic complexes or preventing their formation, and solubilizing membrane proteins. In order to reduce the purification time, the extraction time was reduced from 90' to 30', and also the LTP enriching step by flash HPLC was performed by packing the resin into a funnel and performing a vacuum filtration, thus allowing the crude preparation to transit through the resin faster because of the vacuum. The resin was then transferred to a streamline column and the sample was allowed to elute. The pH of the crude extract and of the MES buffer was about 6, that is lower than the pI of the LTP (~ 9). In this way, basic proteins were positively charged and able to interact with the negatively charged residues of the stationary phase. Then, Na^+ contained in the elution buffer acted as a strong counterion, suitable to elute the proteins linked to the resin.
At the end of the elution, 33 fractions were obtained, 10 mL each. The UV absorbance values of each fraction at 280 nm are plotted in Fig 2.

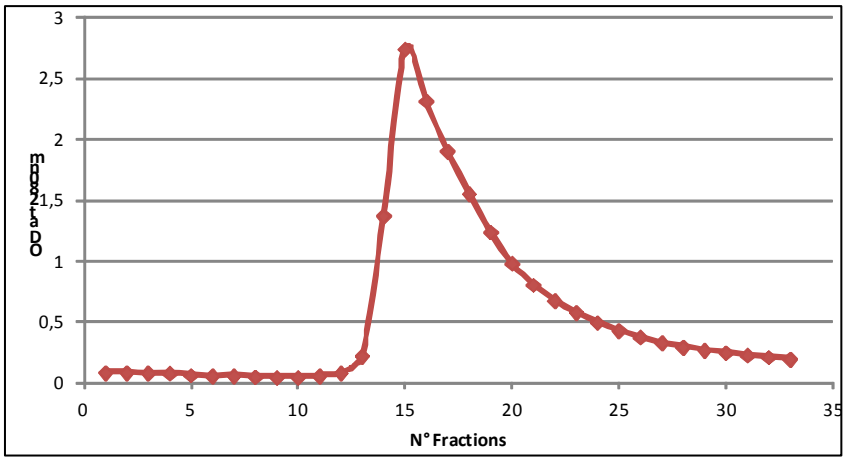


Fig 2. Graphical representation of UV absorbance values of protein fractions off the flash HPLC step of purification.

Fractions from 13 to 33 were pooled together and the whole sample was concentrated about 5 times by diafiltration system, reducing the volume from 220 mL to 46 mL through a 1 kDa nominal molecular weight limit (NMWL) membrane.

ii. Size Exclusion Chromatography

Size exclusion chromatography, also called gel filtration, allows the separation of proteins with differences in molecular size, under mild conditions. Separation is based on differences in the size and/or shape of the analyte molecules, which determines the analytes' access to the pore volume inside the column packing particles. According to their size, smaller analytes have partial to complete access to the pore volume. Among the analytes that partially or fully enter the pore volume, larger molecules with less access to the pore volume elute first, while the smallest molecules elute last.

Thus, protein molecular weight determination by gel filtration can be made by comparing the elution volume of the substance of interest with the ones obtained for several known calibration standards run in the same buffer.

Following this principles, a preparative gel filtration analysis was performed on the LTP enriched extract using a Superdex™ 75 column (16 mmL x 600 mmD). Since the sample does not have to be in exactly the same buffer as that used to equilibrate and run through the column, a low concentration of salt (25 mM phosphate and 150 mM NaCl) was including in the buffer composition, also to avoid nonspecific ionic interactions with the matrix which can be seen as delays in peak elution.

Both samples and buffer were 0.22-µm filtered before being applied to the column, to avoid the risk of blockage due to small particles. A super-loop was loaded with 46 mL of sample and 11 runs of 4 mL each were performed.

The graph showed in Fig 3a represents the overlapped chromatograms related to each run, compared to the chromatogram obtained for the molecular standards in the same conditions.

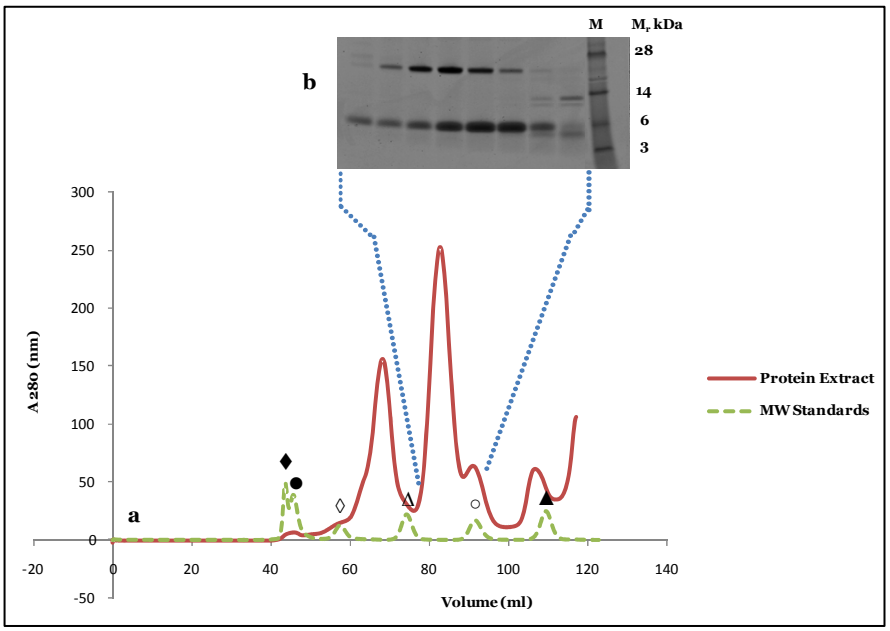


Fig 3. a, chromatogram related to the size exclusion chromatography of the LTP enriched extract. Molecular weight standards were: ♦, Bovine thyroglobulin (M_r 670 kDa); ●, Bovine gamma globulin (M_r 158 kDa); ◊, Chicken ovalbumin (M_r 44 kDa); Δ, Horse myoglobin (M_r 17 kDa); ○, Aprotinin (M_r 6.5 kDa); ▲, Vitamin B12 (M_r 1.35 kDa). **b**, SDS PAGE analysis of fractions containing proteins with molecular weight ranging between 17 kDa and 6.5 kDa. Lane marked 'M' corresponds to molecular weight markers and their relative molecular masses are reported on the right side.

As it can observed from the chromatogram, several peaks were obtained, indicating the presence of several proteins. In order to identify fractions containing LTP, an SDS PAGE analysis was performed on those containing proteins with molecular weight ranging from 17 kDa to 6.5 kDa.
These fractions (Fig 3b) showed a protein band with a relative molecular mass of 6 kDa, which could be an LTP, together with other high molecular weight proteins.
In order to purify the potential LTP, a further chromatographic step was performed, exploiting another chemical features of proteins: the charge.

iii. Ion exchange chromatography

Ion exchange chromatography separates proteins with differences in surface charge to give high-resolution separation. The partition is based on the reversible interaction between a charged protein and an oppositely charged chromatography medium. Typically, when below their isoeletric points, as in this case, proteins will bind to a negatively charged, cation exchanger. Proteins bind as they are loaded onto the column at low ionic strength (20 mM MES buffer pH 6.0). The conditions are then altered so that bound substances are desorbed differentially. Elution is usually performed by increasing salt concentration (from 0 to 0.4 M of 1 M NaCl), so that the presence of Na^+ ions competes with the bound proteins for the binding to the column. In this way, unbound molecules are eluted before the gradient begins, while tightly bound molecules are eluted in a high-salt wash. All the other proteins will elute according to the ion strength during the gradient.
Since the proteins have to be loaded onto the POROS® 20 HS column (4.6 mmD x 100 mmL) at low ionic strength, a dialysis of pooled fractions off the gel filtration was

carried out. After 2 hours of dialysis in 20 mM MES buffer, the conductivity of the sample lowered to 5.86 mS/cm, low enough to be analyzed by a cation exchange chromatography.
5 mL of dialyzed sample were injected onto the column and the injections were repeated until the whole volume (150 mL) was depleted.
Since LTPs are low abundant in aromatic amino acid residues, the eluent was monitored also following the absorbance at 220 nm, in order to detect peptide bounds.

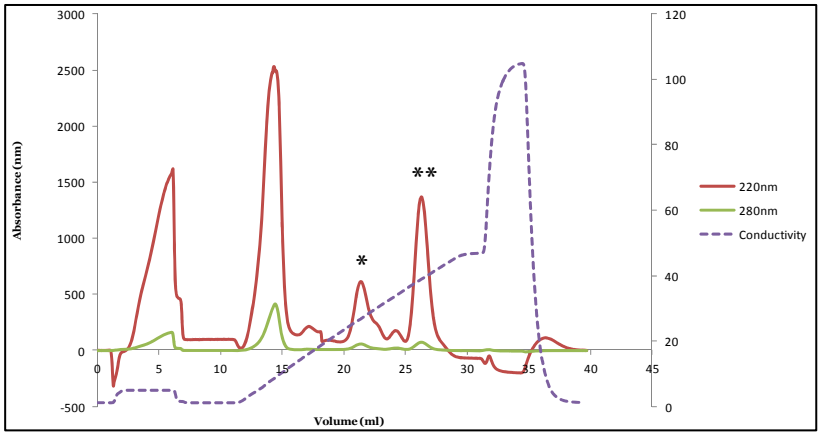


Fig 4. Ion exchange chromatogram of pooled fractions off the gel filtration containing the protein of interest. Stars indicate the range of fractions which were further analyzed by SDS PAGE.

After ion exchange chromatography (Fig 4), peaks eluting during the linear gradient of salt were further investigated by SDS PAGE, in order to understand whether the protein with the expected molecular weight had been purified.

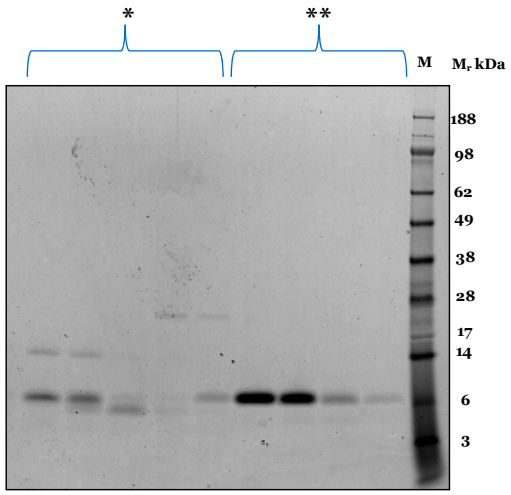


Fig 5. SDS PAGE of fractions off ion exchange chromatography. Stars indicate fractions as referred to in Fig 4. The lane marked as 'M' represents the molecular weight standards and their relative molecular masses are reported on the right side.

As showed by the electrophoresis analysis reported in Fig 5, fractions constituting the peak indicated with '**' revealed the presence of an unique protein band with the

expected molecular weight. So they were pooled and used for purified protein characterization.
Also fractions referred to '*' showed the protein of interest but some other bands, with both higher and lower molecular weights, were also present. Due to their different behavior during the ion exchange chromatography, it was supposed that different tomato LTP isoforms had been detected but, because of their different interactions with the stationary phase, one of them couldn't be purified, since it still co-eluted with other proteins.

iv. Purified protein quantitation

Purified protein was quantified by the bicinchoninic acid (BCA) assay. The principle is similar to the Lowry procedure, since both rely on the formation of a Cu^{2+}-protein complex under alkaline conditions, followed by reduction of the Cu^{2+} to Cu^{1+}. The amount of reduction is proportional to the protein present. It has been shown that cysteine, tryptophan, tyrosine, and the peptide bond are able to reduce Cu^{2+} to Cu^{1+}. BCA forms a purple-blue complex with Cu^{1+} in alkaline environments, thus providing a basis to monitor the reduction of alkaline Cu^{2+} by proteins at absorbance maximum 562 nm.
A calibration curve was obtained by using Bovine Serum Albumine (BSA) standard at increasing concentration, from 0 µg/mL to 1000 µg/mL. The net absorbance at 562 nm was calculated as mean between the three repeats for each standard and sample. A standard curve (Fig 6) was obtained by plotting the net absorbance at 562 nm versus the protein standard concentrations.

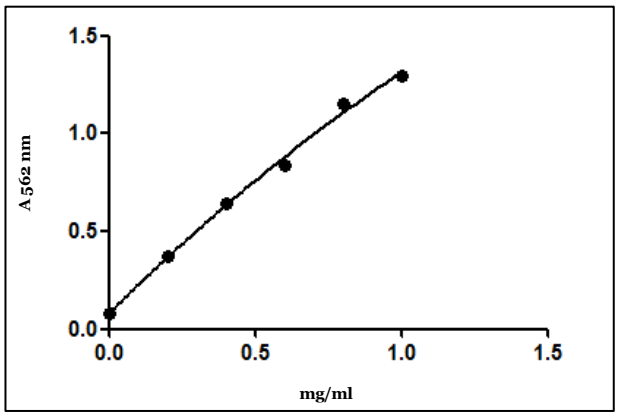


Fig 6. Standard curve obtained by plotting the absorbance values versus the protein standard concentration. The calculated R^2 was 0.99.

Interpolating the values of absorbance obtained for the sample non diluted and diluted 1:1 in Phosphate Buffer Saline (PBS), the protein concentrations were found to be 0.13 mg/mL and 0.065 mg/mL respectively. Accounting for the dilution factor, also the diluted sample gave a value of 0.13 mg/mL for the protein concentration. Considering that the volume of the pooled fraction containing purified proteins was 91 mL, 9.1 mg of pure protein were thus purified. As referred to the amount of starting material (3 kg of whole tomatoes), it can be estimated that the isolation

strategy developed here allowed to obtain a yield for the purified proteins around 3 mg/kg of whole tomatoes.

v. Assessing protein purity

The bands obtained after a SDS PAGE analysis may contain more than a protein, if their molecular weights differ for few Daltons, since they can not be discriminated by gel electrophoresis.

Although a separation according to pH had already been performed, since only basic proteins could have interacted with the stationary phase during the ion exchange chromatography, further analytical checks were carried out in order to value the purity of the isolated protein.

400 µL of the solution containing the purified protein were then applied to a Superdex™ 75 analytical column (10 mmD x 300 mmL) and the run was performed as described in paragraph 3.11.

As shown in Fig 7a, the analytical size exclusion chromatography revealed the presence of two peaks: besides the expected one, between 17 kDa and 6.5 kDa, which was related to the protein of interest, another one was observed whose molecular weight ranged between 6.5 kDa and 1.35 kDa. The last one was proved to belong to the MES buffer by a blank run. So, the analytical gel filtration analysis confirmed the presence of only one protein.

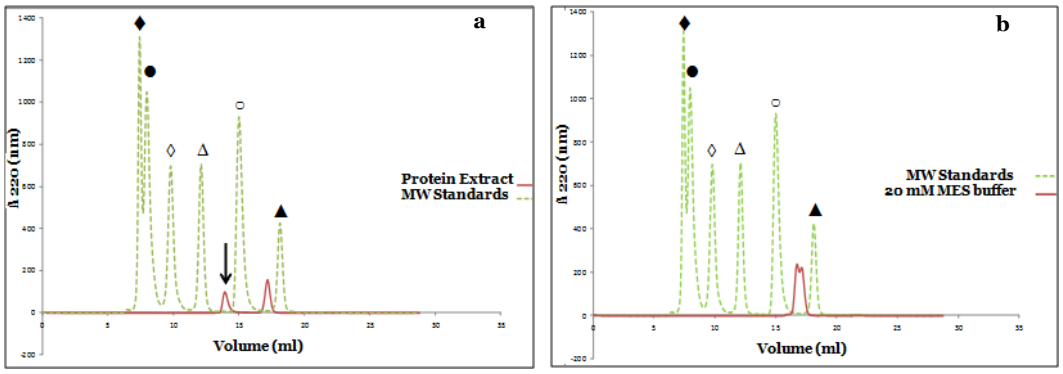


Fig 7. Analytical gel filtration analysis. **a**, chromatogram of purified protein, the arrow indicates the peak related to the protein of interest; **b**, chromatogram of 20 mM MES buffer. Molecular weight standards were: ♦, Bovine thyroglobulin (Mr 670 kDa); ●, Bovine gamma globulin (Mr 158 kDa); ◊, Chicken ovalbumin (Mr 44 kDa); Δ, Horse myoglobin (Mr 17 kDa); ○, Aprotinin (Mr 6.5 kDa); ▲, Vitamin B12 (Mr 1.35 kDa).

To further check the purity of the isolated protein, the highest resolution chromatography technique available was applied: reversed phase chromatography.

In this chromatography, the separation of proteins depends on a reversible hydrophobic interaction between sample molecules in the eluent and the medium. Initial conditions are primarily aqueous, favoring a high degree of organized water structure surrounding the sample molecule. Frequently, a small percentage of organic modifier, is present in order to achieve a "wetted" surface. As sample binds to the medium, the hydrophobic area exposed to the eluent is minimized. Separation relies on sample molecules existing in an equilibrium between the eluent and the surface of the medium. Initially, conditions favor an extreme equilibrium state where essentially 100% of the sample is bound. To bring about elution the amount of organic solvent is

increased so that conditions become more hydrophobic. Binding and elution occur continuously as sample moves through the column. The process of moving through the column is slower for those samples that are more hydrophobic. Consequently, samples are eluted in order of increasing hydrophobicity.

To test the efficiency of the method and of the column, a preliminary separation was performed using a β-lactoglobulin mixture, containing both A and B β-lactoglobulin genetic variants, which differ for two substitutions: 64th (Asp (A)→ Gly (b)) and 118th (Val(A) → Ala(B) (Fig 8a).

The presence of two resolved peaks, indicating the two isoforms of β-lactoglobulin which differ for 86.04 Da, indicated the efficiency of the developed method.

Afterwards, the same method was applied to the purified protein. As shown in Fig 8b, only a peak was obtained, again confirming the purity of the isolated protein.

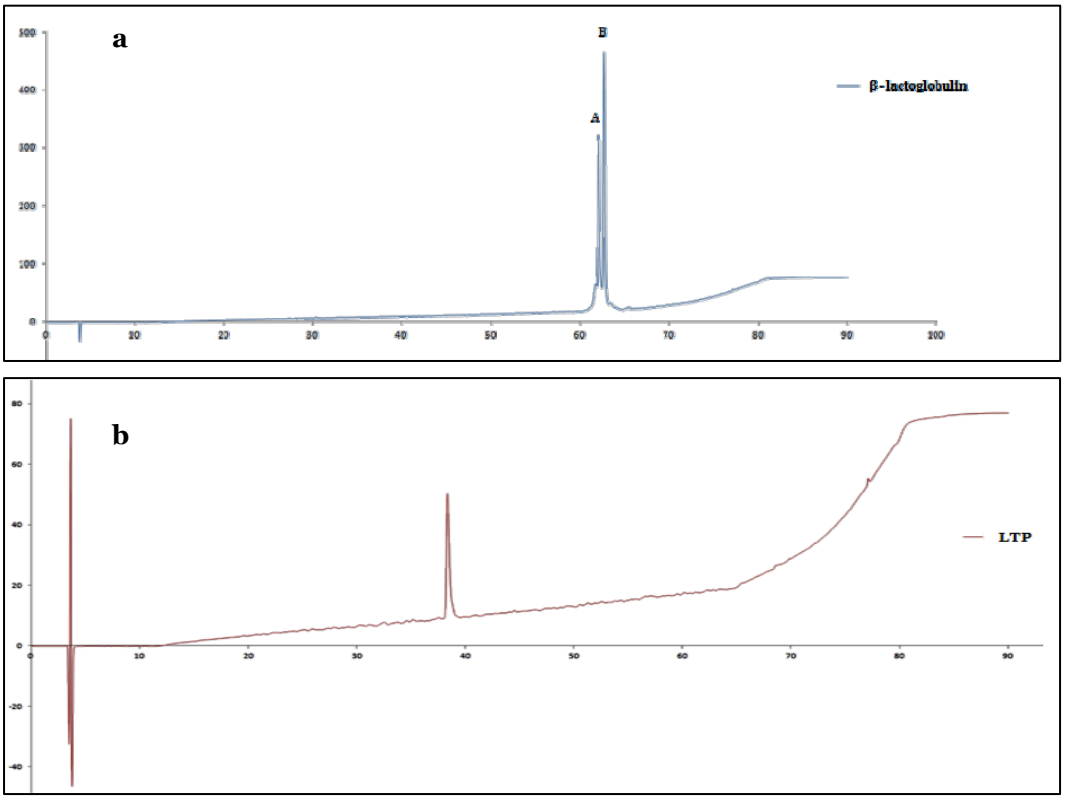


Fig 8. chromatograms of **a,** β-lactoglobulin A and B variants and of **b,** purified protein separated by reversed phase chromatography.

5.2 Tomato peel LTP characterization

After the MES buffer was replaced with Milli Q H_2O and the concentrating procedure was exploited, the purified protein was freezer dried. The net weight of the lyophilized protein was 5.2 mg.

5.3 Identification of a novel LTP from tomato peel by mass spectrometry analyses

In order to definitively identify the purified protein and to clarify its possible nature as tomato LTP, a MALDI-TOF analysis was performed.

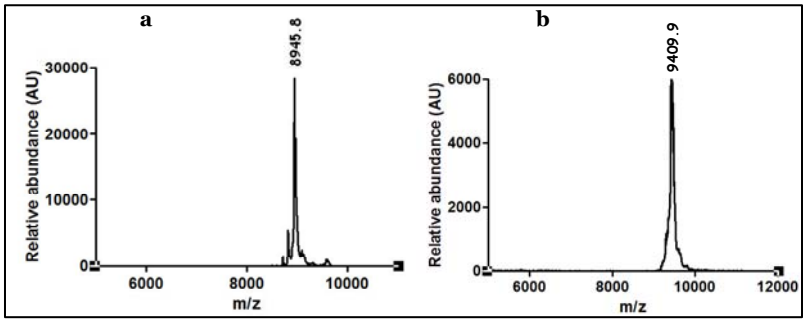


Fig 9. MALDI-TOF spectra of the purified protein; **a**, MALDI-TOF spectrum of the native protein; **b,** MALDI-TOF spectrum of the reduced and alkylated protein.

The spectrum related to the native protein revealed the presence of a main peak with m/z value of 8945.8 (Fig 9a). The same analysis, carried out with the reduced and alkylated purified protein, showed a peak whose m/z value was incremented of 464 Da (Fig 9b), which is the shift in the molecular mass expected for the typical structure of LTPs containing 4 disulphide bridges whose 8 cysteine are linked by a carboxamidomethyl group each to produce carboxymethyl cysteine (58 average mass change).

A search of the UniProt database for this molecular weight didn't find any already characterized LTPs, neither in tomato or in any other plant species.

So, in order to obtain some sequence information, a bottom-up approach of the purified protein was carried out. Spots relating to the purified protein (Fig 5) were in-gel digested by trypsin and the peptides were analyzed by LTQ-Orbitrap, as previously described.

Table 1. List of tryptic peptides derived from in gel-digestion of the spot relating to the purified protein and analyzed by LTQ-Orbitrap analysis.

Sequence	Expected molecular mass (Da)	Observed molecular mass (Da)
GPLGGC*C*R	875,33	875,37
GLLGAAK	628.39	628.39
GLLGAAKTPADR	1168.66	1168.66
TPADRK	686.37	686.37
SAANAIKGLNLGK	1255,73	1255,72
GLNLGK	600,36	600,36
AAGIPSAC*GVSIPYK	1489,76	1489,76
ISPFTDC*SK	1053.48	1053.48

* indicates a carboxamidomethyl-cysteine

Mascot software identified a tomato nonspecific lipid transfer protein (ExPASy E2GLP6) whose intact molecular weight (9381 Da) was not consistent with the one obtained by MALDI-TOF analysis. Besides, the peptide GLLGAAK couldn't belong to that protein, as shown in Fig 10b. Alternatively, the presence of the glutamic acid residue replacing the lysine in position 38 wouldn't allow the release of this peptide.
The same tryptic peptides (listed in Table 1) were used to perform a search in 'Tomato-ESTs' database (from John Innes Center facilities). In this case, the software identified a tomato EST, gi|150438794, containing the sequences of the peptides from the digestion, with a MOWSE (MOlecular Weight SEarch) score of 317 (Protein MASCOT scores greater than 65 are significant).
The loss of amino acids residues observed in the MALDI-TOF spectrum of the protein helped in fixing the C-terminus of the new LTP sequence, as reported in Fig 10a, while the N- terminus was inferred by subtracting the N-terminal amino acids until the expected weight was reached and it was confirmed by the alignment with the other tomato LTPs, already annoted in databases (Fig 10b). The putative N-terminus of this sequence was consistent with the signal peptide cleavage site predicted by SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/).

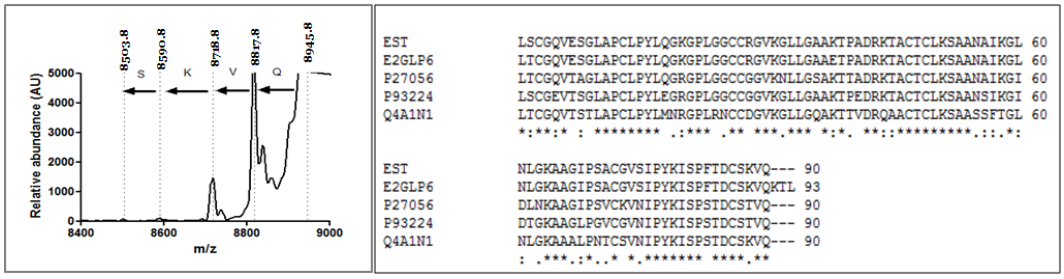


Fig 10. a, particular of MALDI-TOF spectrum of the purified protein showing the loss of amino acid residues at the C-terminus; **b,** alignment of the detected tomato EST with all the already known tomato LTPs.

The new identified LTP from tomato peel was proved to weigh 8944 (±1) Da and to be constituted by 90 residues (Fig 11).

LSCGQVESGLAPCLPYLQGKPLGGCCRGVKGLLGAAKTPADRKTACTCLKSAANAIKGLNLGKAAGIPSACGVSIPYKISPFTDCSKVQ

Fig 11. Sequence of the novel LTP purified from tomato peel.

Finally, in-gel digestion analysis was also performed with the other band of the gel in Fig 5 which eluted with other proteins after both size exclusion and ion exchange chromatography, in order to clarify its nature. LC/MS performed by Orbitrap identified peptides corresponding to another tomato LTP isoform, already present in database (ExPASy P27056) (data not shown). The differences in the primary structure of the two detected sequences (see Fig 10b for the alignment) might be responsible of their diverse interaction with the stationary phase in the ion exchange chromatography, resulting in a different repartition of the two isoforms during the elution step. Being already known in literature and not completely purified after a multidimensional chromatography approach, we decided to focus our efforts only on the characterization of the new tomato LTP isoform.

5.4 Spectroscopy analysis of secondary structure

An assessment was made of the folded state of the purified LTP using CD spectroscopy (Fig 12). The CD spectrum of the purified protein was entirely consistent with the estabilished α-helical structure of LTP, with a positive maximum at around 192 nm and an intense negative double minimum at 210 and 224 nm, already observed for LTPs from apple,[50] peach,[51] pear,[29] and hazelnut.[34]

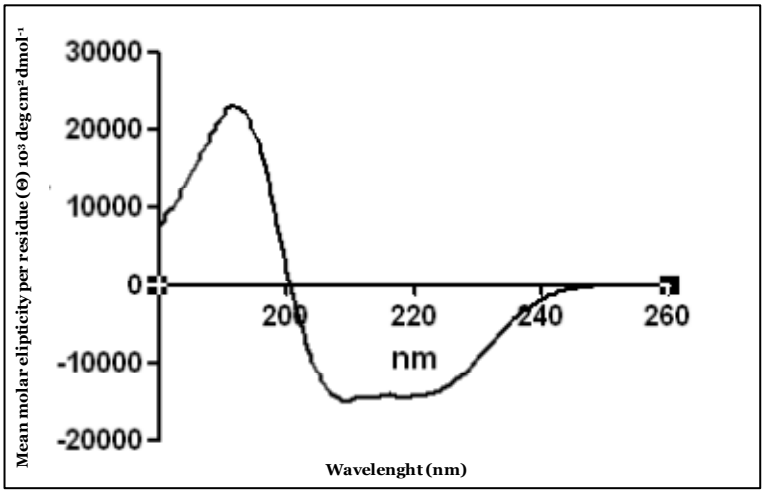


Fig 12. Far-UV circular dichroism (CD) spectroscopy of purified LTP from tomato peel.

All together the high isoeletric point, the molecular weight, the high sequence homology showed with other tomato LTPs, the presence of 4 disulphide bridges and the α-helical structure confirmed the nature of the protein isolated from tomato peel as a new isoform of lipid transfer protein.

5.5 Simulated in vitro gastrointestinal digestion of tomato peel LTP

As already described, one of the feature that makes the LTP a 'true' allergen is the high stability of its tertiary structure to proteolysis, which allows the allergen to maintain its immunogenic and allergenic motifs and thus to interact with the immune system associated with the gastrointestinal epithelia, thereby inducing both sensitization and systemic symptoms after ingestion.

In order to value the stability of the new tomato LTP isoform to proteolysis, the effect of the gastro-intestinal digestion on the purified protein in a model system were studied by means of mass spectrometry.

Firstly, 2 mg of the purified LTP underwent to a gastric phase digestion, in which the results of the activity of the pepsin on the protein were investigated after 1 hour of digestion at 37 °C. As shown by MALDI-TOF analysis in Fig 13, protein was completely resistant to gastric pepsinolysis at pH 2.5, since no differences were noticed before and after the pepsin addition, confirming the marked resistance of the LTP to pepsin digestion, as already reported for the peach.[52,53]

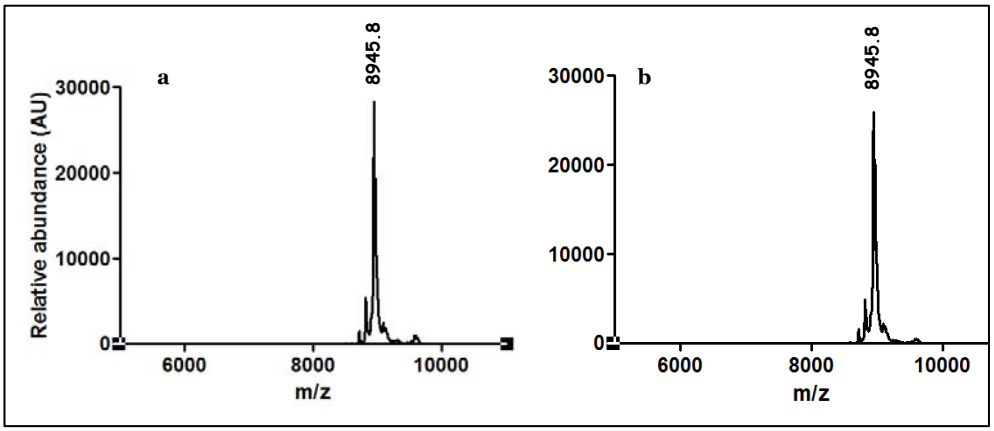


Fig 13. MALDI-TOF spectra of the **(a)** native pure protein before the addition of the pepsin and **(b)** after 60' of digestion at 37 °C.

After gastric pepsinolysis, the digestion of tomato LTP by trypsin and chymotrypsin under duodenal conditions was followed by MALDI-MS and SDS-PAGE, with the masses determined for the proteolysis products being compared with those of peptides predicted by in silico digestion with chymotrypsin and trypsin using mMass.[54]

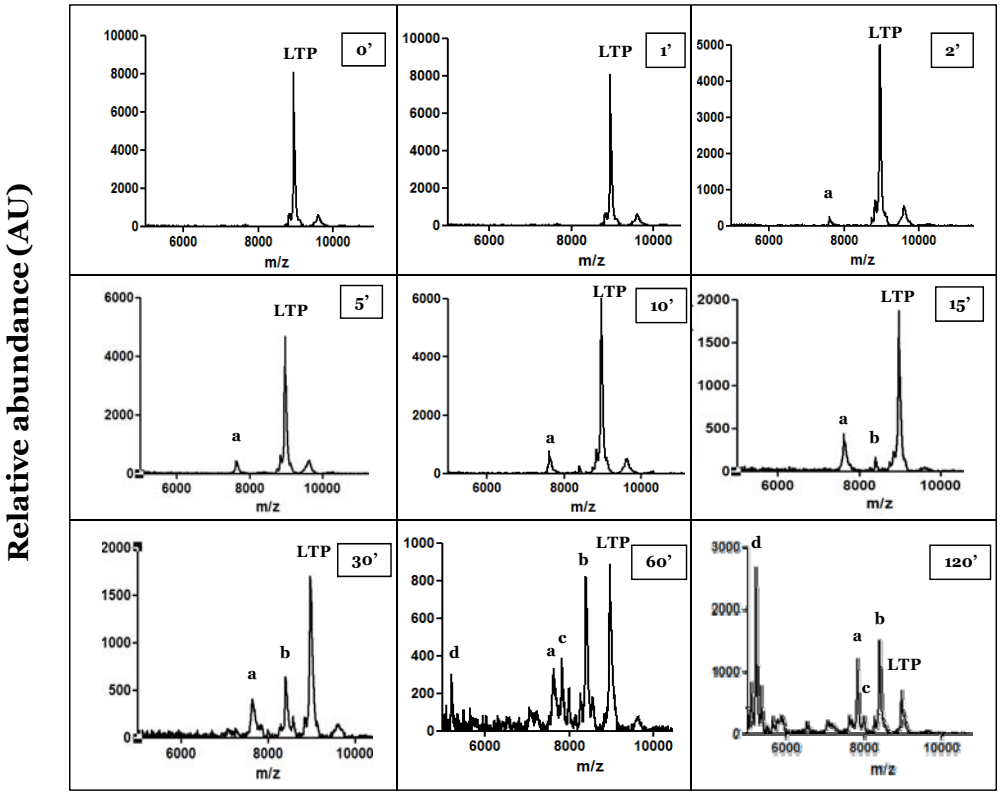


Fig 14. Analysis of tomato LTP duodenal digestion by MALDI-TOF mass spectrometry, under native conditions. The sampling points, expressed in minutes, are the numbers in the boxes; a, b, c, d indicate peptides released through the duodenal digestion (see text).

As reported in Fig 14, after 2' of duodenal digestion the peptide referred to as 'a' was released from LTP structure. This peptide (m/z 7615) is consistent with the sequence 1-78 of the tomato LTP with still 4 cysteine residues bound by 2 disulphide bridges.

Tyr_{78} represents a cleavage site for chymotrypsin; the analogous residue in peach LTP (Tyr_{79}) has been showed to be highly mobile.[36] For this reason it was the most accessible to the proteolytic enzymes and thus the first to be cleaved among all the potential ones. Other peptides, referred to as 'b', 'c' and 'd' were released through the duodenal phase but, after 2 hours of digestion, the intact LTP was still present in all digestion samples, corresponding to a mass of 8945.8 (±1) Da, indicating the high resistance of its tridimensional structure to proteolysis, even under extreme conditions.

The same digested samples also underwent to reduction and alkylation steps, before MALDI-TOF analysis. Also in this case the gastric pepsinolysis did not affect the structure of tomato LTP at all (Fig 15), remarking the already known resistance to pepsin digestion.[53]

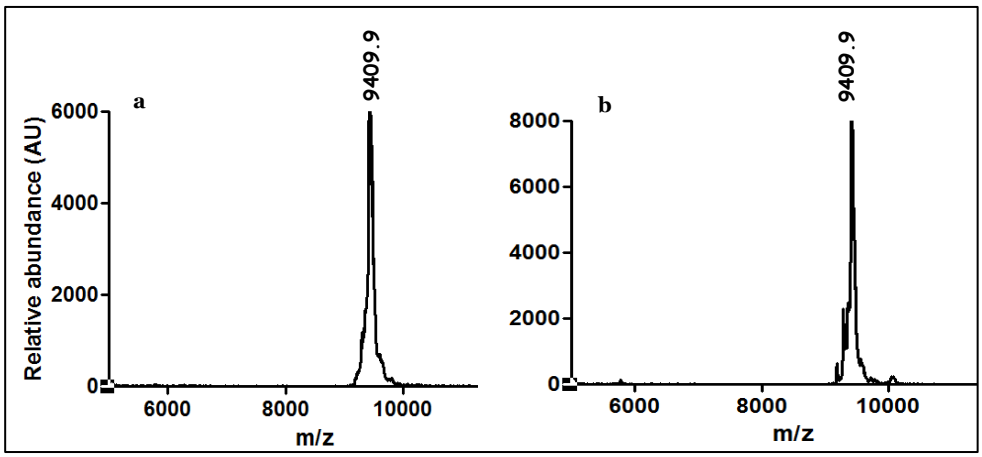


Fig 15. MALDI-TOF spectra of the **(a)** reduced and alkylated pure protein before the addition of the pepsin and **(b)** after 60' of digestion at 37 °C.

On the contrary, an intense proteolysis was observed through the duodenal phase (Fig 16). Immediately after the addition of the enzymes (time point 0), the peptide referred to as 'a*' was released from tomato LTP. This peptide, which is represented in the spectra with a m/z of 8018.3, corresponded to the peptide 1-78. Again the first observed cleavage concerned the Tyr_{78}, which was confirmed to be the cleavage site less tolerating the proteolysis. The relative abundance of this fragment also increased up to 30' and persisted as the predominant one till the end of digestion, while the abundance of the intact protein decreased. After 30' of duodenal digestion other two peptides were revealed: peptide 'b*', corresponding to the sequence 1-60 with a m/z of 6263.1 (expected 6263.1), was generated by the cleavage on the Leu_{60} by chymotrypsin; peptide 'c*', corresponding to the sequence 1-58 with a m/z of 6093 (expected 6092.97), was generated by the cleavage on Lys_{58} by both trypsin or chymotrypsin.

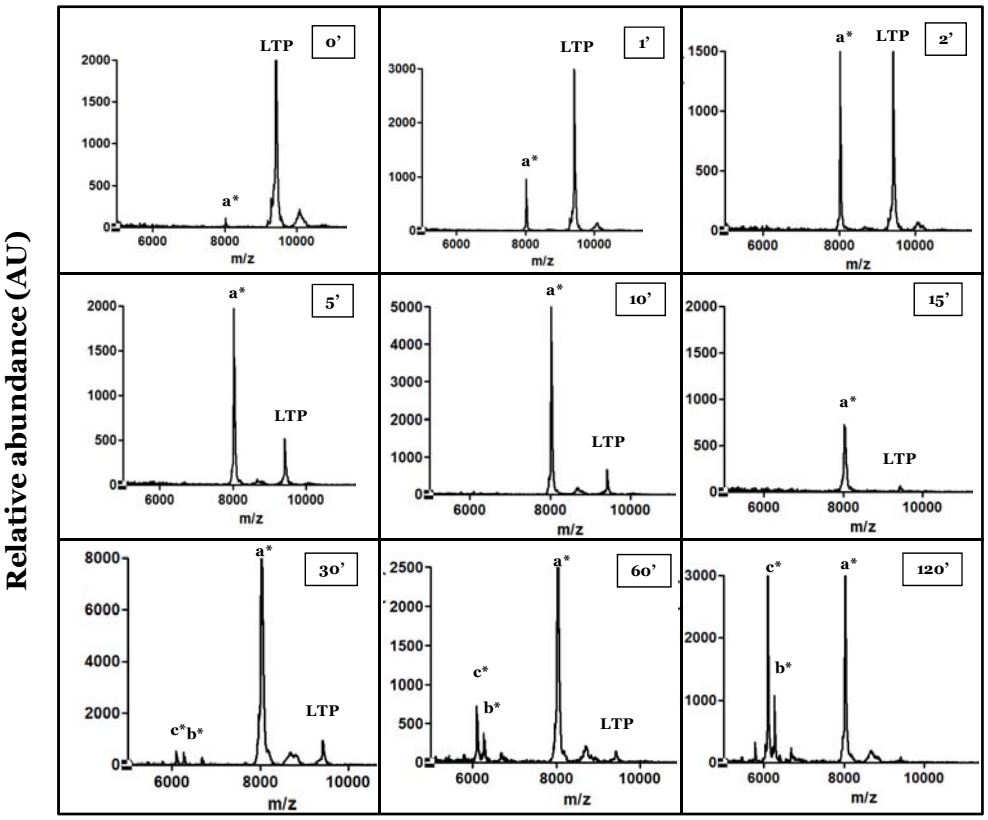


Fig 16. Analysis of tomato LTP duodenal digestion by MALDI-TOF mass spectrometry, under reducing conditions. The sampling points, expressed in minutes, are the numbers in the boxes; a*, b*, c*, indicate peptides released through the duodenal digestion (see text).

For duodenal digestion under reducing conditions, results were consistent between experimental methods, with quantitative analysis by SDS-PAGE (Fig 17, a and b) showing that tomato LTP is rapidly digested after the addition of the intestinal endoprotease and concentration of peptide 1-78 rises over 30' where it drops, as expected due to its further digestion into 1-60 and 1-58 fragments, even if fragments 1-60 and 1-58 are not discriminated by SDS-PAGE, possibly because the low resolution of the gel electrophoresis.

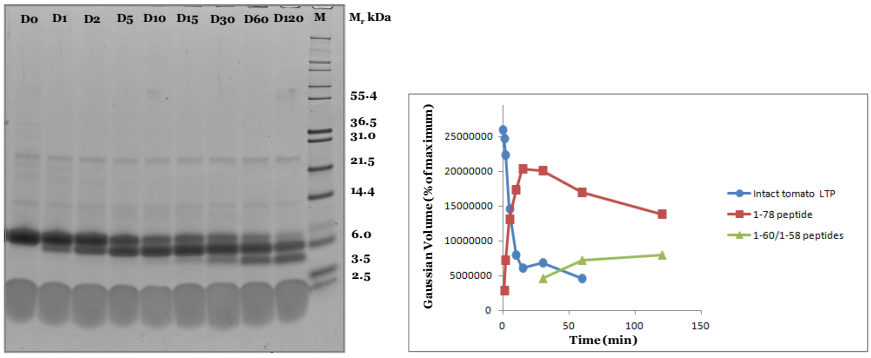


Fig 17. a, SDS PAGE performed on the digested and reduced samples at different times ('D' stands for 'duodenal' and the following numbers indicate the sampling times). The lane marked as 'M' represents the molecular weight standards and their relative molecular masses are reported on the right side. The larger bands on the bottom of the gel are the PMSF, while the ones higher than 14.4 kDa represent the proteolytic enzymes; **b,** relative abundance of tomato LTP and digestion products over time as estimated from densitometry of SDS-PAGE gels. Band intensities were calculated as a percentage of maximum Gaussian volume by scanning of gel images and subsequent analysis using the TotalLab package.

6. *Conclusions - Part I*

The method described here allowed to purify and identify a novel LTP isoform from tomato peel. The same method was also tested on tomato pulp but the gel filtration analysis gave a very low signal for the peak containing the expected protein when compared to the same one obtained from tomato peel. For this reason and according to the literature, in which it is reported that the localization of LTPs is concentrated in the peel of many plant species, we focused our efforts in purifying the protein from tomato peel.

The resolving power of the multidimensional chromatography step, associated with the high resolution of mass spectrometry techniques, allowed to identify a new tomato LTP isoform, also assigning a function to the tomato EST clone present in database.

This novel tomato LTP was the only discernable isoform in our purified protein fractions, and is therefore likely to be a major form *in planta*.

The simulated in vitro gastrointestinal digestion showed that this LTP isolated from tomato peel was still resistant to gastric pepsinolysis, both in the native and in reduced conformations. Concerning the duodenal phase, tomato LTP showed to be still intact at the end of the digestion, even after the release of some peptides, suggesting that most of the fragments remain associated by intramolecular disulfide bridges. However, the presence of observable 1-78 peptide in the non-reduced samples indicates either that some fragments are unassociated after duodenal digestion or that the ionization used in the MALDI experiments partially fragments associated peptides. On the contrary, the duodenal digestion of the reduced samples resulted in an almost complete conversion of the tomato LTP in three types of peptides 1-78, 1-58, 1-60, suggesting that the stability of this protein toward simulated duodenal proteolysis at pH 6.5 is related to the fully folded conformation. Finally, the retention of the overall three-dimensional structure of this protein following simulated gastrointestinal digestion explains the fact that proteolysis does not affect the ability of allergenic LTPs, such as grape,[55] and peach,[52] to bind IgE and

elicit histamine release. The effect of the immunological properties of this new tomato LTP isoform will be investigated further.

7. Results and Discussion – Part II

7.1 LTPs isolation from tomato peel, pulp and seeds

As in the previous part of the work, where common grapevine tomatoes were analyzed, in this second part tomatoes of Piccadilly variety, which is commonly spread in Italian markets, were used. In this way the detected LTP, eventually, could be representative of the main allergen LTP isoforms which are commonly present and recognized by IgE of tomato allergic patients. The analyses were carried out on the three main fractions of tomato fruit (peel, pulp and seeds), in order to know in which of them the LTP is more concentrated.

Since in this part of the work the aim was not isolating a pure LTP, but rather setting up a method to detect them in various vegetal tissues, the procedure for sample preparation was highly simplified, as compared to the first part.

Protein extraction was performed by means of precipitation with cold acetone and pellet extraction in a phosphate-buffered saline solution, as described elsewhere in this manuscript, and slightly modified for the seeds, which were previously reduced to a fine powder by milling in liquid nitrogen.

The quantification of protein extracts was performed by a bench fluorometer and the values obtained for each tissue were related to the amount of the dry material, as reported in Table 2.

Table 2. Quantification of protein extract from tomato peel, pulp and seed performed by Qubit® fluorometer.

Tissue	A_s (g)	A_{dm} (g)	P (mg)	Prot %
Pulp	50	0,9	29,48	3,27%
Peel	10	2,04	78,82	3,86%
Seed	6,37	3,07	94,67	3,08%

A_s= Amount of sample; A_{dm}= Amount of dry sample; P= Amount of extracted proteins, obtained by Qubit® fluorometer; Prot %= Percentage of proteins related to A_{dm}.

As expected, the percentage of extracted proteins was higher for the seeds, if comparing the amount of the starting material among all the other fractions. This result was in line with the function fulfilled by seeds, that is storage organs.

Being known that LTPs are proteins with molecular weights ranging between 7-11 kDa, a simple and fast analytical method was developed here, in order to narrow down the field of analysis to those proteins with appropriate masses, including LTPs. For this reason, a double ultracentrifugation step was carried out with each protein extract: the first one was performed by using centrifugal filter devices containing membranes with a nominal molecular weight limit of 30 kDa; the filtrate solutions, containing all the molecules with molecular weights lower than 30 kDa, were then applied to centrifugal filter devices containing membranes with a nominal molecular weight limit of 3 kDa. In this case the retentates (i.e. the upper part of devices) were recovered, as they contained proteins with molecular weight higher than 3 kDa. In

this way, theoretically, all the molecules with masses including the range 30-3 kDa were isolated from the total protein extract of each tomato tissue. Although the membranes were characterized by a nominal molecular weight limit (NMWL), since these limits were very far from the supposed molecular mass of LTP proteins (7-11 kDa), the possibilities that LTPs could be lost were very low.

7.2 Identification and characterization of tomato LTPs by UPLC/MS analysis

The Ultra Performance Liquid Chromatography (UPLC) takes advantage of technological strides made in particle chemistry performance, system optimization, detector design, and data processing and control. Using sub-2 μm particles and mobile phases at high linear velocities, and instrumentation that operates at higher pressures than those used in HPLC, dramatic increases in resolution, sensitivity, and speed of analysis can be obtained. This new category of analytical separation science retains the practicality and principles of HPLC while creating a marked improvement in chromatographic performance.

As for all the other chromatography system, UPLC allows the separation of molecules in a complex mixture, by partitioning between a stationary phase and a mobile phase on the basis of their relative affinity. Usually the most suitable stationary phases are made up of alkyl chains bonded covalently to the support surface, which are able to interact with apolar proteins. The elution is performed increasing organic solvent concentration of the mobile phase. The detection of proteins is achieved by a mass spectrometer: in this case, the separated molecules are conveyed to an ESI source which provides for their ionization. The ion species are then separated by a quadrupole analyzer, according to their m/z value, and are then revelead by a detector. Currents generated by each ions are profiled in the Total Ion Current (TIC) where each peak represents one or more eluted molecules which can be identified by the interpretation of the relative mass spectrum.

The separation of extracted proteins was performed by using two different chromatographic columns, C8 and C18, which differ in the number of carbons of the alkyl chains bound to the support. Considering the small size of the alkyl chain, in the first column dispersive interactions with the stationary phase are reduced than in the second column. Generally, for proteins with a certain size, this leads to improved chromatographic peak shapes. Being both used for the separation of small-medium proteins, a parallel study was carried out, in order to find out which of them was more appropriate in the separation of tomato LTP.

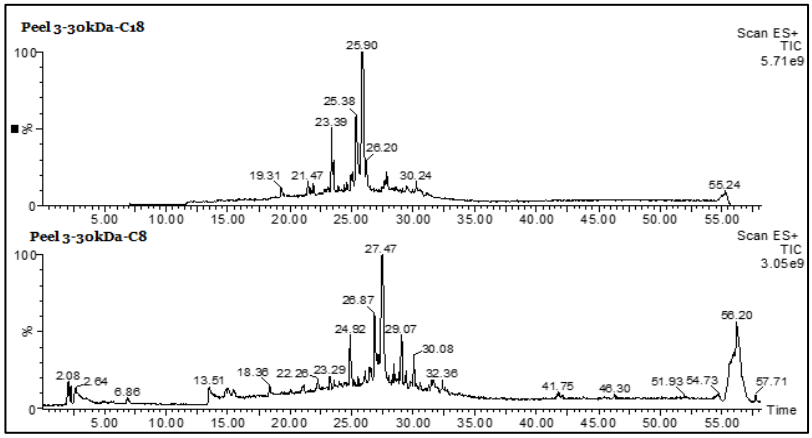


Fig 18. UPLC chromatograms related to peel fraction proteins obtained with a C18 (panel above) and a C8 (panel below) column.

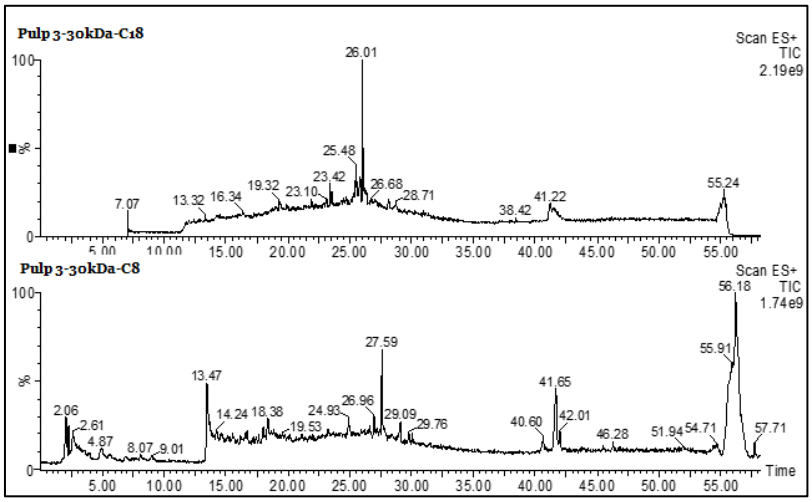


Fig 19. UPLC chromatograms related to pulp fraction proteins obtained with a C18 (panel above) and a C8 (panel below) column.

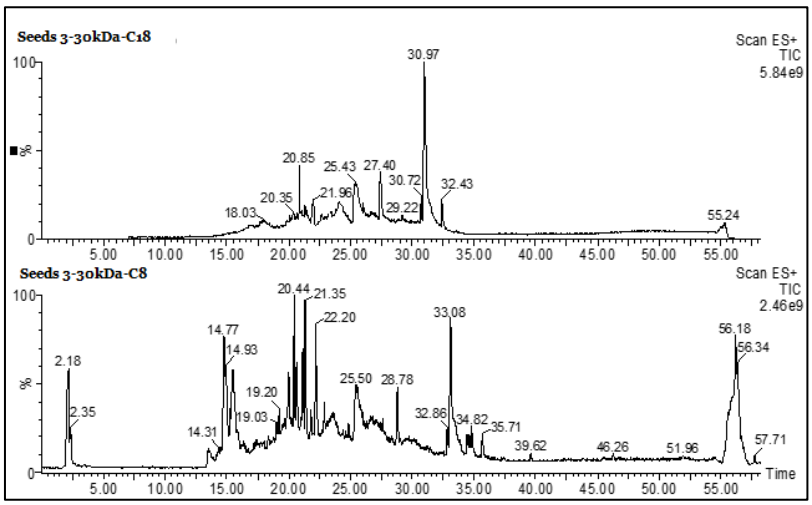


Fig 20. UPLC chromatograms related to seed fraction obtained with a C18 (panel above) and a C8 (panel below) column.

As shown in chromatograms reported in Figg 18, 19 and 20, proteins seemed to be more abundant in the fraction seeds. Besides, it is pretty clear that C8 column showed higher and better resolved peaks. This could be due to the fact that many proteins have hydrophobic residues that interact with the column. Consequently, shorter alkyl chains are generally more appropriate for the separation of these proteins, in order to avoid a prolonged retention leading to a broadening of the chromatographic peak. In contrast, peptides are smaller and require longer hydrophobic chains to be efficiently separated in reversed phase; in this case a C18 column would be ideal.
The search of proteins of interest was carried out for every tomato protein fraction by investigating mass spectra relating to each chromatographic peak and identifying the molecular mass through the multicharged clusters of ions.
In this way, four proteins with molecular weights similar to the ones of plant LTPs, that is 7039 (±1) Da, 9354 (±1) Da, 9441 (±1) Da and 9724 (±1) Da, were detected only in the seed fraction.

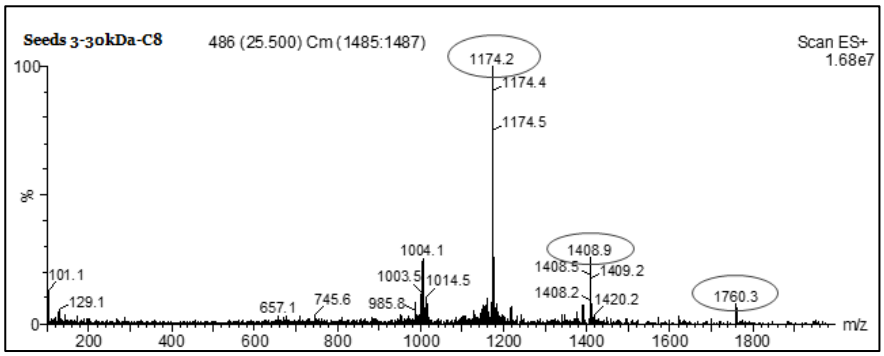


Fig 21. ESI-Q mass spectrum of 7039 (±1) Da protein. Circles indicate multicharged ions.

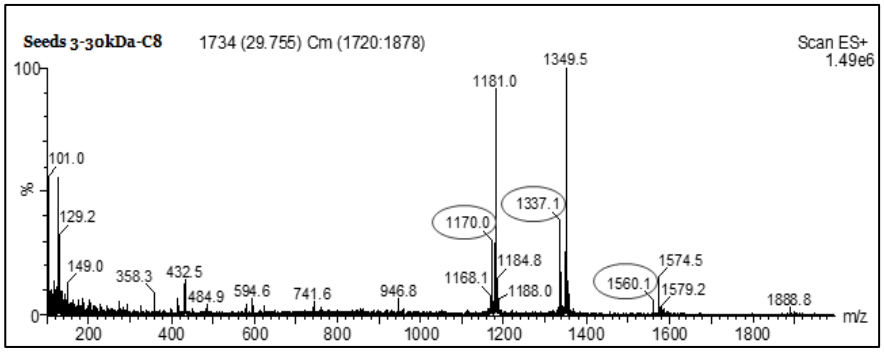


Fig 22. ESI-Q mass spectrum of 9354 (±1) Da protein. Circles indicate multicharged ions.

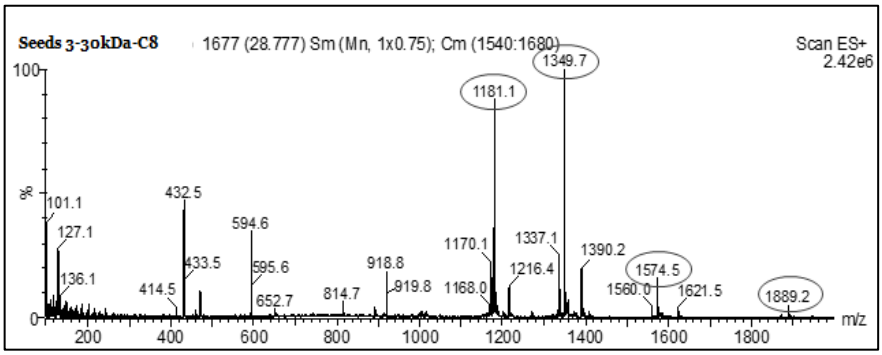


Fig 23. ESI-Q mass spectrum of 9441 (±1) Da protein. Circles indicate multicharged ions.

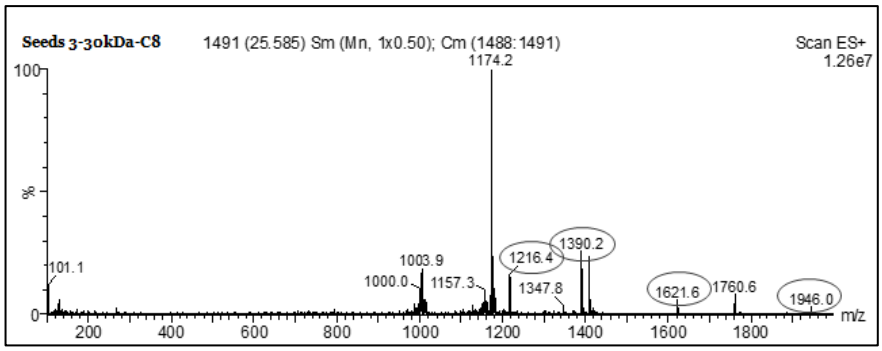


Fig 24. ESI-Q mass spectrum of 9724 (±1) Da protein. Circles indicate multicharged ions.

The retention times and the characteristic ions were summarized in Table 3.
As shown, all the four proteins had a molecular weight close to the one already found for other plant LTPs, such as Pru P 3 from peach (9135 Da),[56] LTPa and LTPb from pear (9252 and 9250 Da),[29] LTPa and LTPb from pomegranate (9342 and 9467),[33] LTP from maize (9046 Da),[57] and Mal d 3 from apple (9076 Da).[50]

Table 3. Retention times and characteristic ions of potential identified LTP

Retention Time (min)	Average Mass (±1 Da)	MS Spectra characteristic ions (m/z)
25.5 26.4	7039	1760.5 (MH_4^{4+}), 1408,9 (MH_5^{5+}) 1174.1 (MH_6^{6+})
26.8 29.7	9354	1560.1 (MH_6^{6+}) 1337.1 (MH_7^{7+}) 1170.2 (MH_8^{8+})
26.6 29.9	9441	1889,2 (MH_5^{5+}) 1574,4 (MH_6^{6+}) 1349.6 (MH_7^{7+}) 1181.1 (MH_8^{8+})
25.6 27.7	9724	1945,9 (MH_5^{5+}) 1621,5 (MH_6^{6+}) 1390.1 (MH_7^{7+}) 1216.5 (MH_8^{8+})

Extraction procedure of characteristic ions allowed to obtain the eXtract Ion Chromatogram (XIC), giving a better understanding of the chromatographic dynamics.

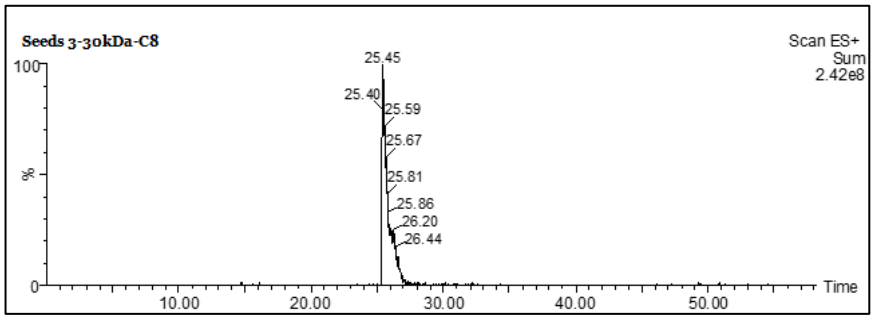


Fig 25. Extracted ion chromatogram of 7039 (±1) Da protein.

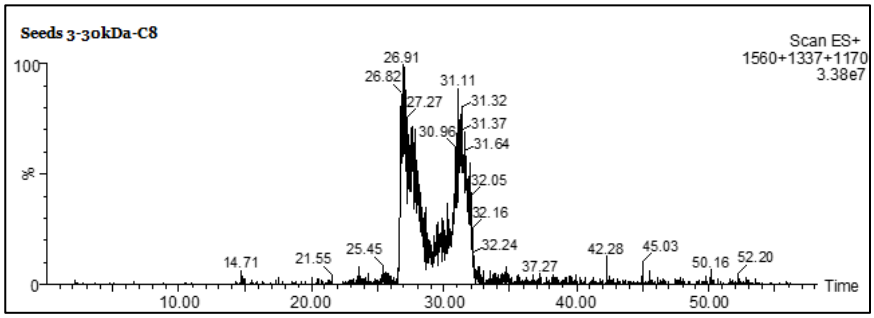


Fig 26. Extracted ion chromatogram of 9354 (±1) Da protein.

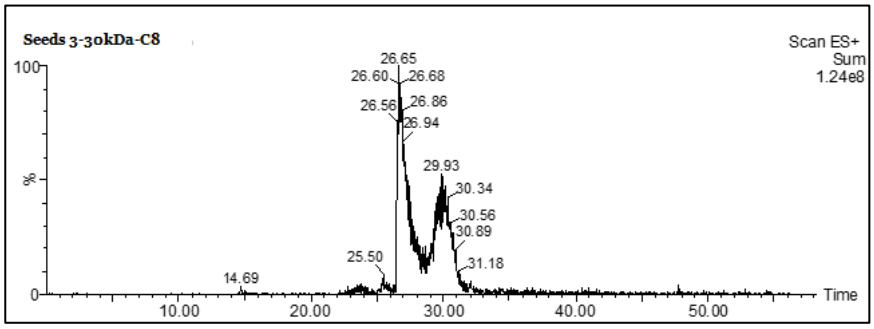


Fig 27. Extracted ion chromatogram of 9441 (±1) Da protein.

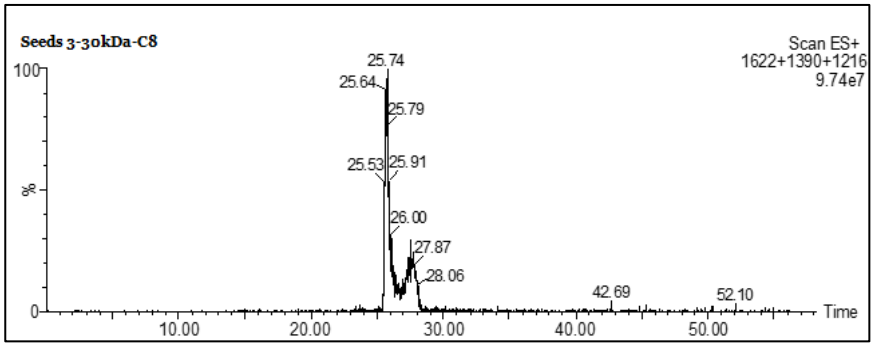


Fig 28. Extracted ion chromatogram of 9724 (±1) Da protein.

As we can infer from the extracted ion chromatograms, all the four potential tomato LTP isoforms co-eluted in two chromatographic peaks, which showed the same mass spectrum. This suggested that each isoform could undergo to a partition between a folded conformation (or native) which is the first to elute, and a partially unfolded conformation (partially denatured), which interacts more with the stationary phase. Beside this feature, the retention time of the extracted peaks was pretty similar to the one obtained for the pure peach LTP, suggesting a similar protein (Fig 29).

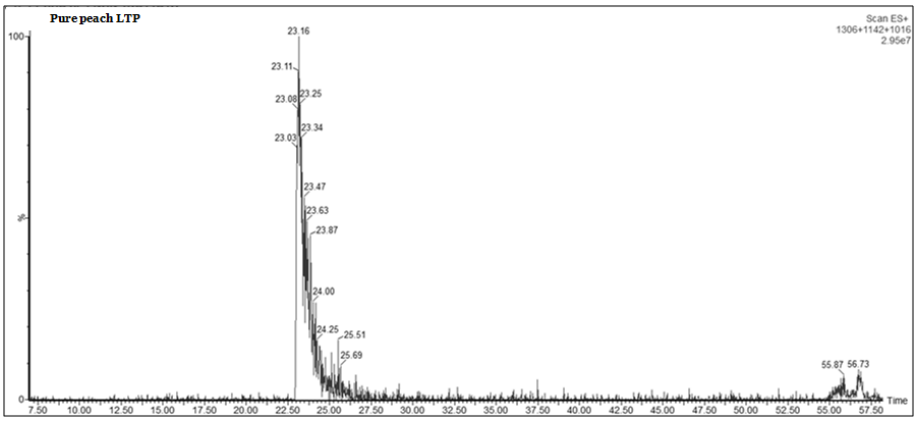


Fig 29. Extracted ion chromatogram of pure Pru p 3 (LTP) from peach (9135 ±1 Da).

As already underlined, these potential tomato LTPs were detected only in seed fraction sharply, while some weak signals, and not for all four, were observed in pulp fraction. By the way, none of the molecular weight of these proteins fit with the masses of the tomato LTPs already known in literature (Table 4).

Table 4. List of all the tomato LTP isoforms annoted in Uniprot database.

Accession number	Organism	Average Mass (Da)	Mono isotopic Mass (Da)
Q3YMR2 - Non-specific lipid-transfer protein 2	Solanum chilense (Lycopersicon chilense)	9139.57	9133.47
O24037 - Non-specific lipid-transfer protein 1	Solanum pennellii (Lycopersicon pennelli)	8968.5	8962.59
O24038 - Non-specific lipid-transfer protein 2	Solanum pennellii (Lycopersicon pennelli)	9114.56	9108.48
E2GLP6 - Non-specific lipid-transfer protein	Solanum lycopersicum var. cerasiformae	9309.98	9303.81
P27056 - Non-specific lipid-transfer protein 1	Solanum lycopersicum (Lycopersicum esculentum)	9025.58	9019.62
P93224 - Non-specific lipid-transfer protein 2	Solanum lycopersicum (Lycopersicum esculentum)	8856.27	8850.41
Q4A1N0 - Non-specific lipid-transfer protein	Solanum lycopersicum (Lycopersicum esculentum)	8856.27	8850.41
Q4A1N1 - Non-specific lipid-transfer protein	Solanum lycopersicum (Lycopersicum esculentum)	9308.84	9302.64

7.3 Structural characterization of identified proteins

As no weight match was observed in the databases for the detected potential tomato LTP isoforms, in order to obtain more information about their nature another investigation was carried out by exploiting the tertiary structure feature which is preserved among all the plant LTPs, that is 4 disulphide bridges holding together 8 cysteine residues.

The reduction and alkylation of these proteins with dithiothreitol and 2-iodoacetamide, in fact, would imply a shift in their molecular weight of 464 (±1) Da, due to the link of a 58 Da carboxymethyl group to each cysteine residues. An UPLC/MS analysis would detect these differences.

Fractionated protein extracts from tomato peel, pulp and seeds, prepared as described in paragraph 4.2, were denatured by urea and then reduced by dithiothreitol, to break the disulphide bridges, and alkylated by 2-iodoacetamide, to avoid a random refold. Also in this case, the performance of both C18 and C8 column in separating whole proteins was tested.

As already observed before, the seed fraction seemed to be the most abundant in proteins, comparing chromatograms from peel and pulp fractions. Also in this case, C8 column was the most suitable in separating proteins.
For each protein of interest, characteristic ions were inferred, considering the shift of mass due to reduction and alkylation; they were extracted from TICs and the multi-charged pattern was investigated for each XIC.
In this way, and once again only in seed fraction, four proteins with molecular weights corresponding to the previous ones with an increase of 464 Da (±1), were found. Mass spectra with the characteristic ions are reported below.

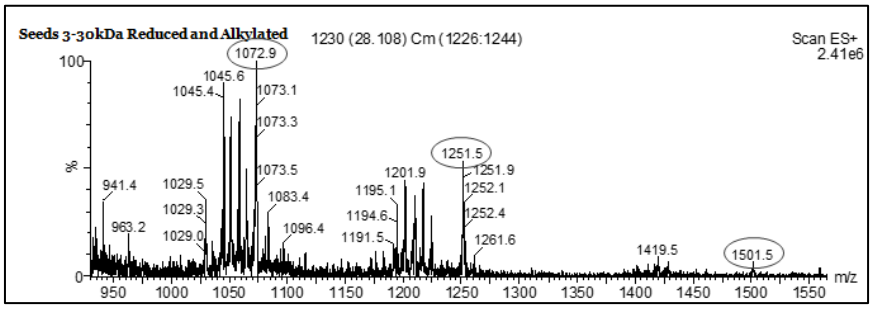


Fig 30. ESI-Q mass spectrum of 7052 (±1) Da protein (from 7039+464). Circles indicate multicharged ions.

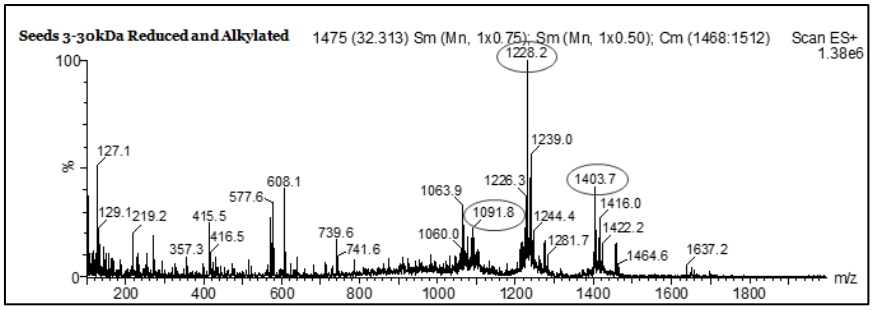


Fig 31. ESI-Q mass spectrum of 9818 (±1) Da protein (from 9354+464). Circles indicate multicharged ions.

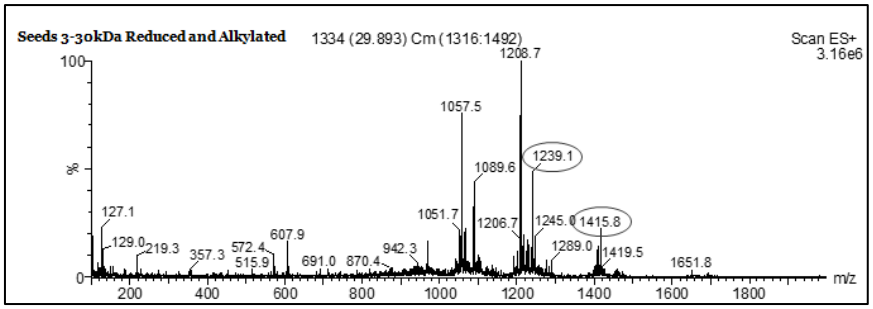


Fig 32. ESI-Q mass spectrum of 9904 (±1) Da protein (from 9441+464). Circles indicate multicharged ions.

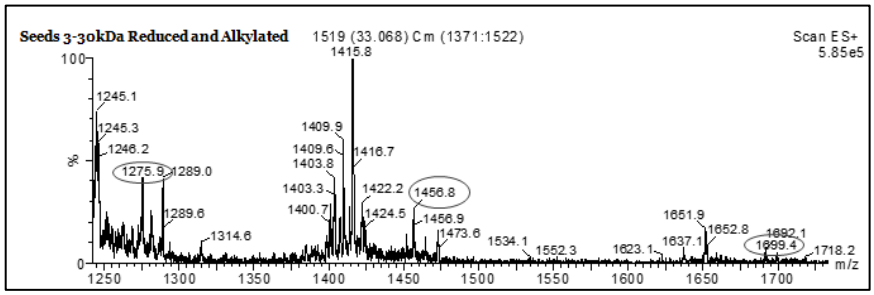


Fig 33. ESI-Q mass spectrum of 10189 (±1) Da protein (from 9724+464). Circles indicate multicharged ions.

The retention times and the characteristic ions were summarized in Table 5.

Table 5. Retention times and characteristic ions of reduced and alkylated potential LTPs

Retention Time (min)	Average Mass (±1 Da)	MS Spectra characteristic ions (m/z)
28.1	7502	1501.1 (MH_5^{5+}), 1251.5 (MH_6^{6+}) 1072.9 (MH_7^{7+})
32.3	9818	1403.7 (MH_7^{7+}) 1228.2 (MH_8^{8+}) 1091.8 (MH_9^{9+})
29.9 32.3	9906	1415.8 (MH_7^{7+}) 1239.1 (MH_8^{8+})
32.5	10189	1699.4 (MH_6^{6+}) 1456.5 (MH_7^{7+}) 1274.7 (MH_8^{8+})

Observed retention times were delayed, if compared to the ones of native proteins. This could be ascribed to the longer interaction between denatured and alkylated proteins with the stationary phase of column.

Albeit not a certainty, the retention times and shape of the peaks similar to the one of peach *Pru p 3*, the molecular weights ranging between 7-10 kDa and the presence of 4 disulphide bridges represent meaningful evidences of these proteins as tomato LTP isoforms. Further investigations will be performed by MS^n to obtain sequence information and definitively assess their nature.

8. Conclusions – Part II

The method of extraction and fractionation of proteins from the three main tomato tissues (peel, pulp and seeds) and the UPLC/MS analysis, described in the second part of this chapter, allowed to identify four proteins in the seed fraction, which could be ascribed to tomato LTP isoforms. Despite the lack of information about the primary structure, the retention time, the molecular weights and the proof of 4 disulphide bridges in their tertiary structures are important clues which allowed us to assume their nature as LTPs. Nevertheless, none of the molecular weight of these potential LTPs were already present in the database or did fit with the one purified and

characterized from tomato peel, as described in the first part of this chapter. The presence of so many LTP isoforms could be due to the biological and physiological roles of LTPs in the cell system. Considering their lipid binding capacity and their antimicrobial activity, for example, the primary structure of LTPs may need some arrangements, set by the evolutionary pressure.
Besides, in peel and pulp fractions no potential tomato LTP isoforms was detected. This may sound in contrast with the most frequent localization of LTP in the peel of many plant species, especially in the *Rosaceae* family,[29,56] even if it has been found in the seeds of pepper (*Capsicum annum*),[58] and it is commonly concentrated in the seeds of hazelnut.[34] Considering that the localization of LTPs may be related to the belonging family, in this case the finding of potential tomato LTPs has to be ascribed to the method of protein extraction and analysis which could allow the detection of proteins only in the tissue where they are most concentrated.

9. *Conclusive remarks*

The data presented in this chapter describe a very complex scenario, as yet not fully understood, depicting the family of non-specific lipid transfer proteins (LTPs). As described before, different tomato LTP isoforms were detected, besides the ones already reported in the literature.[47,49] A new isoform was purified from the peel of a common grapevine variety, whilst four proteins, which shared several features with well-known plant LTPs, such as molecular weights, retention times and 8 cysteine residues bound together by 4 disulphide bridges, were revealed in the seeds of the Piccadilly variety.
First of all, it can be observed that the occurrence of so many tomato LTP isoforms is not unusual across the plant kingdom. It is known, in fact, that in the three species for which the full genome sequence is available (rice, Arabidopsis, and poplar) approximately a dozen of the family LTP1 genes are present in each genome.[15] The reason for such a plethora of LTP isoforms is still unclear. Considering the amount of functional properties which have been hypothesized for LTPs, by observing their behaviour in vitro, it may be speculated that one, or more, isoform(s) may be involved in the fulfilment of these functions. The ability to detect one isoform or another could be related, instead, to several other aspects, which affect proteomic analysis, such as the genetic characteristics and the phenological phase of the individual under examination, as well as the efficiency of the protein extraction method and the sensitivity of the analytical system employed. All these variables may allow the preferential detection of one isoform over the others, making it difficult to gain a whole and complete understanding of the expression pattern of such proteins.
Secondly, the LTP distribution across different tomato fruit tissues has been shown to be very variable. Unlike what has been previously observed by Pravettoni and co-workers,[47] in this study tomato LTP isoforms have been found in peel or seeds of different variety, and no LTP isoform was revealed in pulp tissue. Again, this variability in LTP distribution may be related to the genotype, to the stage of ripening of the fruit, as well as some technical issues, such as the experimental design, which could just give a ‘snapshot’ of the protein expression pattern at a certain moment. LTP spatial and temporal expression, in turn, may depend on the functional and biological roles fulfilled by these proteins, which could be different according to the tissue, the

developmental stages and the physiological conditions. Moreover, the fact that LTPs are concentrated in the peel of many plant species, especially in the *Rosaceae* family,[29,56,59] but they have been detected also in the seeds,[58,34] and in the pulp,[47] of other plants, suggests a possible correlation of their distribution with the genus and the species.

Thirdly, the structural stability of purified tomato LTP toward gastric pepsinolysis at pH 2.5 and simulated gastroduodenal proteolysis at pH 6.5 was investigated. Despite the presence in the protein of numerous predicted pepsin, trypsin, and chymotrypsin cleavage sites, tomato LTP was only partially digested, since an amount of the intact protein was still present at the end of the digestive process. The three-dimensional structure of LTP is responsible for its resistance to gastric pH and peptic digestion, making this protein an ideal food allergen, as confirmed by the digestion under reducing conditions, which led to the complete proteolysis of the protein.

Finally, the characterization of tomato LTPs still represents a difficult challenge, since these peptides comprise a multigenic family with different genes that are expressed in different tissues, in different developmental stages of plants and that also react differently to an array of stimuli. Studies investigating the presence of isoforms and their temporal and spatial distribution within the plant may produce particularly complicated results. The characterization and the purification of one isoform rather than another, in turn, can affect the efficacy of these proteins for the screening of LTP-allergic patients, whenever the tested protein is not the same one to which the patient is sensitized. This would require the need to screen for every different LTP isoform found, which could be time and cost ineffective.

10.References

[1] de Oliveira Carvalho A, Moirera Gomes V. Role of plant lipid transfer proteins in plant cell physiology – A concise review. *Peptides*, 2007, **28:**1144-1153.

[2] Kader JC. Lipid-transfer proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 1996, **47**:627–654.

[3] Castro MS, Gerhardt IR, Orru S, Pucci P, Bloch JC. Purification and characterization of a small (7.3 kDa) putative lipid transfer protein from maize seeds. *Journal of Chromatography B,* 2003, **794**:109–114.

[4] Douliez JP, Pato C, Rabesona H, Molle D, Marion D. Disulfide bond assignment, lipid transfer activity and secondary structure of a 7-kDa plant lipid transfer protein, LTP2. *European Journal of Biochemistry*, 2001, **268**:1400–1403.

[5] Liu YJ, Samuel D, Lin CH, Lyu PC. Purification and characterization of a novel 7-kDa non-specific lipid transfer protein-2 from rice (Oryza sativa). *Biochemical and Biophysical Research Communications,* 2002, **294**:535–540.

[6] Arondel V, Tchang F, Baillet B, Vignols F, Grellet F, Delseny M, Kader JC, Puigdomenech P. Multiple mRNA coding for phospholipid-transfer protein form Zea mays arise from alternative splicing. *Gene*, 1991, **99**:133–136.

[7] Suelves M, Puigdomènech P. Different lipid transfer protein mRNA accumulates in distinct parts of Prunus amygdalus flower. *Plant Science*, 1997, **129**:49–56.

[8] Vignols F, Lund G, Pammi S, Trèmousaygue D, Grellet F, Kader JC, Puigdomènechc P, Delseny M. Characterization of a rice gene coding for a lipid transfer protein. *Gene*, 1994, **142**:265–270.

[9] Garcìa-Garrido JM, Menossi M, Puigdimenèch P, Martìnez-Izquierdo JA, Delseny M. Characterization of a gene encoding an abscisic acid-inducible type-2 lipid transfer protein from rice. *FEBS Letters*, 1998, **428**:193–199.

[10] Kalla R, Shimamoto K, Potter R, Nielsen PS, Linnestad C, Olsen OA. The promoter of the barley aleurone-specific gene encoding a putative 7 kDa lipid transfer protein confers aleurone cell-specific expression in transgenic rice. *Plant Journal*, 1994, **6**:849–860.

[11] Lee JY, Min K, Cha H, Hwang DHSKY, Suh SW. Rice nonspecific lipid transfer protein: the 1.6Å crystal structure in the unliganded state reveals a small hydrophobic cavity. *Journal of Molecular Biology,* 1998, **276**:437–448.

[12] Shin DH, Lee JY, Hwang KY, Kim KK, Suh SW. High-resolution crystal structure of the non-specific lipid transfer protein from maize seedlings. *Structure*, 1995, **3**:189–199.

[13] Samuel D, Liu YJ, Cheng CS, Lyu PC. Solution structure of plant nonspecific lipid transfer protein-2 from rice (Oryza sativa). *Journal of Biological Chemistry*, 2002, **277**:35267–35273.

[14] Han GW, Lee JY, Song HK, Chang C, Min K, Moon J, Shin DH, Kopka ML, Sawaya MR, Yuan HS, Kim TD, Choe J, Lim D, Moon HJ, Suh SW. Structural basis of non-specific lipid binding in maize lipid transfer protein complexes revealed by high-resolution X-ray crystallography. *Journal of Molecular Biology,* 2001, **308**:263–278.

[15] Yeats TH, Rose JKC. The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). *Protein Science*, 2008, **17**:191–198.

[16] Pelèse-Siebenbourg F, Caelles C, Kader J-C, Delseny M, Puigdomènech P. A pair of genes coding for lipid-transfer proteins in Sorghum vulgare. *Gene*, 1994, **148**:305–308.

[17] Jung HW, Kim W, Hwang BK. Three pathogen-inducible genes encoding lipid transfer protein from pepper are differentially activated by pathogens, abiotic and environmental stresses. *Plant, Cell and Environment,* 2003, **26**:915–928.

[18] Arondel V, Vergnolle C, Cantre C, Kader JC. Lipid transfer proteins are encoded by a small multigene family in Arabidopsis thaliana. *Plant Science*, 2000, **157**:1–12.

[19] Vignols F, Wigger M, Garcìa-Garrido JM, Grellet F, Kader JC, Delseny M. Rice lipid transfer protein (LTP) genes belong to a complex multigene family and are differently regulated. *Gene*, 1997, **195**:177–186.

[20] Garcìa-Olmedo F, Molina A, Segura A, Moreno M. The defensive role of nonspecific lipid-transfer proteins in plants. *Trends in Microbiology,* 1995, **3**:72–74.

[21] Pyee J, Yu H, Kolattukudy PE. Identification of a lipid transfer protein as the major protein in the surface wax of broccoli (Brassica oleracea) leaves. *Archives of Biochemistry and Biophysics,* 1994, **311**:460–468.

[22] Thoma S, Hecht U, Kippers A, Botella J, De Vries S, Somerville C. Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from Arabidopsis. *Plant Physiology*, 1994, **105**:35–45.

[23] Jang CS, Lee HJ, Chang SJ, Seo YW. Expression and promoter analysis of the TaLTP1 gene induced by drought and salt stress in wheat (Triticum aestivum L). *Plant Science*, 2004, **167**:995–1001.

[24] Broekaert W F, Cammue BPA, De Bolle MFC, Thevissen K, De Samblanxa GW, Osbornb RW, Nielsonc K. Antimicrobial peptides from plants. *Critical Reviews in Plant Sciences*, 1997, **16**:297 –323.

[25] Hoffmann-Sommergruber K. Pathogenesis-related (PR)-proteins identified as allergens. *Biochemical Society Transactions,* 2002, **30**:930–935.

[26] Dìaz-Perales A, Garcia-Casado G, Sanchez-Monge R, Garcia-Salles FJ, Barber D, Salcedo G. cDNA cloning and heterologous expression of the major allergens from peach and apple belonging to the lipid-transfer protein family. *Clinical* & *Experimental Allergy,* 2002, **32**:87–92.
[27] Pastorello EA, Farioli L, Pravettoni V, Ortolani C, Ispano M, Monza M, Baroglio C, Scibola E, Ansaloni R, Incorvaia C, Conti A. The major allergen of peach (Prunus persica) is a lipid transfer protein. *Journal of Allergy and Clinical Immunology*, 1999, **103**:520–526.
[28] Pastorello EA, Farioli L, Pravettoni V, Giuffrida MG, Ortolani C, Fortunato D, Trambaioli C, Scibola E, Calamaria AM, Robino AM, Conti A. Characterization of the major allergen of plum as a lipid transfer protein. *Journal of Chromatography B*, 2001, **756**:95–103.
[29] Ramazzina I, Amato S, Passera E, Sforza S, Mistrello G, Berni R, Folli C. Isoform identification, recombinant production and characterization of the allergen lipid transfer protein 1 from pear (Pyr c 3). *Gene*, 2012, *491*: 173-181.
[30] Pastorello EA, Farioli L, Pravettoni V, Ortolani C, Fortunato D, Giuffrida MG, Perono Garoffo L, Calamari AM, Brenna O, Conti A. Identification of grape and wine allergens as an endochitinase 4, a lipid-transfer protein, and a thaumatin LTP cross-reacting with the peach major allergen. *Journal of Allergy and Clinical Immunology*, 2003, **111**:350–359.
[31] Pastorello EA, Farioli L, Pravettoni V, Ispano M, Scibola E, Trambaioli C, Giuffrida MG, Ansaloni R, Godovac-Zimmermann J, Conti A, Fortunato D, Ortolani C. The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. *Journal of Allergy and Clinical Immunology*, 2000, **106**:744–751.
[32] Palacin A, Quirce S, MD, Armentia A, Fernàndez-Nieto M, Pacios LF, Asensio T, Sastre J, Diaz-Perales A, Salcedo G. Wheat lipid transfer protein is a major allergen associated with baker's asthma. *Journal of Allergy and Clinical Immunology*, 2007, **120**:1132-1138.
[33] Zoccatelli G, Dalla Pellegrina C, Consolini M, Fusi M, Sforza S, Aquino G, Dossena A, Chignola R, Peruffo A, Olivieri M, Rizzi C. Isolation and identification of two lipid transfer proteins in pomegranate (Punica granatum). *Journal of Agricultural and Food Chemistry*, 2007, **55**:11057–11062.
[34] Rigby NM, Marsh J, Sancho AI, Wellner K, Akkerdaas J, van Ree R, Knulst A, Fernández-Rivas M, Brettlova V, Schilte PP, Summer C, Pumphrey R, Shewry PR, Mills EN. The purification and characterisation of allergenic hazelnut seed proteins. *Molecular Nutrition* & *Food Research,* 2008, **52**:251-261.
[35] Bernardi ML, Giangrieco I, Camardella L, Panico MR, Zennaro D, Ferrara R, Palazzo P, Tuppo L, Tamburrini M, Carratore V, Santoro M, Ciardiello MA, Mari A. Biochemical and immunological characterisation of Act d 10 a lipid transfer proteins from green kiwifruit. *Clinical and Translational Allergy*, 2011, **1**:O23.
[36] Wijesinha-Bettoni R, Alexeev Y, Johnson P, Marsh J, Sancho AI, Abdullah SU, Mackie AR, Shewry PR, Smith LR, Mills ENC. The structural characteristics of nonspecific lipid transfer proteins explain their resistance to gastroduodenal proteolysis. *Biochemistry*, 2010, **49**:2130–2139.
[37] van Ree R. Clinical importance of non-specific lipid transfer proteins as food allergens. *Biochemical Society Transactions*, 2002, **30**:910–913.

[38] Astwood JD, Leach JN, Fuchs RL. Stability to food allergens to digestion in vitro. *Nature Biotechnology,* 1996, **14**:1269–1273.

[39] Asero R, Mistrello G, Roncarolo D, Amato S, van Ree R. A case of allergy to beer showing cross-reactivity between lipid transfer proteins. *Annals of Allergy, Asthma and Immunology,* 2001, **87**:65–67.

[40] Duffort OA, Polo F, Lombardero M, Díaz-Perales A, Sánchez-Monge R, García-Casado G, Salcedo G, Barber D. Immunoassay to quantify the major peach allergen Pru p 3 in foodstuffs. Differential allergen release and stability under physiological conditions. *Journal of Agricultural and Food Chemistry*, 2002, **50**:7738–7741.

[41] Pastorello EA, Pompei C, Pravettoni V, Farioli L, Calamari AM, Scibilia J, Robino AM, Conti A, Iametti S, Fortunato D, Bonomi S, Ortolani C. Lipid-transfer protein is the major maize allergen maintaining IgE-binding activity after cooking at 100 °C, as demonstrated in anaphylactic patients with positive double-blind, placebo-controlled food challenge results. *Journal of Allergy and Clinical Immunology*, 2003, **112**:775–783.

[42] Pastorello EA, Vieths S, Pravettoni V, Farioli L, Trambaioli C, Fortunato D, Lüttkopf D, Calamari M, Ansaloni R, Scibilia J, Ballmer-Weber BK, Poulsen LK, Wütrich B, Skamstrup Hansen K, Robino AM, Ortolani C, Conti A. Identification of hazelnut major allergens in sensitive patients with double-blind, placebo-controlled food challenge results. *Journal of Allergy and Clinical Immunology*, 2002, **109**:563–570.

[43] Asero R, Mistrello G, Roncarolo D, Amato S, Falagiani P. Analysis of the heat stability of lipid transfer protein from apple. *Journal of Allergy and Clinical Immunology*, 2003, **112**:1009 –1011.

[44] Ebner C, Birkner T, Valenta R, Rumpold H, Breitenbach M, Scheiner O, Kraft D. Common epitopes of birch pollen and apples. Studies by Western and Northern blot. *Journal of Allergy and Clinical Immunology*, 1991, **88**:588–595.

[45] Blanco C, Carrillo T, Castillo R, Quiralte J, Cuevas M. Latex allergy: clinical features and crossreactivity with fruits. *Annals of Allergy*, *Asthma & Immunology*, 1994, **73:**309–314.

[46] Fernandez-Rivas M, van Ree R, Cuevas M. Allergy to Rosaceae fruits without related pollinosis. *Journal of Allergy and Clinical Immunology*, 1997, **100**:728–733.

[47] Pravettoni V, Primavesi L, Farioli L, Brenna OV, Pompei C, Conti A, Scibilia J, Piantanida M, Mascheri A, Pastorello EA. Tomato Allergy: detection of IgE-binding lipid transfer proteins in tomato derivatives and in fresh tomato peel, pulp, and seeds. *Journal of Agricultural and Food Chemistry*, 2009, **57**:10749–10754.

[48] Asero R, Mistrello G, Amato S. Anaphylaxis caused by tomato lipid transfer protein. *European annals of allergy and clinical immunology*, 2011, **43**:125-126.

[49] Le LQ, Lorenz Y, Scheurer S, Fötisch K, Enrique E, Bartra J, Biemelt S, Vieths S, Sonnewald U. Design of tomato fruits with reduced allergenicity by dsRNAi-mediated inhibition of ns-LTP (Lyc e 3) expression. *Plant Biotechnology Journal*, 2006, **4**:231-242.

[50] Sancho AI, Rigby NM, Zuidmeer L, Asero R, Mistrello G, Amato S, González-Mancebo E, Fernández-Rivas M, van Ree R, Mills EN. The effect of thermal processing on the IgE reactivity of the non-specific lipid transfer protein from apple, Mal d 3. *Allergy*, 2005, **60**:1262-1268.

[51] Gaier S, Marsh J, Oberhuber C, Rigby NM, Lovegrove AL, Alessandri S, Briza P, Radauer C, Zuidmeer L, van Ree R, Hemmer W, Sancho AI, Mills C, Hoffmann-Sommergruber K, Shewry PR. Purification and structural stability of the peach

allergens Pru p 1 and Pru p 3. *Molecular Nutrition & Food Research,* 2008, **52**:S220-S229.

[52] Cavatorta V, Sforza S, Aquino G, Galaverna G, Dossena A, Pastorello EA, 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. *Molecular Nutrition & Food Research,* 2010, **54**:1–6.

[53] Asero R, Mistrello G, Roncarolo D, de Vries SC, Gautier MF, Ciurana CL, Verbeek E, Mohammadi T, Knul-Brettlova V, Akkerdaas JH, Bulder I, Aalberse RC, van Ree R. Lipid transfer protein: a pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. *International Archives of Allergy and Immunology,* 2000, **122**:20–32

[54] Strohalm M, Kavan D, Novak P, Volny M, Havlicek V. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Analytical Chemistry,* 2010, **82**:4648-4651.

[55] Vassilopoulou E, Rigby N, F. Moreno FJ, Zuidmeer L, Akkerdaas J, Tassios I, Papadopoulos NG, Saxoni-Papageorgiou P, van Ree R, Mills C. Effect of in vitro gastric and duodenal digestion on the allergenicity of grape lipid transfer protein. *Journal of Allergy and Clinical Immunology*, 2006, **118**:473-480.

[56] Cavatorta V, Sforza S, Mastrobuoni G, Pieraccini G, Francese S, Moneti G, Dossena A, Pastorello EA, Marchelli R. Unambiguous characterization and tissue localization of Pru P 3 peach allergen by electrospraymass spectrometry and MALDI imaging. *Journal of Mass Spectrometry,* 2009, **44**:891–897.

[57] Kuppannan K, Albers DR, Schafer BW, Dielman D, Young SA. Quantification and characterization of maize lipid transfer protein, a food allergen, by liquid chromatography with ultraviolet and mass spectrometric detection. *Analytical Chemistry*, 2011, **83**:516–524.

[58] Cruz LP, Ribeiro SF, Carvalho AO, Vasconcelos IM, Rodrigues R, Da Cunha M, Gomes VM. Isolation and partial characterization of a novel lipid transfer protein (LTP) and antifungal activity of peptides from chilli pepper seeds. *Protein & Peptide Letters*, 2010, **17**:311-318.

[59] Marzban G, Puehringer H, Deya R, Brynda S, Ma Y, Martinelli A, Zaccarini M, van der Weg E, Housley Z, Kolarich D, Altmann F, Laimer M. Localisation and distribution of the major allergens in apple fruits. *Plant Science*, 2005, **169**: 387-394.

Chapter V - SECTION II

Genomics in food analysis

1. *Food authenticity*

Consumers worldwide, and European ones in particular, are showing an increasing interest in issues related to food, diet and nutrition. The globalization of food markets implies, in fact, that consumers come into contact with a great variety of foods and that they are more and more concerned about the origin of the food they eat. There is a growing enthusiasm among consumers for high quality food with a clear regional identity as a consequence of patriotism, specific culinary, organoleptic qualities or purported health benefits associated with regional products, a decreased confidence in the quality and safety of foods produced outside their local region, country or the EU or concern about animal welfare and 'environmentally friendly' production methods.[1] The determination of food authenticity is, therefore, of paramount importance in food quality control and safety.

Food authentication is the process by which a food is verified as complying with its label description. Therefore, labelling and compositional regulations, which may differ from country to country, have a fundamental place in determining which scientific tests are appropriate for a particular issue. In general, food authenticity issues fall into one of the following categories: economic adulteration of high value foods, misdescription of the geographical, botanical or species origin, non-compliance with the established legislative standards and implementation of non-acceptable process practices. Labelling legislation is there to ensure that food is properly described and ensures that correctly described products remain available to the consumer and that consumer confidence is maintained, which in turn ensures a market place for these foods.[1]

The incorrect labelling of food represents a commercial fraud, considering the consumer acquisition. It is very important to establish that species of high commercial value declared are not substitute, partial or entirely, by other lower value species. The misleading labelling might also have negative implications concerning health, especially for sensitive consumers to non-declared potential allergens. The information given to consumers is also essential for them choosing certain foods over others. That choice might be the reflection of lifestyles, such as vegetarianism, or religious practices, such as Jews and Muslims, where pork meat should be absent. The recent occurrence of several food crises has emphasized food safety and protection of consumer's health as main objectives for the food labelling legislation.[2]

Most food legislation is harmonized throughout the European Union through a number of European Commission (EC) Directives and Regulations. The labelling of food is subject to the general rules laid down in Council Directive 2000/13/EC. The main provision of such Directive is to require the following particulars in the labelling of food: the name under which a product is sold, the list of ingredients, the quantity of certain ingredients or categories of ingredients, the net weight, and for alcoholic drinks with more than 1.2% alcohol by volume, the alcoholic strength by volume, the date of minimum durability i.e. 'use by' for those highly perishable foods from a microbiological point of view, or 'best before', any special storage conditions of use, the name of business name and address of the manufacturer, packager, or seller established within the EU, and instructions.[3] Other particulars need to be given where to omit them would be misleading to the consumer. These include any physical process such as freezing, drying or irradiation of ingredients, and the geographical origin of the food. There is also a requirement to declare the presence of any approved GM ingredients above 0.9% (non-

approved are prohibited), and warn consumers of certain allergens not named in the list of ingredients. A quantitative ingredient declaration is also required for those ingredients highlighted in the name of the food or which consumers would associate with the food.[1]

The EC also adopted special provisions as regards the protection of geographical indications and designations of origin for agricultural products and foodstuffs. The Council Regulation (EC) 510/2006 of 20th March 2006 establishes, in fact, the rules for protecting designations of origin and geographical indications for agricultural products and foodstuffs intended for human consumption. The main provision of Council Regulation (EC) 510/2006 is to ensure that only products genuinely originating in a specific region are allowed in commerce as such, to make Protected Designation of Origin (PDO) and Protected Geographical Indication (PGI) symbols or indications obligatory and to enable an easier identification of these products on the market so as to facilitate controls.[4]

Issues such as tradition and identity play an important role in the perception of food authenticity, and the paradox might occur that a food might be safer if produced with more modern methods although the traditional method of producing the food is what makes it authentic. In the light of serious food safety incidents relating to food adulterations and frauds, consumers think of authentic food as being safe, and food safety and authenticity are undoubtedly linked. Also, food authenticity is often synonym of a positive quality even if people in general are able to recognize that non-traditional high quality foods exist as well (e.g. probiotic yoghurts, functional foods, etc.). In conclusion, adulteration and food fraud are without any doubt interconnected with the concept of food authenticity, even if a more appropriate definition describes food fraud as 'the deliberate and illegal mislabelling of food for economic gain'.[1]

1.1 Authenticity issue

As showed elsewhere,[5] there can be many different and indeed subtle issues concerning labelling which it may be desirable to check by performing chemical tests. However, it is possible to classify the issues into a number of similar topics.

A common authenticity problem is for the species from which a food was made to be misdescribed. This may take the form of substitution of one species for another with a less commercial value, or the mixing of one species with similar material from a cheaper species. Thus, the use of soft wheat in the Italian traditional pasta manufactured,[6] the replacement of Basmati rice with cheaper long grain-varieties,[7] the use of non-premium olive cultivars in the production of olive oils with certificated geographical origin,[8] the employment of cheap fruit pulps (e.g. blackberry or apple) rather than more expensive pulps (e.g. strawberry and raspberry) as ingredients of jams, yoghurts and fruit pies and desserts,[9] are all examples of misdescription.

Claims concerning the species of origin are effectively claims which concern the genetic of the ingredients.[5] The application of recent developments in DNA technology (described further below) for authenticity testing are thus being developed and used in order to verify the match between labelling and the presence or the absence of certain food materials.

2. *Food traceability*

Food authenticity, in general terms, is perceived as an important aspect by consumers at an emotional level because it involves their trust in what they buy. As a consequence, it is well controlled by food manufacturers and legislators. For this reason, traceability within food chains is gaining in importance to the European economy, especially because consumers exert pressure by wanting to know the provenance and authenticity of raw materials used for the production of foods, not just the nutritional value of those foods. In this scenario, the traceability allows to protect the consumer from possible fraud and to preserve individual food choices. In addition, the ability to check the origin of raw materials by using objective and scientific tools increases the value of quality certification (such as trademarks PGI, PDO). This encourages the development of economic marginal areas through the enhancement of local and niche products and also provides incentives to conserve and maintain the local ecotype biodiversity, through increased protection for authentic products.

Nowadays, the development of organisational and management systems allows the food chain to be completely accessible and unambiguous. The instrument to ensure the full transparency of the food chain is the food traceability, i.e., the documented identification of incoming materials and all of the operations that contribute to the formation of the finished product as it is sold to the consumers. It is a system that allows the consumer to know, as quickly and as well as possible, origin and methods of production, transfer and commercialization of food. The main purpose is to protect consumers' health and to assign responsibilities to those who are involved in agricultural and agro-industry productions.

This checking policy requires the ability to track raw materials at every stage and every step of the food chain. This was established by the Regulation 178/2000 (EC): 'traceability means the ability to trace and follow a food, feed, food-producing animal or substance intended to be, or expected to be incorporated into a food or feed, through all stages of production, processing and distribution'. In practice, tracking materials/products through a food chain means collecting data that are generated along the path from farm to table, every time a stage production has been completed anywhere in the chain: seed or nursery industry, farm, first processing company, enterprise transformation, distribution and consumer.

From 1 January 2005 the European Directive 2000/13/EC and the European Regulation 178/2002 came into effect, according to which and to ensure transparency, all production and trading companies should adopt such procedures, exposed in the directive, in order to pinpoint the source of all the elements used in a given agro-food production process. This means that individual companies must provide data collected along the chain, in various stages of processing, organizing them in order to make available to the authorities all of the information concerning the whole production process.

The benefits of adopting a system of traceability are multiple; primarily, they can affect both consumers, who are protected from being sold an inferior product with a false description, and honest traders, who are protected from unfair competition, thus increasing competitiveness and facilitating their inclusion in the certification systems.

3. Tools for assessing food authenticity

One of the crucial point of the food traceability systems relies on the availability of simple, reliable and efficient tools allowing the immediate identification of ingredients in processed or composite mixtures, during any step of the food chain.

Many different chemical and biochemical techniques have been developed for determining the authenticity of food and in recent years methods based on DNA analysis have become more important. This is because some techniques, such as immunoassays, work well with raw foods but lose their discrimination when applied to cooked or highly processed food. Also many techniques do not easily distinguish between closely related materials at the chemical level. For example, olive and hazelnut oils are similar chemically so the usual analytical methods cannot be applied to detect the adulteration of olive oils with hazelnut oil. Conventional chemical methods are also not always able to detect country or region of origin of olive oil.[10]

DNA provides many advantages that make it especially attractive in studies of diversity and relationships. These advantages include: freedom from environmental and pleiotropic effects (molecular markers do not exhibit phenotypic plasticity, while morphological and biochemical markers can vary in different environments and are under polygenic control); a potentially unlimited number of independent markers are available, unlike morphological or biochemical data; DNA characters can be more easily scored as discrete states of alleles or DNA base pairs, while some morphological, biochemical and field evaluation data must be scored as continuously variable characters that are less amenable to robust analytical methods;[11] DNA is more resilient to destruction by food processing (particularly cooking and sterilization) than other marker substances; DNA analysis has discriminating power because ultimately the definition of a variety or species is dependent on the sequence of the DNA in its genome.

These advantages do not imply that other more traditional data used to characterize biodiversity are not valuable. On the contrary, morphological, ecological and other "traditional" data will continue to provide practical and often critical information needed to characterize genetic resources. DNA analysis has also the great disadvantage that, given a genetically defined material, it can not assess where it has been cultivated or the different environmental conditions in which it has been cultivated. The same specific variety of tomato, just to name an example, cultivated in Italy or in China, will probably be very different, given the different environmental conditions which in turn will result in different phenotypes, but will yield the same results at the genetic analysis. At the same time, nucleic acid-based technologies are developing rapidly and the informed adoption of suitable methods by the food industry has the potential to greatly simplify methods of authentication.

3.1 DNA hybridization techniques

Initial studies using DNA to detect specific species used relatively simple methods, in which labelled DNA probes were hybridized to samples of amplified genomic DNA covalently attached to nylon membranes in a slot- or dot-blot format.[12] This most basic regime lacked of specificity, due to the high cross-reactivity among the DNA sequences of related species. The use of further species-specific DNA sequence probes has subsequently been described and the successful discrimination of meat from different

species, even closely related such as sheep and goat, has been accomplished, both in raw materials and in processed, heated and canned meat.[13,14]
Being quite cumbersome, DNA testing by hybridization has been largely replaced by other DNA-based approaches to species identification.

3.2 DNA sequencing

DNA sequencing based techniques provide better discrimination among the different species/varieties than crude hybridization tests.
Currently, dye-terminator sequencing technique based on the Sanger method is the standard method in automated sequencing analysis. The basic technique related with dye terminator sequencing and phylogenetic analysis is illustrated in Fig 1. In dye-terminator sequencing, the four dideoxynucleotide chain terminators are labelled with fluorescent dyes, each with a different wavelength of fluorescence emission. The different oligonucleotides having different length are then separated by capillary electrophoresis with fluorescence detection. The main advantages of this technique are its robustness, automation and high accuracy (>98%). On the other hand, the limitations of this technique include dye effects due to differences in the incorporation of the dye-labeled chain terminators into the DNA fragment. Such incorporation of dye can result in unequal peak heights and shapes in the electronic DNA sequence trace chromatogram after capillary electrophoresis. Another drawback is its inability to handle long sequences.[15]

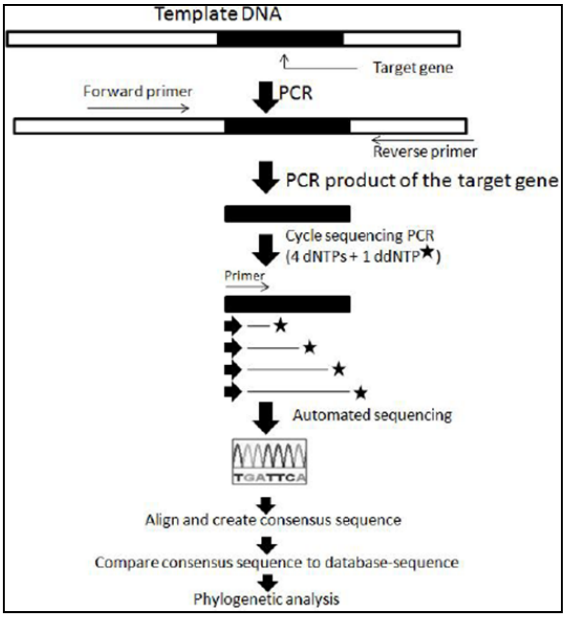


Fig 1. Schematic diagram summarizing the sequencing of a target DNA sequence. Adapted from reference 16.

A new generation of non-Sanger based sequencing technologies has been evolving on its promise of sequencing DNA at unprecedented speed, thereby also enabling impressive scientific achievements and novel biological applications. These techniques are based on complete genomes rather than just short sequences of a single gene. Briefly, the procedure works by first capturing billions of single molecules of sample DNA on an application-specific proprietary surface within two flow cells. These captured strands serve as templates for the sequencing-by-synthesis. Polymerase and one fluorescently

labelled nucleotide (C/G/A/T) are added. The polymerase catalyzes the sequence-specific incorporation of fluorescent nucleotides into nascent complementary strands on all the templates. After a wash step, which removes all free nucleotides, the incorporated nucleotides are imaged and their positions are recorded. The fluorescent group is removed in a highly efficient cleavage process, leaving behind the incorporated nucleotide. The process continues through each of the other three bases (Fig 2). DNA sequencing data from next generation platforms typically present shorter read lengths, higher coverage and different error profiles compared with Sanger sequencing data. Several software packages have been created especially to cope with the next generation sequencing data.[15]

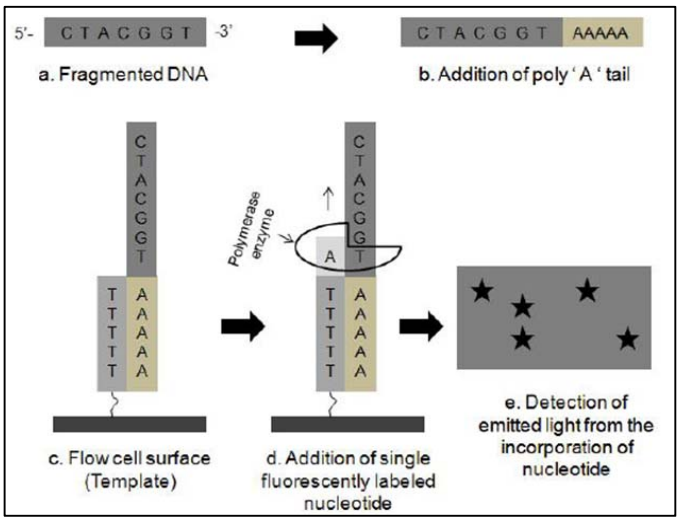


Fig 2. Basic workflow of next generation sequencing technique. Adapted from reference 17.

Current interest is also in the DNA barcoding, especially if applied to plants with the aim to classify unknown plants in known classification schemes. DNA barcoding is a technique for characterizing species of organisms using a short DNA sequence from a standard and agreed-upon position in the genome. DNA barcode sequences are very short relative to the entire genome and they can be obtained reasonably quickly and cheaply.[18] The success of species-level assignment of plants using Basic Local Alignment Search Tool (BLAST) with individual barcodes has already obtained with different species, such as woody trees, shrubs and palms.[19] There are some limitations which hamper the wide spread use of DNA barcode. First of all, a DNA sequencing- based identification system can work only if every species to be identified has its own sequences annotated in databases. An incomplete database, in fact, will allow to determine if the unknown sequence is different from those already known, but it will not imply the definitive identification of a species or its affiliation to a new species.[20] Secondly, in order that the molecular approach is fulfil, the discrimination between intra- and inter-specific variability must be accomplished.

Limitations aside, DNA barcoding has many chances to be applied in the field of biological systematic, as a tool in diagnosis and food traceability, and for the phylogenesis analysis.

3.3 DNA molecular markers

During the last few decades, the use of molecular markers, revealing polymorphisms at the DNA level, has been playing an increasing part in the field of food identification, to prevent the adulteration of target plant with other plant species. Due to the rapid developments in the field of molecular genetics, varieties of different techniques are now available, differing in genomic abundance, level of polymorphism detected, locus specificity, reproducibility, technical requirements and financial investment.

A molecular marker is a DNA sequence that is readily detected and whose inheritance can be easily monitored. The uses of molecular markers are based on the naturally occurring DNA polymorphism, which forms basis for designing strategies to exploit for applied purposes. A marker must be polymorphic, i.e. it must exit in different forms so that chromosome carrying the mutant genes can be distinguished from the chromosomes with the normal gene by a marker it also carries. Genetic polymorphism is defined as the simultaneous occurrence of a trait in the same population of two discontinuous variants or genotypes. DNA markers seem to be the best candidates for efficient evaluation and selection of plant material. Unlike protein markers, DNA markers segregate as single genes and they are not affected by the environment. DNA is easily extracted from plant materials and its analysis can be cost and labour effective.[11]

An ideal molecular marker must have some desirable properties. For example, it must be high polymorphic, as it is polymorphism that is measured for genetic diversity studies; it should have a codominant inheritance, allowing the determination of homozygous and heterozygous states of diploid organisms; an ideal marker should be evenly and frequently distributed throughout the genome; it should be easy, fast and cheap to detect by means of easy and fast assays, providing high reproducibility even between different laboratories.[11]

It is extremely difficult to find a molecular marker, which would meet all the above criteria. A wide range of molecular techniques is available that detects polymorphism at the DNA level. Various types of molecular markers are utilized to evaluate DNA polymorphism and are generally classified as hybridization-based markers and polymerase chain reaction (PCR)-based markers. In the former, DNA profiles are visualized by hybridizing the restriction enzyme-digested DNA, to a labelled probe, which is a DNA fragment of known origin or sequence. PCR-based markers involve in vitro amplification of particular DNA sequences or loci, with the help of specifically or arbitrarily chosen oligonucleotide sequences (primers) and a thermostable DNA polymerase enzyme. The amplified fragments are separated electrophoretically and banding patterns are detected by different methods such as staining and autoradiography. This technique presents a high potential due to its fastness, simplicity sensibility and specificity.[11]

There are several technical considerations specific to the use of PCR for amplifying DNA derived from food. First, in many instances the test samples will be highly processed and might have been heated to temperatures over than 100 °C to cook or sterilize them. This results in DNA degradation and so PCR primers should be designed to amplify fragments of 200 bp or less.[21] Second, across the spectrum of foodstuffs to be examined, many different food matrices will be encountered including those high in oils, fats, vegetable material, animal tissue and various additives and fillers. The test sample might contain only raw ingredients or some or all of the components might have been milled, boiled, dried and so on. This means that DNA extraction procedures have to be optimized before

analysis begins, in order to ensure that sufficient test DNA is extracted and that inhibitors of the PCR are reduced or eliminated.[10] To date, no single extraction method has proved useful with all the different matrices encountered. This problem is made more acute because quantification of the amounts of DNA from each species present often is required.

3.3.1 Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) is a technique in which organisms may be differentiated by analysis of patterns derived from cleavage of their DNA. If two organisms differ in the distance between sites of cleavage of particular restriction endonucleases, the length of the fragments produced will differ when the DNA is digested with a restriction enzyme. Size fractionation is achieved by gel electrophoresis and, after being transferred to a membrane by Southern blotting, fragments of interest are identified by hybridization with radioactive labelled probe (Fig 3). The similarity of the patterns generated can be used to differentiate species from one another. Such a polymorphism can be used to distinguish plant species, genotypes and, in some cases, individual plants.[22] They have their origin in the DNA rearrangements that occur due to evolutionary processes, point mutations within the restriction enzyme recognition site sequences, insertions or deletions within the fragments, and unequal crossing over.[23]

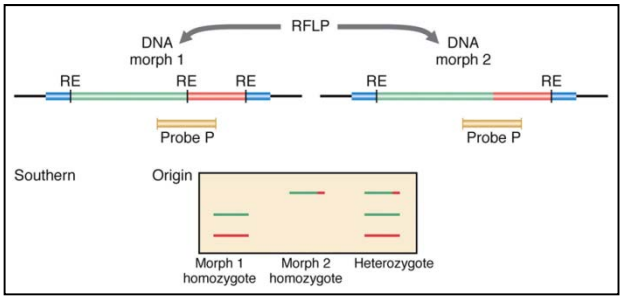


Fig 3. Schematic representation of RFLP technique.

RFLPs are generally found to be moderately polymorphic. In addition to their high genomic abundance and their random distribution, RFLPs have the advantages of showing codominant alleles and having high reproducibility. They are very reliable markers in linkage analysis and breeding and can easily determine if a linked trait is present in a homozygous or heterozygous state. On the other side, the utility of RFLPs has been hampered due to the large quantities (1–10 μg) of purified, high molecular weight DNA required for each DNA digestion and Southern blotting. The requirement of radioactive isotope makes the analysis relatively expensive and hazardous. The assay is time-consuming and labour-intensive and only one out of several markers may be polymorphic. Despite these limitations, RFLPs have been applied in diversity and phylogenetic studies ranging from individuals within populations or species, to closely related species.[11] RFLPs have been widely used in gene mapping studies because of their high genomic abundance.[24] They also have been used to investigate relationships in the tomato genus *Lycopersicon*,[25] as fingerprinting tool,[26] and for diversity studies.[27]

3.3.2 Random Amplified Polymorphic DNA (RAPD)

RAPD is a PCR-based technology. The method is based on enzymatic amplification of target or random DNA segments with arbitrary primers. In 1990 Welsh and McClelland developed a new PCR-based genetic assay namely randomly amplified polymorphic DNA (RAPD).[28] This procedure detects nucleotide sequence polymorphisms in DNA by using a single primer of arbitrary nucleotide sequence. In this reaction, a single species of primer anneals to the genomic DNA at two different sites on complementary strands of DNA template. If these priming sites are within an amplifiable range of each other, a discrete DNA product is formed through thermo cyclic amplification. Amplified products are separated on agarose gels in the presence of ethidium bromide and view under ultraviolet light and presence and absence of band will be observed (Fig 4).

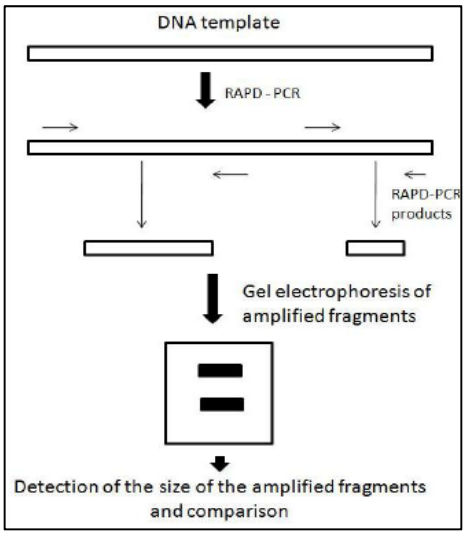


Fig 4. The principle of RAPD-PCR technique. Arrows indicate primer annealing sites. Adapted from reference 29.

On an average, each primer directs amplification of several discrete loci in the genome, making the assay useful for efficient screening of nucleotide sequence polymorphism between individuals. These polymorphisms are considered to be primarily due to variation in the primer annealing sites, but they can also be generated by length differences in the amplified sequence between primer annealing sites. The main advantage of RAPDs is that they are quick and easy to assay. Because PCR is involved, only low quantities of template DNA are required. Since random primers are commercially available, no sequence data for primer construction are needed. Moreover, RAPDs have a very high genomic abundance and are randomly distributed throughout the genome. They are dominant markers and hence have limitations in their use as markers for mapping. The main drawback of RAPDs is their low reproducibility. Besides, RAPD analyses generally require purified, high molecular weight DNA, and precautions are needed to avoid contamination of DNA samples because short random primers are used that are able to amplify DNA fragments in a variety of organisms. Finally, RAPD markers are not locus-specific, and band profiles can not be interpreted in terms of loci and alleles (dominance of markers).[11] RAPD markers have been employed for the construction of intra-specific tomato genetic map,[30] and for the assessment of phylogenetic diversity and relationships of some tomato varieties.[31]

3.3.3 Amplified Fragment Length Polymorphism (AFLP)

AFLP is a DNA fingerprinting technique, which detects DNA restriction fragments by means of PCR amplification. It involves the restriction of genomic DNA, followed by ligation of adaptors complementary to the restriction sites and selective PCR amplification of a subset of the adapted restriction fragments. These fragments are viewed on denaturing polyacrylamide gels either through autoradiographic or fluorescence methodologies.[32] The PCR primers consist of a core sequence (part of the adapter), and a restriction enzyme specific sequence and 1–5 selective nucleotides (the higher the number of selective nucleotides, the lower the number of bands obtained per profile). The AFLP banding profiles are the result of variations in the restriction sites or in the intervening region (Fig 5).

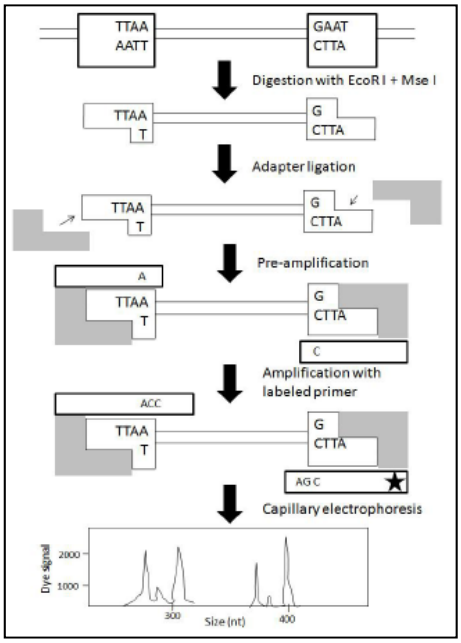


Fig 5. A schematic flow chart showing the principle of the AFLP method. Adapted from reference 33.

The strengths of AFLPs lie in their high genomic abundance, considerable reproducibility and automatism, the generation of many informative bands per reaction, their wide range of applications, and the fact that no sequence data for primer construction are required. On the other side, AFLP method suffers from the need for purified, high molecular weight DNA and the dominance of alleles, which does not allow the discrimination between the homozygous and the heterozygous genotype.[11] AFLPs have been applied in studies involving the construction of an intraspecific map of the tomato genome,[34] and for the identification of markers linked to a tomato gene resistance.[35]

3.3.4 Minisatellites (VNTRs and HVRs)

The term minisatellites was introduced by Jeffrey in 1985. These loci contain tandem repeats that vary in the number of repeat units between genotypes and are referred to as variable number of tandem repeats (VNTRs) (i.e. a single locus that contains variable number of tandem repeats between individuals) or hypervariable regions (HVRs) (i.e. numerous loci containing tandem repeats within a genome generating high levels of

polymorphism between individuals).[36] They consist of chromosomal regions containing tandem repeat units of a 10–50 base motif, flanked by conserved DNA restriction sites. Locus specific probes can be developed by molecular cloning of DNA restriction fragments, subsequent screening with a multilocus minisatellite probe and isolation of specific fragments (Fig 6). Variation in the number of repeat units, due to unequal crossing over or gene conversion, is considered to be the main cause of length polymorphisms. Due to the high mutation rate of minisatellites, the level of polymorphism is substantial, generally resulting in unique multilocus profiles for different individuals within a population.

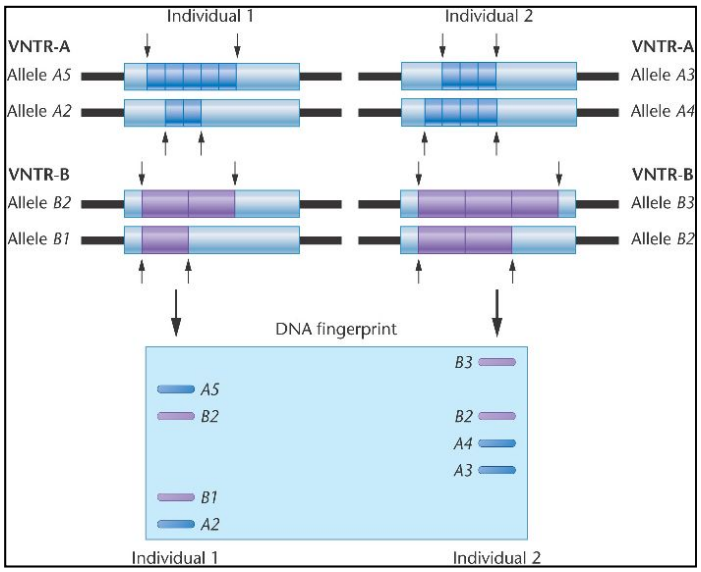


Fig 6. The principle of VNTR technique. Arrows indicate restriction recognition sites.

The main advantages of minisatellites are their high level of polymorphism and high reproducibility while they hold the disadvantages of RFLPs, due to the high similarity in methodological procedures. Moreover, highly informative profiles are generally observed due to the generation of many informative bands per reaction. In that case, band profiles can not be interpreted in terms of loci and alleles.[11]

Concerning the application of VNTR markers for the characterization of tomato genome, several studies showed the $(GATA)_4$ motif as the one with the high polymorphic power,[37,38] although it is not uniformly spread all along the genome. Other polymorphic motives along the tomato genome have been found, such as $(GGAT)_4$, $(CCTA)_4$ and $(CA)_8$.[38] The origin and the function of these motives have not been disclosed, yet, although they can contribute to the polymorphisms. Recent experiments of FISH (Fluorescent In Situ Hybridization) showed that they are mainly placed in the pericentromeric heterochromatin and in the NOR regions.[39]

3.3.5 Microsatellites or Simple Sequence Repeat (SSR)

They are sections of DNA, consisting of repeating mono-, di-, tri-, tetra- or penta-nucleotide units in tandem that are arranged throughout the genomes of most eukaryotic species.[40] Microsatellite markers, developed from genomic libraries, can belong to either the transcribed region or the non transcribed region of the genome, and rarely is there information available regarding their functions. Microsatellites, like minisatellites,

represent tandem repeats, but their repeat motifs, or *core*, are shorter (1–6 base pairs). If nucleotide sequences in the flanking regions of the microsatellite are known, specific primers (generally 20–25 bp) can be designed to amplify the microsatellite by PCR (Fig 7).

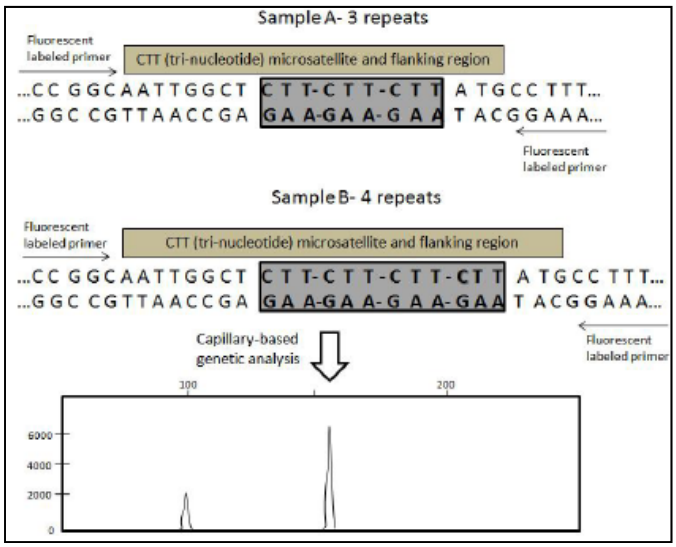


Fig 7. Representation of a CTT (tri-nucleotide) microsatellite and flanking region and the detection method. Arrows indicate positions of PCR primers. Two length variants are shown (A and B). Adapted from reference 41.

Polymerase slippage during DNA replication, or slipped strand mispairing, is considered to be the main cause of variation in the number of repeat units of a microsatellite, resulting in length polymorphisms that can be detected by gel electrophoresis or capillary electrophoresis.

The strengths of microsatellites include the codominance of alleles, their high genomic abundance in eukaryotes and their random distribution throughout the genome. Because the technique is PCR-based, only low quantities of template DNA (10–100 ng per reaction) are required. Due to the use of long PCR primers, the reproducibility of microsatellites is high and analyses do not require high quality DNA. One of the main drawbacks of microsatellites is that high development costs are involved if adequate primer sequences for the species of interest are unavailable. Although microsatellites are in principle codominant markers, mutations in the primer annealing sites may result in the occurrence of null alleles (no amplification of the intended PCR product), which may lead to errors in genotype scoring. A very common observation in microsatellite analysis is the appearance of stutter bands that are artefacts in the technique that occur by DNA slippage during PCR amplification. These can complicate the interpretation of the band profiles because size determination of the fragments is more difficult and heterozygotes may be confused with homozygotes.[11]

Microsatellite sequences are especially suited to distinguish closely related genotypes; because of their high degree of variability, they are, therefore, favoured in population studies and for the identification of closely related cultivars.

Concerning tomato, despite its low level of genetic diversity, SSRs can be highly informative and represent a powerful tool for variety discrimination.[42,43,44] Moreover, several studies have found that genic SSRs are useful for estimating genetic diversity

within tomato collection,[45,46] and a very useful tool to trace tomato cultivars in tomato food chains.[47]

3.3.6 Cleaved Amplified Polymorphic Sequence (CAPS)

CAPS are DNA fragments amplified by PCR using specific 20–25 bp primers, followed by digestion of the PCR products with a restriction enzyme. Subsequently, length polymorphisms resulting from variation in the occurrence of restriction sites are identified by gel electrophoresis of the digested products (Fig 8). CAPS have also been referred to as PCR-Restriction Fragment Length Polymorphism (PCR-RFLP).

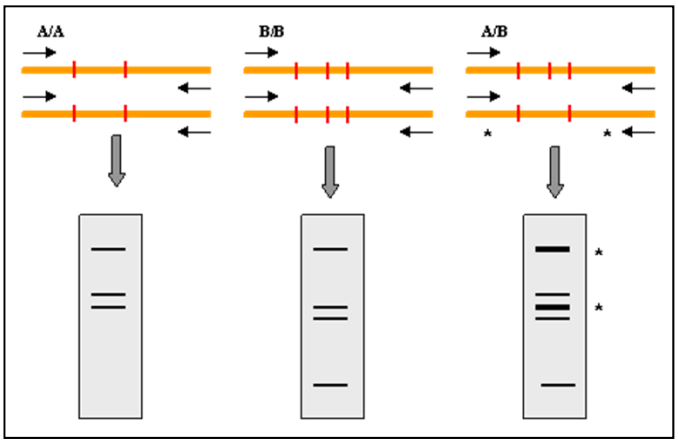


Fig 8. A schematic representation showing the principle of the CAPS method. Arrows indicate positions of PCR primers.

Advantages of CAPS include the involvement of PCR requiring only low quantities of template DNA (50–100 ng per reaction), the codominance of alleles and the high reproducibility. Compared to RFLPs, CAPS analysis does not include the laborious and technically demanding steps of Southern blot hybridization and radioactive detection procedures. In comparison with RFLP analysis, CAPS polymorphisms are more difficult to find because of the limited size of the amplified fragments (300–1800 bp vs 0.5-3.0 kb in size of RFLPs).[11]

Due to their suitability, CAPS markers have been deeply employed in tomato for gene mapping studies,[48,49] especially those associated to resistance factors,[50,51] for breeding purpose,[52] and for the distinction of closely related tomato accessions.[43]

3.3.7 Sequence Characterized Amplified Region (SCAR)

SCARs are DNA fragments amplified by the PCR using specific 15–30 bp primers, designed from nucleotide sequences established from cloned RAPD fragments linked to a trait of interest (Fig 9). By using longer PCR primers, SCARs do not face the problem of low reproducibility generally encountered with RAPDs. Length polymorphisms are detected by gel electrophoresis.[53]

The main advantage of SCARs is that they are quick and easy to use. In addition, SCARs have a high reproducibility and are locus-specific. Due to the use of PCR, only low quantities of template DNA are required (10–100 ng per reaction). The main drawback includes the need for sequence data to design the PCR primers and the need to clone the sequence to be amplified, which can be cost and time-dispending.[11]

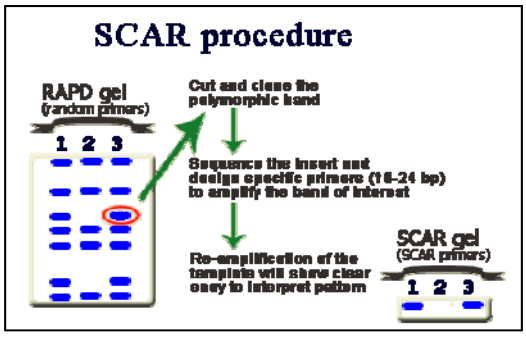


Fig 9. A schematic representation showing the principle of the SCAR method.

As for the CAPS markers, SCARs are locus specific and have been applied in tomato gene mapping,[54,55] and for marker assisted selection.[56]

3.4 Single Nucleotide Polymorphism (SNP)

In recent years, there has been emphasis on the development of newer and more efficient molecular marker systems involving inexpensive non gel-based assays with high throughput detection systems. Availability of single nucleotide polymorphism is one such development that will be addressed in this paragraph.

Single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide (A, T, G or C) differs among members of a species. SNP is the most abundant marker system both in animal and plant genomes.

The extraordinary abundance of SNPs largely makes them the most attractive molecular marker system developed so far. In plant systems, the SNPs seem to be more abundant than even those in the human genome, so that in preliminary studies conducted in wheat, one SNP per 20 bp,[57] and in the maize genome, one SNP per 70 bp,[58] have been recorded in certain regions of these genomes. SNPs may be found both in the non-repetitive coding or regulatory sequences and in the repetitive non-coding sequences. When present in the coding sequences, they may or may not determine the mutant phenotype, but will show 100% association with the trait and will therefore, be very useful, both for MAS (Marker-Assisted Selection) and for gene isolation. In tomato, the frequency of SNP has been estimated to be relatively lower, one SNP per 4000-8500 bp in coding regions:[59] the paucity of SNPs in tomato ESTs may due to the low level of polymorphism, well know in tomato species, and to a limited genotype sampling in the databases. On the other side, tomato genetic variability has been showed to be higher in non-coding sequences (non translated regions and introns), where polymorphisms are better tolerated than in coding regions.[59] This implies that the analysis of non coding regions may be a precious tool for the detection of diversity among tomato varieties.

3.4.1 Methods used for discovery and identification of new SNPs

RFLPs, RAPDs and SSRs which were the markers of choice during the last two decades, need gel-based assays and are, therefore, time consuming and expensive. Therefore, emphasis is now shifting towards the development of SNP molecular markers, which can be detected through non gel-based assays and which allow a high-throughput in the amount of data.

Locus specific-PCR amplification
In this approach, locus specific PCR primers are synthesized from the available genomic sequences and PCR amplification is undertaken using DNA samples from several individuals. PCR amplified products are sequenced and sequence differences are used for discovery of new SNPs. Since prior sequence information is necessary, the method can be used only for genomic regions with known sequences.[60]

Alignment among available genomic sequences
The development of SNPs involving non gel-based assays has recently been facilitated by the availability of genome-wide sequences and EST databases. Alignment of genomic and EST sequences is the most convenient method for discovery of SNPs, if genomic/EST sequences from a heterozygote or from more than one individuals of the same species are available in the databases. The alignment of sequences is automated through computer software, and will allow detection of SNPs in a cost-effective manner. However, a comparison of genomic sequences will detect SNPs, both in coding and non-coding regions, while that of ESTs will detect SNPs only in the coding region. Sequences experimentally obtained from a shotgun library may also be aligned to available genomic sequences to discover new SNPs.[60]

Whole genome shotgun sequences
In this approach, random clones from the genomic library prepared from a mixture of DNA from several individuals are sequenced. This should require several fold coverage of the whole genome before SNPs can be detected by alignment of sequences belonging to the same locus.[60]

Overlapping regions in BACs and PACs
This is one of the common methods for the detection of SNPs in organisms that have been used for genome sequencing (e.g. Arabidopsis, rice). Since the overlapping sequences from BACs/PACs (bacterial/P1 artificial chromosomes) may be derived from genomes of different individuals, an SNP in the overlapping region can be detected by a mismatch.[61]

3.4.2 Methods used for genotyping individuals at SNP loci
A number of methods for SNP genotyping are now available, relying on the ability to distinguish a perfect match from a single base mismatch. The instrumentation and the techniques which recently became available, are allowing accurate genotyping of individual plants. The assays used for SNP genotyping can be broadly classified into two groups, the gel-based assays and the non gel-based assays, the latter being preferred in most laboratories to economize on time and money.

Gel-based assays for SNP detection

RFLP, AFLP and CAPS-based assay: the presence of SNPs can be detected by RFLP, AFLP or CAPS conducted on PCR products, whenever such SNP generates or destroys a specific restriction site for an enzyme. The PCR product in this method is subjected to restriction digestion with individual enzymes to detect differences in patterns, which will be due to changes in the restriction sites.[60]

Single-strand conformation polymorphism: single-strand conformation polymorphism (SSCP) technology allows detection of polymorphism due to differences of one or more base pairs in the PCR products and is therefore suitable for SNP detection. The technique relies on the secondary structure being different for single strands derived from PCR products that differ by one or more nucleotides at an internal site. In order to detect SNPs by this method, PCR product carrying the SNP site is denatured and electrophoretically separated in neutral acrylamide gel (Fig 10).[62] Any difference between the wild strain and a mutant genotype will suggest the presence of SNP.

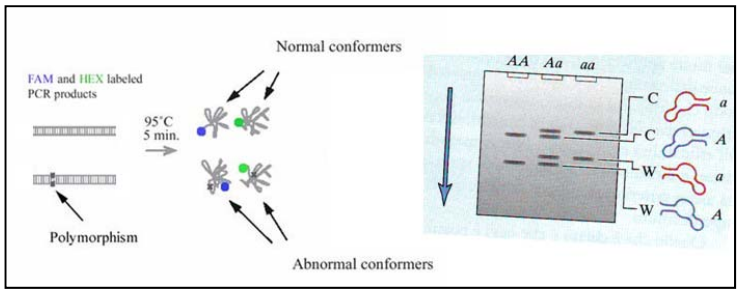


Fig 10. A schematic representation showing the principle of the SSCP method.

Allele-specific amplification for SNP genotyping: SNPs can also be detected by designing allele-specific primers for individual SNP sites. Different fluorochromes are attached at the 5' ends and the polymorphic nucleotides are attached at the 3' ends of the two primers. These allele-specific primers, when used for PCR with pre-amplified DNA as the template, the amplified product will be allele-specific and could be identified either on polyacryamide gel or on an automatic sequencer (Fig 11).[63]

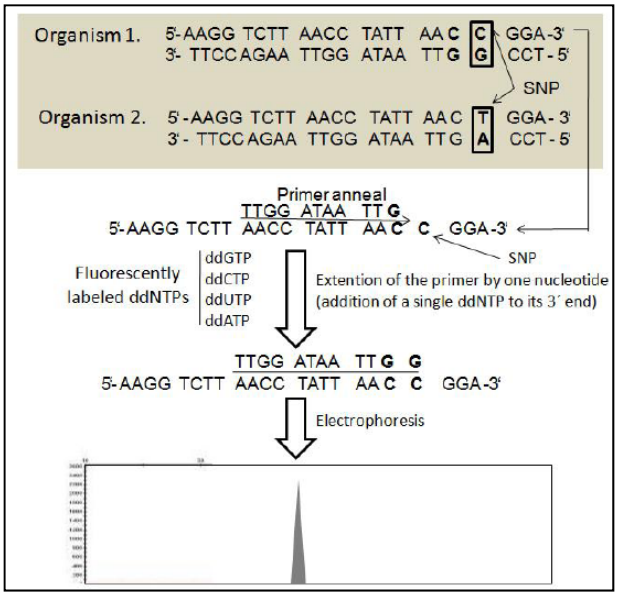


Fig 11. A flow-chart showing the basic principle of allele-specific amplification for SNP genotyping. Adapted from reference 64.

Mismatch at 3' end leading to failure in PCR reaction: if SNP is present at the 3' end of an amplicon template, it can be detected simply by the failure of amplification due to mismatch between the primer sequence and the binding site in the template, although it

may be difficult to distinguish this failure of PCR due to SNP from PCR failure due to other reasons.[60]

Non gel-based assays for SNP detection

TaqMan assay: in an assay described as TaqMan, an oligonucleotide TaqMan probe is labelled with a fluorescent reporter molecule (e.g. FAM or VIC) at the 5' end and a quencher (e.g. TAMRA) at the 3' end. The probe on hybridization to the template DNA is degraded at its 5' end due to exonuclease activity of Taq polymerase (TaqMan), so that the reporter is released leading to a rise in fluorescence signal (Fig 12).

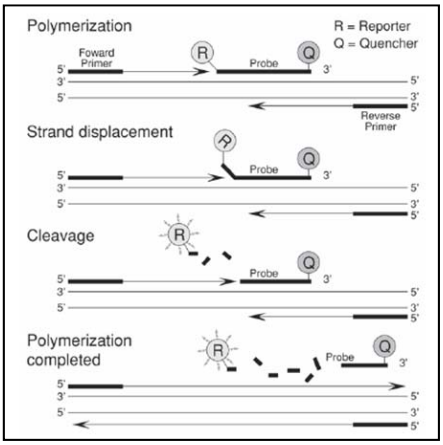


Fig 12. A schematic representation showing the principle of the TaqMan method.

However, when due to the presence of an SNP, the probe mismatches with the template leading to failure in duplex formation, no such degradation at the 5' end of the probe is possible and there is no rise in fluorescence signal.[65]

Molecular beacons: molecular beacons are allele-specific hairpin-shaped oligonucleotide hybridization probes that become fluorescent upon target binding.[66]
The probe (molecular beacon) will be specific for the target SNP sequence, and the sequences at its two ends will be complementary to each other. The two ends of the oligonucleotide are labelled just like the oligonucleotide probe used in TaqMan assay. The probe in isolation (when not forming a duplex with the target DNA) generates a hairpin structure due to self-annealing of its two ends, thus quenching the reporter. But when the probe anneals with the template, it gets linearized, thus separating the reporter from the quencher and permitting fluorescence signal (Fig 13).

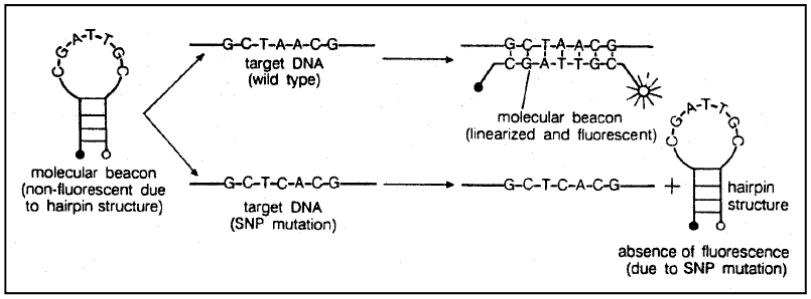


Fig 13. Diagrammatic representation of the operation of molecular beacons. Adapted from reference 67.

Several molecular beacons, each designed to use a different target and labelled with a different fluorophore, can be used to distinguish multiple targets in the same reaction mix.[67]

Oligonucleotide ligation assay: in this approach, two independent probes (one is 5' biotinylated and the other 3' fluorescent-labelled) are used for hybridization with PCR product, so that when the probe matches the product, the two probes anneal with the PCR product and undergo ligation resulting in an oligonucleotide which is biotinylated at the 5' end and fluorescent-labelled at the 3' end.[68] The ligation product, which is fluorescent labelled at the 3' end, is captured on a solid streptavidin-coated matrix due to biotinylation at its 5' end, and the signal is detected by autoradiography (Fig 14). However, when there is a mismatch due to the presence of SNP, fluorescent-labelled and biotinylated oligonucleotides are unable to ligate, so that the oligonucleotides captured by streptavidin, carry no signal.

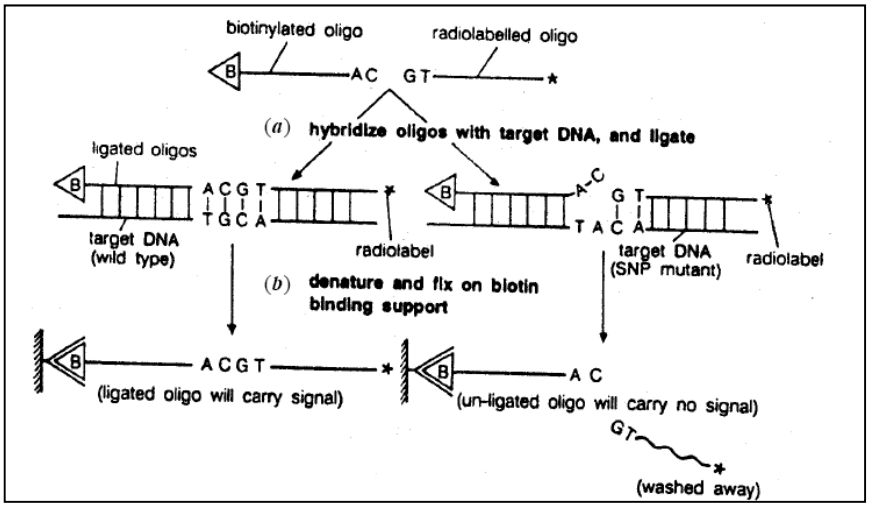


Fig 14. Diagrammatic representation of the oligonucleotide ligation assay depicting gene detection through ligation of hybridized oligonucleotide probes. Adapted from reference 68.

DNA chips and microarrays: DNA chips and microarrays of immobilized oligonucleotides of known sequences, which differ at specific sites of individual nucleotides (at the site of SNP), can also be used for detection of SNPs. The technique is actually suitable to score several SNPs in parallel from each sample in a multiplexed fashion. Four oligonucleotides in a column of an array will differ only at the SNP site and only one will be fully homologous. When such an array is hybridized with biotinylated PCR product, the perfect match will allow binding and mismatched products will be washed away. The perfect match in each case can be detected through a detection system.[69]

Nowadays, the development of new high performance probes, analogues of DNA, has allowed the improvement of such a method. Such a probe, named Peptide Nucleic Acid (PNA) has several advantages which bypass many problems often encountered when using oligonucleotide probes: higher affinity and selectivity towards the complementary DNA sequences, outstanding chemical and biological stability, higher independence of hybridization from environment conditions (pH, ionic strength), easiness of

functionalization in order to obtain the desired chemical, physical or biological characteristics.[70]

The selectivity of the probes can be further increased by using modified PNA monomers.[71] One of the most efficient modifications was the introduction of PNA monomers derived from chiral amino acids with positively charged side chains (chiral PNAs),[72,73] which were shown to be selective in recognition of single point mutations with different techniques.[74,75]

Dynamic allele-specific hybridization: this technique is based on the differences in melting temperatures between duplexes resulting due to perfect match and mismatch between the PCR product and an allele-specific oligonucleotide, 15–21 bases long.[76] The differences in the melting temperatures are detected by the use of a dye, which intercalates in a duplex DNA molecule and emits fluorescence. For PCR amplification, two primers are used, one of which is biotinylated to allow immobilization of the PCR product on a solid support. The immobilized PCR product is denatured, so that only biotinylated single-stranded DNA is retained on the solid support and the other strand is washed away. The biotinylated single-stranded DNA on the solid support is then hybridized to allele-specific oligonucleotide probe containing the SNP site. The duplex formed with the probe is detected by a low-cost fluorescent intercalating dye specific for double-stranded DNA (Fig 15). The dye emits fluorescence proportionate to the amount of double stranded DNA produced due to probe–target hybridization. When this hybrid duplex is denatured steadily by increasing temperature, the melting can be followed by the reduction in the above fluorescence. A rapid and sudden fall in fluorescence indicates the melting temperature (T_m) of the duplex. A single base mismatch results in a dramatic lowering of this melting temperature, thus permitting detection of SNP.

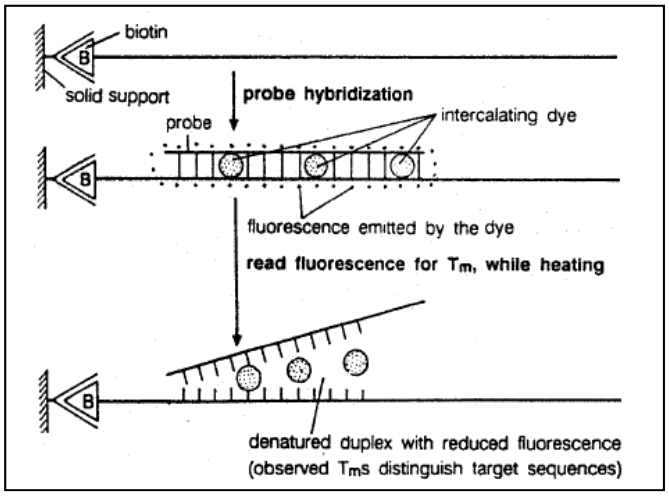


Fig 15. Diagrammatic representation of the dynamic allele-specific hybridization, used for SNP genotyping. Adapted from reference 76.

Temperature modulated heteroduplex analysis usin dHPLC WAVE™ system: SNP detection sometimes requires screening for a sequence variant without any a priori knowledge of the exact location of a mutation in a given sequence. This will be possible if the wild type DNA sample is available for comparison with SNP mutant. For this purpose, denaturing high pressure liquid chromatography (dHPLC) has been used,

where each SNP yields a unique chromatography pattern with temperature modulated heteroduplex analysis or chromatography. The characteristic chromatography pattern can be used not only to identify a novel sequence variant, but also as a diagnostic tool, if an SNP is already characterized. It has been shown that this approach has extraordinary sensitivity to distinguish heteroduplexes from homoduplexes with perfect accuracy, and it is this property which has been utilized for SNP detection.[60]

Individuals which are heterozygotes for SNP will have 1:1 ratio of wild type and variant DNA, so that the PCR products will also have 1:1 ratio of wild type and variant sequence. In case of homozygous mutant, the DNA will have to be mixed with wild type PCR amplified DNA. In either case, the mixture of wild type and variant DNA is heated and cooled again, so that the sample will then have a mixture of homo- and heteroduplexes. The heteroduplexes partially denature due to single base pair mismatch and can then be distinguished from corresponding homoduplexes by ion-pair reversed phase liquid chromatography (Fig 16).

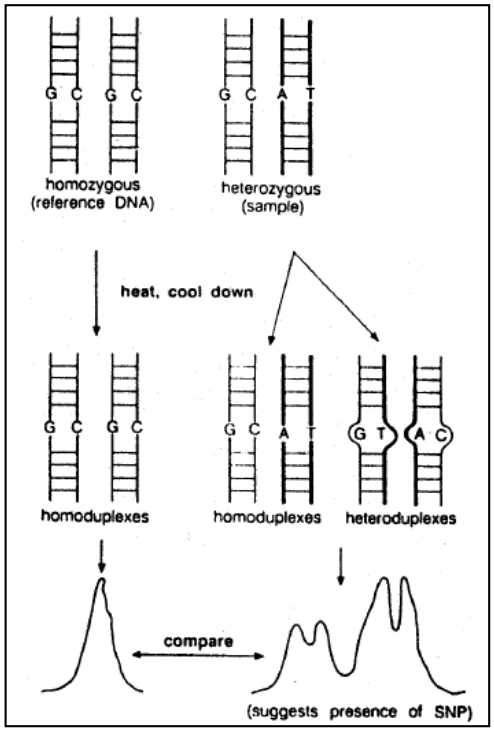


Fig 16. Diagrammatic representation of the principle involved in temperature modulated heteroduplex analysis using dHPLC.

Pyrosequencing for SNP genotyping: Pyrosequencing is actually a new sequencing method for obtaining sequences of short DNA segments (up to ~20 nucleotides). The method relies on step-wise addition of individual dNTP (with simultaneous release of pyrophosphates, i.e. PPi) and monitoring their template guided incorporation into the growing DNA chain via chemiluminescent detection of the formation of pyrophosphate.[77,78] Incorporation of a nucleotide into DNA will be possible, only if it is complementary to the next base in the template strand and the quantity incorporated will depend on the number of one or more consecutive complementary bases. Unincorporated dNTP is degraded using the enzyme apyrase. Pyrophosphate released is utilized to convert 5' amino phosphosulfurate (APS) into ATP with the help of the enzyme ATP sulfurylase, and the ATP produced drives luciferase mediated conversion of

luciferin into oxyleciferin, generating light. The light produced is proportionate to ATP produced, which in its turn will be directly proportionate to the dNTP consumed. The emitted light is detected by a CCD camera and seen in a program as a peak, whose height will tell us about the number of molecules of dNTP incorporated (Fig 17).

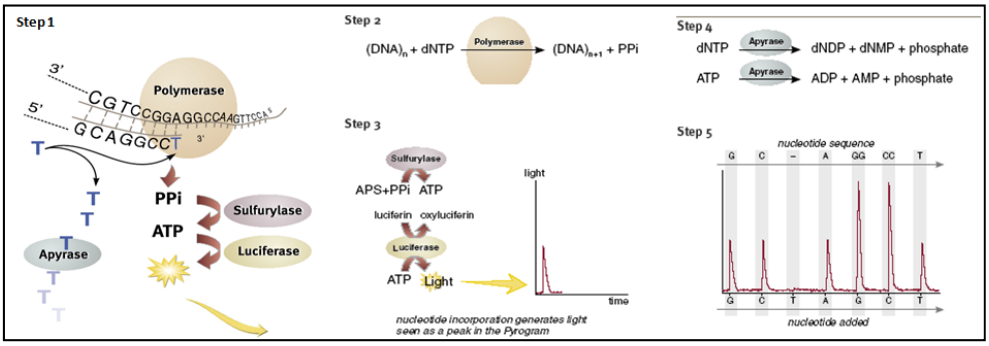


Fig 17. A schematic representation showing the principle of the Pyrosequencing method.

Invasive cleavage assay – A non-PCR assay: PCR does have significant limitations, when used in a high throughput approach, so that approaches involving simpler and more direct analysis of DNA without prior PCR amplification have been developed. One such approach employs 'invasive cleavage assay' for nucleic acids and a MALDI-TOF-MS detection system.[79] The invasive cleavage assay involves hybridization of genomic DNA with two sequence-specific oligonucleotides, one termed the invader oligonucleotide, and the other termed the probe oligonucleotide. The invader oligonucleotide has a sequence homology with a segment of genomic DNA upstream of the SNP site. The probe oligonucleotide, on the other hand, has a segment at its 3' end that is homologous to the target DNA, and another segment at its 5' end, that has no homology with the target DNA. On hybridization, a duplex is formed between the homologous segment of the probe oligonucleotide and the target DNA. The invader oligonucleotide now invades into the duplex for at least one nucleotide, thus forming an overlap at this point of invasion. The flap endonucleases cleave the unpaired region (including the overlap) available on the 5' end of the probe, resulting in a 3'-hydroxyl DNA cleavage product (Fig 18).

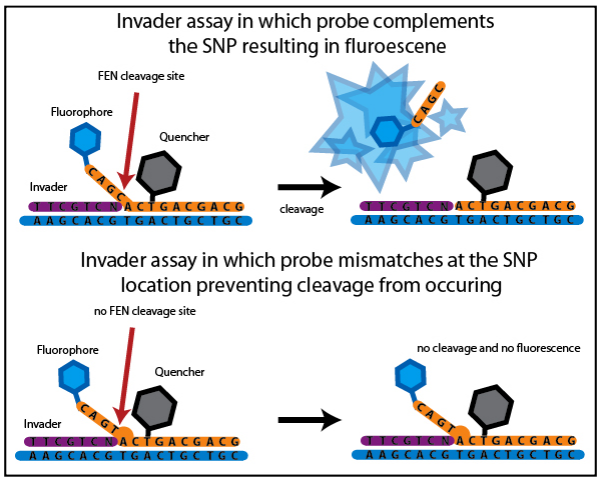


Fig 18. A schematic representation showing the principle of the Invader assay. Adapted from reference 80.

4. Conclusive remarks

Food authenticity is presently a subject of great concern to food authorities, as the incorrect labelling of foodstuff can represent a commercial fraud. Among many labelling declarations which claim 'quality' characteristic of a given food most concern varieties of vegetables used as ingredients. Their identity in processed or composite mixtures is not always readily apparent and verification that the components are authentic and from sources acceptable to the consumer may be required. In this scenario, DNA molecular markers have been proved to be indispensable tools for traceability purpose, due to the high stability of DNA compared with proteins, even in the hostile environments applied during many processing steps used for food production, and also to their presence in most biological tissue.[81] Nowadays, a variety of DNA-based methods is potentially available for use in food authentication, varying in their complexity and cost. Among these, SNPs have emerged as the new generation of molecular markers, which have already been developed in a large number of crop plant genome, not only for studies involving associations with a number of traits of economic value, but also for the study of genetic diversity and variety identification. Enormous genomic and cDNA sequence data that are accumulating in the databases will be extremely useful in future for discovery of new SNPs. A number of gel-based and non gel-based methods will also be used for detection of already characterized SNPs and for genotyping of populations at these SNP sites.[60] The genetic variation due to the alternation of single nucleotide allows easy comparisons between the results collected by different laboratories using different chemicals and technologies, without the needs of extensive controls. Finally, the availability of emergent technologies, such as microarrays, coupled to the development of new high performance probes, such as PNA, represents a high potential due, not only to their easiness, speed and specificity, but also to their miniaturised and automated nature, which makes them suitable for a great number of targets. Thus, these new technologies are promising tools for multiple species identification in complex matrices, answering to the constant needs for food fraud detection and quality assessment.

5. References

[1] Carcea M, Brereton P, Hsu R, Kelly S, Marmiroli N, Melini F, Soukoulis C, Wenping D. Food authenticity assessment: ensuring compliance with food legislation and traceability requirements. *Quality Assurance and Safety of Crops & Foods*, 2009, 93-100.

[2] Cheftel JC. Food and nutrition labelling in the European Union. *Food Chemistry*, 2005, **93**:531-550.

[3] http://eur-lex.europa.eu/LexUriServ.

[4] http://www.tpe.gov.tr/dosyalar/ABcografi/510_2006_Sayili_AB_Tuzugu.pdf.

[5] Dennis MJ. Recent developments in food authentication. *Analyst*, 1998, **123**:151–156.

[6] Terzi V, Morcia C, Giovanardi D, D'Egidio MG, Stanca AM, Faccioli P. DNA-based analysis for authenticity assessment of monovarietal pasta. *European Food Research and Technology,* 2004, **219**:428–431.

[7] Bligh HFJ. Detection of adulteration of Basmati rice with non-premium long-grain rice. *International Journal of Food Science and Technology*, 2000, **35**:257-265.

[8] Pafundo S, Agrimonti C, Marmiroli N. Traceability of plant contribution in olive oil by amplified fragment length polymorphisms. *Journal of Agricultural and Food Chemistry*, 2005, **53**:6995–7002.
[9] Palmieri L, Bozza E, Giongo L. Soft fruit traceability in food matrices using Real-Time PCR. *Nutrients*, 2009, **1**:316-328.
[10] Woolfe M, Primrose S. Food forensic: using DNA technology to combat misdescription and fraud. *Trends in Biotechnology*, 2004, **22**:222-226.
[11] Kumar P, Gupta VK, Misra A, Modi DR, Pandey BK. Potential of molecular markers in plant Biotechnology. *Plant Omics Journal*, 2009, **2**:141-162.
[12] Chikuni K, Ozutsumi K, Koishikawa T, Kato S. Species identification of cooked meats by DNA hybridization assay. *Meat Science*, 1990, **27**:119-128.
[13] Hunt DJ, Parkes HC, Lumley ID. Identification of the species of origin of raw and cooked meat products using oligonucleotide probes. *Food Chemistry,* 1997, **60**:437-442.
[14] Buntjer JA, Lamine A, Haagsma N, Lenstra JA. Species identification by oligonucleotide hybridisation: the influence of processing on meat products. *Journal of Food Science,* 1999, **79**:53-57.
[15] Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Bahkali AH, Al Sadoon M, Shobrak M. A brief review of molecular techniques to assess plant diversity. *International Journal of Molecular Sciences*, 2010, **11**:2079-2096.
[16] Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K. A system for rapid DNA sequencing with fluorescent chain terminating dideoxynucleotides. *Science*, 1987, **238**:336–341.
[17] Blow N. DNA sequencing: generation next-next. *Nature Methods*, 2008, **5**:267–274.
[18] Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences USA*, 2005, **102**:8369–8374.
[19] Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences USA*, 2009, **106**:18621–18626.
[20] Dayrat B. Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 2005, **85**:407-415.
[21] Unseld M, Beyermann B, Brandt P, Hiesel R. Identification of the species origin of highly processed meat products by mitochondrial DNA sequences. *PCR Methods and Applications*, 1995, **4**:241–243.
[22] Karp A, Isaac PG, Ingram GS. Molecular tools for screening biodiversity: plants and animals. *Chapman & Hall*, 1998.
[23] Schlotterer C, Tautz D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, 1992, **20**:211-215.

[24] Neale DB, Williams CG. Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. *Canadian Journal of Forest Research*, 1991, **21**:545–554.
[25] Miller JC, Tanksley SD. RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theoretical and Applied Genetics*, 1990, **80**:437–448.
[26] Fang DQ, Roose ML, Krueger RR, Federici CT. Fingerprinting trifoliate orange germplasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. *Theoretical and Applied Genetics,* 1997, **95**:211–219.
[27] Dubreuil P, Dufour P, Krejci E, Causse M, De Vienne D, Gallais A, Charcosset A. Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. *Crop Science*, 1996, **36**:790–799.
[28] Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, 1990, **18**:7213–7218.
[29] Charney J, Vershon D, Bill Sofer B. In: RAPD PCR. *Genes, Genomes and Human Genetics*, 2007. Available at
http://avery.rutgers.edu/WSSP/StudentScholars/project/archives/onions/rapd.html
[30] Foolad MR, Jones RA, Rodriguez RL. RAPD markers for constructing intraspecific tomato genetic maps. *Plant Cell Reports*, 1993, **12**:293-297.
[31] Abd El-Hady EAA, Haiba AAA, Abd El-Hamid NR, Rizkalla AAR. Phylogenetic diversity and relationships of some tomato varieties by electrophoretic protein and RAPD analysis. *Journal of American Science*, 2010, **6**:434-441.
[32] Matthes MC, Daly A, Edwards KJ. Amplified fragment length polymorphism (AFLP). In: Molecular Tools for Screening Biodiversity. *Chapman and Hall*, 1998.
[33] Romero G, Adeva C, Battad Z. Genetic fingerprinting: advancing the frontiers of crop biology research. *Philippine Science Letters*, 2009, **2**:8–13.
[34] Saliba-Colombani V, Causse M, Gervais L, Philouze J. Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome*, 2000, **43**:29–40.
[35] Thomas CM, Vos P, Zabeau M, Jones DA, Norcott KA, Chadwick BP, Jones JDG. Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato Cf-9 gene for resistance to Cladosporium fulvum. *Plant Journal*, 1995, **8**:785–794.
[36] Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. *Nature*, 1985, **314**:67–73.
[37] Rao R, Corrado G, Bianchi M, Di Mauro A. (GATA)4 DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. *Plant Breeding*, 2006, **125**:173-176.
[38] Kaemmer D, Weising K, Beyermann B, Borner T, Epplen J T, Kahl G. Oligonucleotide fingerprinting of tomato DNA. *Plant Breeding*, 1995, **114**-12-17.

[39] Chang SB, Yang TJ, Datema E, van Vugt J, Vosman B, Kuipers A, Meznikova M, Szinay D, Klein Lankhorst R, Jacobsen E, de Jong H. FISH mapping and molecular organization of the major repetitive sequences of tomato. *Chromosome Research*, 2008, **16**:919–933.

[40] Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science,* 1996, **1**:215–222.

[41] Allender C. Microsatellite Information Exchange for Brassica. 2004. Available at: http://www.brassica.info/resource/markers/ssr-exchange.php

[42] Bredemeijer G, Cooke R, Ganal M, Peeters R, Isaac P, Noordijk Y, Rendell S, Jackson J, Röder S, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B. Construction and testing of a microsatellite database containing more than 500 tomato varieties. *Theoretical and Applied Genetics*, 2002, **105**:1019-1026.

[43] Caramante M, Rao R, Monti LM, Corrado G. Discrimination of 'San Marzano' accessions: a comparison of minisatellite, CAPS and SSR markers in relation to morphological traits. *Scientia Horticulturae*, 2009, **120**:560-564.

[44] He C, Poysa V, Yu K. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among Lycopersicon esculentum cultivars. *Theoretical and Applied Genetics*, 2003, **106**:363-373.

[45] Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien MA. Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theoretical and Applied Genetics*, 2005, **110**:819-831.

[46] Ruiz JJ, García-Martínez S. Genetic variability and relationship of closely related Spanish traditional cultivars of tomato as detected by SRAP and SSR markers. *Journal of the American Society for Horticultural Science*, 2005, **130**:88-94.

[47] Caramante M, Corrado G, Monti LM, Rao R. Simple sequence repeats are able to trace tomato cultivars in tomato food chains. *Food Control*, 2011, **22**:549-554.

[48] Fulton TM, Xu Y, Siew FL, Tanksley SD. Efficiency of using CAPS as an alternative and potentially automatable mapping system. Report of the Tomato Genetics Cooperative, 1999, **49**:15–17,

[49] Ganal MW, Czihal R, Hannappel U, Kloos DU, Polley A, Ling HQ. Sequencing of cDNA clones from the genetic map of tomato (Lycopersicon esculentum). *Genome Research*, 1998, **8**:842–847.

[50] Pérez de Castro A, Blanca JM, Díez MJ, Nuez Viñals F. Identification of a CAPS marker tightly linked to the tomato yellow leaf curl disease resistance gene Ty-1 in tomato. *Biomedical and Life Sciences European Journal of Plant Pathology*, 2007, **117**:347-356.

[51] Doganlar S, Dodson J, Gabor B, Beck-Bunn T, Crossman C, Tanksley SD. Molecular mapping of the py-1 gene for resistance to corky root rot (Pyrenochaeta lycopersici) in tomato. *Theoretical and Applied Genetics*, 1998, **97**:784-788

[52] Frary A, Xu Y, Liu J, Mitchell S, Tedeschi E, Tanksley SD. Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their

usefulness for genetics and breeding experiments. *Theoretical and Applied Genetics*, 2005, **111**:291–312.

[53] Martin GB, Williams JGK, Tanksley SD. Rapid identification of markers linked to a Pseudomonas resistance gene in tomato by using random primers and near-isogenic lines. *Proceedings of the National Academy of Sciences USA*, 1991, **88**:2336–2340.

[54] Zhang Y, Stommel JR. Development of SCAR and CAPS markers linked to the Beta gene in tomato. *Crop Science*, 2001, **41**:1602–1608.

[55] Ohmori T, Murata M, Motoyoshi F. Molecular characterization of RAPD and SCAR markers linked to the Tm-1 locus in tomato. *Theoretical and Applied Genetics*, 1996, **92**:151–156.

[56] Dianese EC, de Fonseca MEN, Goldbach R, Kormelink R, Inoue-Nagata AK, Resende RO, Boiteux LS. Development of a locus-specific, co-dominant SCAR marker for assisted-selection of the Sw-5 (Tospovirus resistance) gene cluster in a wide range of tomato accessions. *Molecular Breeding*, 2010 **25**:133-142.

[57] Wolters P, Powell W, Lagudah E, Snape J, Henderson K. In: Plant and Animal Genome VIII Conference, 2000. Available at http://www.intl-pag.org/pag/8/ Abstracts/

[58] Bhattramakki D, Ching A, Morgante M, Dolan M, Register J, Smith H, Tingey S, Rafalski JA. In: Plant and Animal Genome VIII Conference, 2000. Available at http://www.intl-pag.org/pag/8/ abstracts/

[59] van Deynze AE, Stoffel K, Buell RC, Kozik A, Liu J, van der Knaap E, Francis D. Diversity in conserved genes in tomato. *BMC Genomics*, 2007, **81**:465-472.

[60] Gupta PK, Roy JK, Prasad M. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science*, 2001, **80**:524-535.

[61] Taillon-Miller P, Gu Z, Li Q, Hillier L, Kwok PY. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Research*, 1998, **8**:748–754.

[62] Orita M, Suzuki Y, Sekiya T, Hayashi K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*, 1989, **5**:874–879.

[63] See D, Kanazin V, Talbert H, Blake T. Primer mediated detection of single nucleotide polymorphisms. *Barley Newsletter*, 1998. Available at http://wheat.pw.usda.gov/ggpages/BarleyNewsletter/42/post26.html

[64] Ingelsson M, Shin Y, Irizarry MC, Hyman BT, Lilius L, Forsell C, Graff C. Genotyping of apolipoprotein E: comparative evaluation of different protocols. *Current Protocols in Human Genetics,* 2003, **9**:914-913.

[65] Lee LG, Livak KJ, Mullah B, Graham RJ, Vinayak RS, Woudenberg TM. Seven-color, homogeneous detection of six PCR products. *Biotechniques*, 1999, **2**:342–349.

[66] Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology,* 1996, **14**:303–308.

[67] Tyagi S, Bratu DP, Kramer FR. Multicolor molecular beacons for allele discrimination. *Nature Biotechnology*, 1998, **16**:49–54.

[68] Landergren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. *Science*, 1988, **244**:1077–1080.
[69] Lemieux B, Aharoni A, Schena M. Overview of DNA chip technology. *Molecular Breeding*, 1998, **4**:277–289.
[70] Sforza S, Corradini R, Tedeschi T, Marchelli R. Food analysis and food authentication by peptide nucleic acid (PNA)-based technologies. *Chemical Society Reviews*, 2011, **40**:221–232.
[71] Ganesh KN, Nielsen PE. Peptide nucleic acids analogs and derivatives. *Current Organic Chemistry*, 2000, **4**:931–943.
[72] Sforza S, Corradini R, Ghirardi S, Dossena A, Marchelli R. DNA binding of a D-lysine-based chiral PNA: direction control and mismatch recognition. *European Journal of Organic Chemistry*, 2000, **2000**:2905–2913.
[73] Menchise V, De Simone G, Tedeschi T, Corradini R, Sforza S, Marchelli R, Capasso D, Saviano M, Pedone C. Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA–DNA duplex. *Proceedings of the National Academy of Sciences USA*, 2003, **100**:12021–12026.
[74] Corradini R, Feriotto G, Sforza S, Marchelli R, Gambari R. Enhanced recognition of cystic fibrosis W1282X DNA point mutation by chiral peptide nucleic acid probes by a surface plasmon resonance biosensor. *Journal of Molecular Recognition*, 2004, **17**:76–84.
[75] Chiari M, Galaverna G, Sforza S, Cretich M, Corradini R, Marchelli R. Detection of the R553X DNA single point mutation related to cystic fibrosis by a "chiral box" D-lysine-peptide nucleic acid probe by capillary electrophoresis. *Electrophoresis*, 2005, **26**:4310–4316.
[76] Howell WM, Jobs M, Gyllensten U, Brookes AJ. Dynamic allele-specific hybridization. *Nature Biotechnology*, 1999, **17**:87–88.
[77] Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science*, 1998, **281**:363–365.
[78] Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, Lundeberg J. Single-nucleotide polymorphism analysis by pyrosequencing. *Analytical Biochemistry*, 2000, **280**:103–110.
[79] Griffin TJ, Hall JG, Prudent JR, Smith LM. Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Proceedings of the National Academy of Sciences USA*, 1999, **96**:6301–6306.
[80] Olivier M. The Invader® assay for SNP genotyping. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 2005, **573**:103-110.
[81] Mafra I, Ferreira I, Beatriz M, Oliveira PP. Food authentication by PCR-based methods. *European Food Research and Technology*, 2008, **227**:649–665.

Chapter VI

Advanced genomic tools for tomato genotyping: evaluation of 'chiral box' PNA affinity and selectivity in solution and on microarrays

1. Introduction

In the field of the molecular probes concerning food analysis, Peptide Nucleic Acids (PNAs) can be considered very promising tools in DNA detection, due to their unique properties.

PNAs are synthetic DNA analogues in which the phosphodiester backbone is replaced by a flexible pseudopeptide, made of repetitive units of N-(2-aminoethyl) glycine to which the purine and pyrimidine bases are attached via a methyl carbonyl linker.[1] Since in the PNA structure the backbone length and the distance of the nucleobases from the backbone are the same as in natural DNA, PNAs can bind complementary DNA or RNA sequences, following standard Watson-Crick rules.[2]

Unlike DNA and RNA, the PNA backbone is not charged. Consequently, there is no electrostatic repulsion when PNA hybridizes to its target nucleic acid sequences, giving a higher stability to the PNA-DNA or PNA-RNA duplexes than the natural homo- and heteroduplexes. This great stability also results in a higher thermal melting temperature (T_m) values as compared to DNA-DNA and DNA-RNA duplexes.[3]

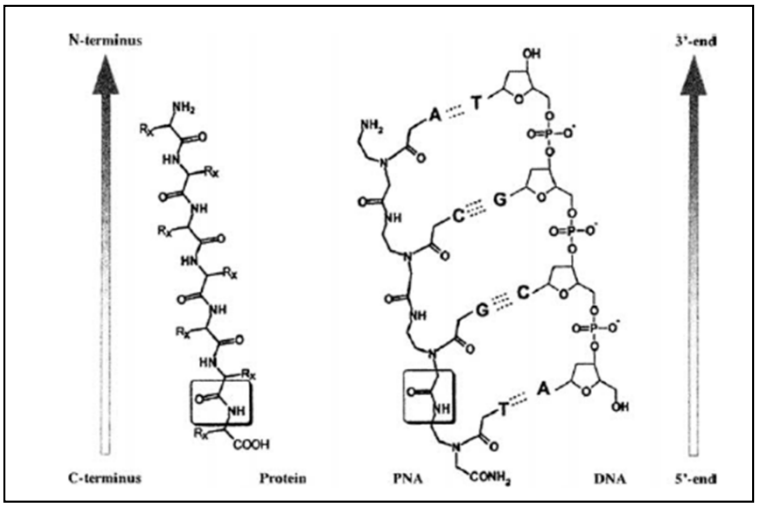


Fig 1. Chemical structures of PNA as compared to DNA and protein. The backbone of PNA displays 2-aminoethyl glycine linkages in place of the regular phosphodiester backbone of DNA, and the nucleotides bases are attached to this backbone at the amino-nitrogens through methylene carbonyl linkages. The amide bond characteristic for both PNA and protein is boxed in. By convention, PNAs are depicted like peptides, with the N-terminus at the left (or at the top) position and the c-terminus at the right (or at the bottom) position. PNAs hybridize to complementary DNA or RNA sequences in a sequence-dependant manner, following the Watson–Crick hydrogen bonding scheme. PNAs can bind to complementary nucleic acids in both parallel and anti-parallel orientation, which is illustrated in this figure. The Watson–Crick hydrogen bonds are indicated by "…". Adapted from reference 4.

An additional consequence of the neutral polyamide backbone is that PNAs hybridize independently of the presence of salts, which are necessary to attenuate electrostatic repulsions in DNA duplex. Indeed, PNA–DNA binding can be efficiently achieved even under very low salt concentrations, a condition that promotes the destabilization of RNA and DNA secondary structures, improving access to target sequences and facilitating the hybridization with the PNAs.[5] The unnatural backbone of PNAs also means that PNAs are particularly resistant to protease and nuclease degradation.[6] Because of this

resistance to the enzymatic degradation, the lifetime of PNAs is extended in biological fluids, both in vivo and in vitro.

As already said, PNAs hybridize to complementary DNA or RNA in a sequence dependant manner but, in contrast to DNA, PNA can bind in either parallel (N-terminus of *PNA* facing the 5'-end of DNA) or anti-parallel (N-terminus of *PNA* facing the 3'-end of DNA) fashion, both adducts being formed at room temperature, with the anti-parallel orientation showing higher stability.[7] These data indicate that PNA backbone is more flexible than native nucleic acid backbone. Actually, the ability of PNAs to strongly bind DNA has been interpreted as a result of the "constrained flexibility" in its structure, since too flexible, as well as too rigid, analogs showed very poor binding properties.

Also, PNA-DNA hybridization is significantly more affected by base mismatches than DNA-DNA hybridization. It has been shown that a single mismatch in a mixed PNA–DNA 15-mer duplex decreases the T_m by up to 15 °C, whereas in the corresponding DNA–DNA complex, a single mismatch decreases the T_m by only 11 °C.[8] Due to this selectivity and specificity which allow a high level of discrimination at single base level, PNA probes represent a reliable tool for an easy detection of single mutations and single nucleotide polymorphisms.[9]

Introduction of different functional groups with different charges/polarity/flexibility has been described, giving rise to modified PNAs. In particular, the role of chirality in these compounds has been deeply investigated by using modified PNAs with stereogenic center either in the C2 or in the C5 carbon position of the backbone or in both, by inserting aminoacid-derived side chains (Fig 2).[10-12]

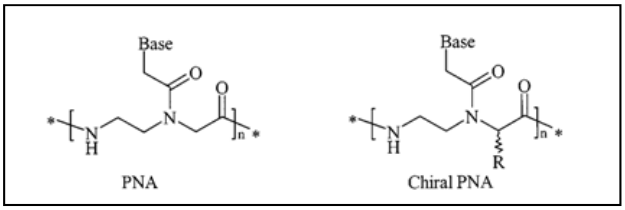


Fig 2. Achiral PNA (left) and chiral PNA substituted at C2 carbon (right). The latter is obtained by inserting a monomer based on a chiral amino acid. Adapted from reference 13.

The main effect of the addition of substituents at C2 or C5 carbon is to shift the PNA preference towards a right-handed or left-handed helical conformation, according to the configuration of the new stereogenic centers, in turn affecting the stability of PNA-DNA duplex through a control of the helix handedness. In particular, when using PNAs with an amino acid derived stereogenic centre at the C2 position, the intrastrand steric hindrance of the amino acid chain influences the PNA helical structure, which thus imparts a preference for right-handed (when D-amino acids are used) or a left-handed (for L-amino acids) helicity. On the contrary, the use of C5-modified monomers derived from L-amino acids were shown to induce a preferential right-handedness.[14] As DNA is a right-handed helix (in the common B-form), it preferentially bound to 2D- or 5L-PNAs, which showed the same preferred handedness. 2L- or 5D-PNAs, which favour the "wrong" left-handed helix, were found to bind to DNA more weakly as were forced to assume an unfavourable conformation. As a result, C2 and C5-modified PNAs were found to bind very effectively to DNA when the stereochemistry was 2D, 5L.[12] Chiral PNAs were generally found to form slightly less stable PNA-DNA duplexes than their achiral analogues,[15] the effect being more pronounced for backbone containing negatively

charged monomers (derived from aspartic and glutamic acid), probably because of their repulsive interactions with the DNA negatively charged phosphate group.[16] On the other side, the introduction of positively charged monomers (derived from lysine,[13] and arginine,[17]) improved the stability of PNA-DNA duplexes. It has also been noted that the stereogenic centre was more efficient in effecting selectivity when it was positioned in the middle of the PNA strand.[18]

Since high sequence selectivity is one of the properties for which PNAs are considered excellent probes, in order to achieve a specific recognition of complementary antiparallel DNA sequences and to improve the selectivity in the recognition of single base mismatch or single point mutation, a series of three adjacent C2-modified chiral monomers, termed "chiral box", were introduced in the middle of the PNA backbone, resulting in PNA probes showing very high direction control (antiparallel vs. parallel binding) and exceptional sequence selectivity (Fig 3).[13,17]

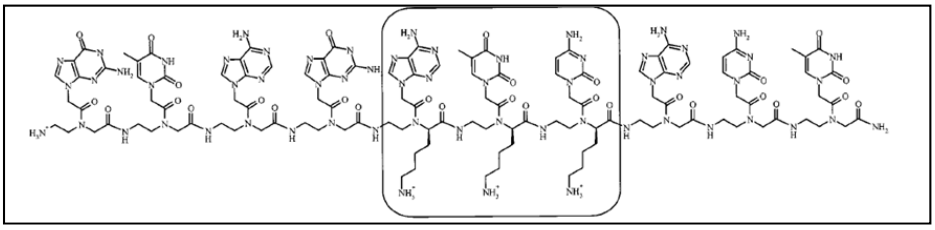


Fig 3. Example of chiral box PNA, incorporating three chiral adjacent monomers of C2-D-lysine in the middle of the strand. Adapted from reference 13.

Therefore, it appears from the literature data that PNAs can usefully replace DNA probes in the revelation of specific gene sequences, being exploitable both on surface and in solution technologies, as already demonstrated in the detection of DNA from GMOs,[19] and from hidden allergens,[20,21] or in the revelation of SNPs in the human genome.[17]

2. *Aim of the work*

In this chapter several PNA structures were tested in order to value which one could provide the best performance in terms of specificity, in binding complementary DNA sequences, and selectivity, in mismatch recognition, both in solution and on surface. For this aim, a PNA probe, designed to recognize a specific SNP placed along tomato genome, was first synthesized, both in the standard format, fully achiral, and as two different chirally modified variants, bearing a "chiral box" composed of three C2- or three C5-modified arginine monomers in the middle of the sequence. Their binding affinity and mismatch recognition were tested both in solution and in microarray technology, using synthetic DNA oligomers mimicking the sequence containing the mutation.

3. *Experimental Part*

3.1 PNA design and synthesis

Complementary reverse sequences of DNA strand to be detected were obtained using on line available FastPCR v 6.0 software.[22] For the DNA sequence to be recognized, several PNA probes were tested by on line PNA designer support,[23] designed by keeping the mismatch nucleotides in the central position of the strand and extending the ends of the

probes to reduce the difference in their T_m, by balancing the purine/pirimidine ratio. Afterwards each of them was first checked to minimize any secondary structure due to self- or hetero-annealing, in order to avoid a loss of hybridization efficiency, by using the on line program IDT – Integrated DNA Technologies.[24] The sequence specificity was then evaluated for each probe using NCBI-BLAST homology tool, to avoid any possible hybridization on other non-target regions. PNA probes have been obtained by the laboratories of the Department of Organic and Industrial Chemistry, after having being synthesized by solid phase methodologies.

3.2 PNA spotting on microarray slides

3.2.1 Chemicals

- Sodium carbonate (Na_2CO_3) (Carlo Erba, Italy)
- Sodium bicarbonate ($NaHCO_3$) (Carlo Erba, Italy)
- Carbonate buffer 100 mM (pH 9)
- Sodium dodecyl Sulphate SDS (0.1% and 10% solutions in Milli Q H_2O) (Sigma-Aldrich, USA)
- Printing solution:
  - PNA probes (30 and 50 μM)
  - 100 mM Carbonate buffer
  - 0.001% SDS
- Ammonia (NH_3) (1% aqueous solution) (Carlo Erba, Italy)
- Milli Q H_2O obtained with Millipore Alpha Q system
- Control probe ($(AC)_{11}$-Cy5)) (Thermo Scientific, Italy)

3.2.2 Instrumentation

- Genorama® SAL Enhanced Slides (25 x 75 x 1 mm) (Asper Biotech Ltd, Estonia)
- SpotArray™ 24 Microarray Printing System (PerkinElmer™ Life Sciences, Inc, USA)
- Express™ Microarray Scanner (PerkinElmer™ Life Sciences, Inc, USA)
- Reciprocating shaker SSL3 (Stuart Scientific, UK)
- ALC PK110 Centrifuge (Thermo Electron, USA)

3.2.3 Procedure

The manufacturer's instructions for the deposition protocol were slightly changed to comply with the special requirement of the chemical structures of PNAs: in particular a 100 mM carbonate buffer (pH 9.0) containing 0.001% SDS was used as printing solution for 30 or 50 μM of PNA probe. Moreover, after every deposition, the pin printing system was washed in three solutions containing 100 mM carbonate buffer (pH 9.0) and SDS at different concentrations (0.2%, 1%, 0.001%), to avoid carry-over of the probes in subsequent depositions. The probes were coupled to the surface by leaving the slides in a humid chamber (relative humidity 75%) at room temperature for 12 hours, and the remaining reactive sites were blocked by immersion in a 1% aqueous solution of NH_3 at room temperature for 15' under gently shaking. The slides were slowly shaken for 15' with 0.1% SDS solution pre-warmed at 40 °C and then washed twice with Milli Q H_2O for 5' each at room temperature. Each slide was spin-dried twice in a plastic tube at 1200 r.p.m. for 5', to remove the remaining washing solution. The efficiency of the deposition

step was then checked by acquiring slides at λ_{ex}=633 nm and λ_{em}=670 nm and displaying the fluorescent control probe ($(AC)_{11}$-Cy5). The slides were then ready to undergo the hybridization protocol or could be stored under dried atmosphere for future use. All the previously described operations were carried out away from direct light in order to prevent degradation of the Cy5 fluorophore.

3.3 Hybridization on microarray

3.3.1 Chemicals

- Sodium chloride (NaCl) (AnalaR Normapur®, Italy)
- Sodium citrate ($C_6H_7NaO_7$) (Carlo Erba, Italy)
- Chloridric acid (HCl) (Sigma-Aldrich, USA)
- Saline-Sodium Citrate buffer (SSC 20 x):
  - 17.5 g NaCl (3 M)
  - 10.3 g $C_6H_7NaO_7$ (0.35 M)
  - Adjust the pH to 7.0 with a solution of HCl.
  - Add Milli Q H_2O to 100 mL
- Cy5-labeled DNA oligonucleotides full match (FM) and mismatch (MM) (Thermo Scientific, Italy)
- Sodium dodecyl Sulphate SDS (10% solution) (Sigma-Aldrich, USA)
- Hybridization solution:
  - 0.1 µM DNA Oligonucleotides
  - 2 x SSC buffer
  - 0.1% SDS solution

3.3.2 Instrumentation

- Array hybridation cassette 1 x 16 (Arrayit Corporation, USA)
- Gasket array hybridation cassette 1 x16 (Arrayit Corporation, USA)
- Reciprocating shaker SSL3 (Stuart Scientific, UK)
- ALC PK110 Centrifuge (Thermo Electron, USA)
- Express™ Microarray Scanner (PerkinElmer™ Life Sciences, Inc, USA)

3.3.3 Procedure

The Cy5-labeled DNA-FM and DNA-MM oligonucleotides were used at a final concentration of 0.1 µM in 2 x saline-sodium citrate (SSC) solution and 0.1% SDS buffer. Hybridization was performed by loading 65 µL of each hybridization solution to array hybridization cassett, using a silicone gasket to avoid cross contamination, and leaving the slides under slow shaking for 2.5 hours at 40 °C. After the hybridization step, the slides were washed under slow shaking for 5' at 40 °C with a 2 x SSC, 0.1% SDS buffer pre-warmed at 40 °C, followed by treatment for 1' with 0.2 x SSC and for 1' with 0.1 x SSC at room temperature. The slides were then spin-dried twice at 1200 r.p.m. for 5'. All post-hybridization steps were performed in a dark environment to prevent degradation of the Cy5 fluorophore used to label the oligonucleotides. The fluorescent signal deriving from the hybridization was acquired using an Express™ Microarray Scanner at λ_{ex}=633 nm and λ_{em}=670 nm. To correctly compare the hybridization data, all the images reported were acquired with laser power set at 100 and photomultiplier gain set at 70.

Images were analyzed using the ScanArray program with integration of 170 μm diameter circular area entirely containing the fluorescent spots.

4. Results and Discussion

4.1 Trial PNA design and synthesis

In order to elucidate which PNA structure could be the more appropriate in terms of DNA binding and mismatch recognition, both in solution and on surface, a trial experiment was carried out with a PNA probe designed on the tomato sequence LeOH 31.3 (see the following chapter for further details), containing the mismatch G/A, as illustrated in Fig 4a.

The trial PNA probe was first designed by using PNA-T_m calculator software,[23] as shown in Fig 4b.

a

LEOH 31.3

5′-AAGCTCTCCTTTCACTGAACTTCCATTGCAATGGCTTCTCTCCTCAACACTGTGCCATCT
ATTAAACTATCAAACTTCAACTACAACAACCCACTTCGCTCTTCACAAATTTCATTCTCC
CTCTCTCGAAGAAGACTCGTTGTTAGAGCAACAGAGACTGAAAAAGAAGCTAAAGCAGAG
GCACC**G/A**GATAAGGCACCAGCTGCTGGTGGCTCAAGTTTAAATCAGATTCTTGGAATC
AAAGGAGCCAAGCAAGAAACGGACAAGTGGAAGATTCGGGTTCAGCTTACAAAACCTGTT
ACTTGGCCTCCCCTTATCTGGGGTGTGGTCTGTGGAGCTGCTGCTTCTGGGAACTTCCAC
TGGACTCCAGAGGATGTGGCCAAATCAATTGTTTGTATGTTGATGTCTGGTCCATTTCTA
ACTGGCTATACTCAGACTATTAATGATTGGTATGATAGAGAAATCGATGCTATAAACGAA
CCTTACCGTCCAATTCCTTCAGGGGCGATAGCTGAACAAGAGGTCATTACTCAAATATGG
GTGCTTCTTTTAGGAGGCCTTGGGTTAGCTGGTATTTTAGATGTTTGGGCAGGGGCATGA
CTTTCTGGTATATTTTACCTTGCACTTGGTGGGATCCTGCTCTCCTACATCTACTCAGCT
CCCCACATTAAGCTCCAACAGAACGGA-3′

b

PNA Sequence (H-NH_2)	GTGCCTTATCCGG	GCCTTATCCGGTG	CTTATCCGGTGCC	TATCCGGTGCCTC	TCC GGTGCCTCTG
T_m c= [1 μM]	[1 μM] = 67 °C	[1 μM] = 67 °C	[1 μM] = 65 °C	[1 μM] = 67 °C	[1 μM] = 71 °C
Molecular Weight	3527.9	3527.9	3487.9	3487.9	3503.9
Length	Residues: 13	Residues: 13	Residues: 13	Residues: 13	Residues: 13
Purine %	38.5%	38.5%	30.8%	30.8%	30.8%
Polymorphic SNP	GTGCCTTATC**C**GG	GCCTTATC**C**GGTG	CTTATC**C**GGTGCC	TATC**C**GGTGCCTC	TC**C**GGTGCCTCTG
Complementarity	High	High	High	High	High

Fig 4. a, genomic sequence of locus LeOH 31.3, with the polymorphic bases in bold red; **b**, results from calculator software of the possible PNA probes, the underlined nucleotide is the one implying in the SNP recognition.

Five different PNA probes, all recognizing the polymorphic base 'G' of tomato LeOH 31.3 sequence, were used for T_m calculation: they were designed by shifting 2 and 4 bases toward 5' end, first, and toward 3' end, then, in order to have the nucleotide to be recognize along all the sequence. All the tested PNA probes had a percentage of purine content below the 60%, since purine-rich PNA oligomers tend to aggregate and have low solubility in aqueous solutions. Since these types of probes are prone to interact, as PNA/PNA interactions are even stronger than PNA/DNA interactions, self-complementary sequences with inverse repeats, hairpins and palindromes should be avoided. Unfortunately, all the tested probes showed a high self-complementary, due to the motif 'CCGG' containing the polymorphic nucleotide and, for its own geometry, the self-complementarity of the probes could not be avoided. Moreover, the lowest T_m belonged to the PNA probe with the polymorphic SNP placed in the central position of the strand. Since previous studies showed that the presence of 'chiral box' in the middle of the strand overcomes the lack of binding selectivity and displayed very good single-

point mismatch recognition,[13,17] finally a PNA probe containing the polymorphic nucleotide in the central position of the strand was chosen to be synthesized.

All the PNA probes used in this work were synthesised at the Department of Organic and Industrial Chemistry at University of Parma using different strategies.

The synthesis of the achiral PNA was performed using Fmoc-based chemistry and standard protocols, which take the name from the N-terminal amine protecting group (9-fluorenylmethoxycarbonyl, Fmoc). PNA synthesis proceeded from C-terminal to N-terminal ends and it started from the first monomer loaded on the resin by a peptide bound formation between carboxylic acid group of the monomer and amino group of the resin. The resin used in this experiment was the Rink-Amide, a polystyrene resin with a linker which is protected on the N-terminal amine by Fmoc group. A solution containing piperidine/NMP (N-Methyl-2-pyrrolidone) allowed the removal of the protecting group Fmoc from the resin and the insertion of the first monomer.

A coupling agent, named HBTU/DIPEA (2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium Hexafluoro-phosphate/DiIsopropylEthylAmine), helped the formation of the peptide bound, by the production of an active ester site in situ. After the first monomer has been introduced, protocol required the introduction of subsequent monomers (Fig 5), as follows:

- ✓ removal of protecting group Fmoc from the last monomer loaded on the resin;
- ✓ coupling reaction with the following monomer;
- ✓ capping reaction with acetic anhydride in order to block all the amino groups which did not react

Each step was preceded by some washes with dichloromethane and N- methylpyrrolidone in order to clean the resin off from impurities and excess reagents, increasing the yield. In this case, PNA was synthesized by an automatic ABI 433A Synthesizer, except for the first monomer loading step on the resin, which was hand-made. At the end of the sequence, two 2-(2-aminoetoxy)etoxyacetyl spacers were inserted, to anchor PNA to the solid surface, keeping it at the right distance. As the synthesis was terminated, PNA was cleaved from the resin with a trifluoroacetic acid:m-cresol (9:1) solution. Trifluoroacetic acid is a strong acid which breaks the bound between PNA and the resin. After being treating with trifluoroacetic acid, PNA was precipitated with ethyl alcohol and, afterwards, it was air-dried for further characterizations.

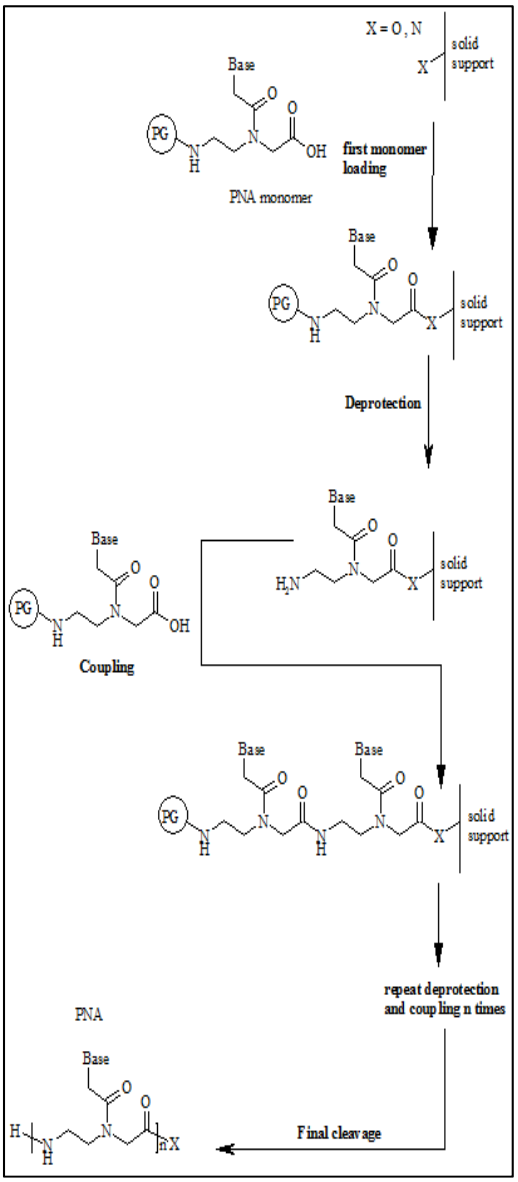


Fig 5. Scheme of PNA synthesis on solid solid support. Adapted from reference 25.

The synthesis of 2D-chiral-box PNA was carried out by using the submonomeric synthesis previously described for other chiral-box.[13] Briefly, the synthesis was carried out on an MBHA (4-Methylbenzhydrylamine) resin; the chiral monomers were directly assembled onto the resin into the middle of the PNA strand by coupling the protected backbone and then, after deprotection, inserting on the backbone the nucleobase, always on resin. The final cleavage was carried out with trifluoromethanesulfonic acid (TFMSA). Unlike the already reported method, D-lysine monomers were substituted by D-argine monomers.

The 5L-chiral box PNA was synthesized by a Boc (butyloxycarbonyl)-based solid phase peptide synthesis using a polystyrene resin bearing MBHA group derivatized with Boc-L-arginine. The coupling reaction was carried out with two preformed chiral monomers,[26] and commercially available Boc-PNA monomers. Boc-deprotection was performed with trifluoroacetic acid and final cleavage was carried out in TFMSA.

All crude PNAs were purified by RP-HPLC with UV detection at 260 nm using a semipreparative column C18 and the resulting pure PNA oligomers were characterized by ESI-MS-RP-HPLC.[26]

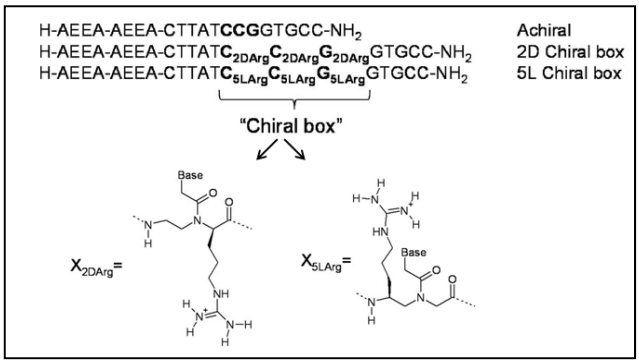


Fig 6. Achiral PNA, 2D-chiral-box PNA, 5L-chiral box PNA, and sequences used in this study. AEEA refers to the 2-(2-aminoetoxy) etoxyacetyl spacer. Adapted from reference 26.

4.2 PNA recognition and binding properties in solution

After synthesis and quantification by UV absorbance have been carried out,[26] the PNA probes complexation properties in solution were investigated by determining the T_m, in order to evaluate the PNA ability in binding a DNA complementary sequence and in recognizing a mismatch. The determination of T_m, indeed, allowed to assess PNA:DNA interaction by recording the absorbance in a temperature range between 18-90 °C and subsequent cooling down. The experiments were performed using synthetic DNA oligonucleotides having full complementary (FM) or a single-mismatch (MM) sequence, corresponding to the polymorphic SNPs observed at locus LeOH 31.3 (Fig 4a and Fig 7).

Full-match DNA-FM: 5'-GGCACC**G**GATAAG-3'
Mismatch DNA-MM: 5'-GGCACC**A**GATAAG-3'

Fig 7. DNA full matched (DNA-FM) and mismatched (DNA-MM) sequences used in this study.

The UV melting curves of the two chiral PNA probes are reported in Fig 8, and the corresponding melting temperatures are reported in Table 1 and compared with those of the achiral PNA.

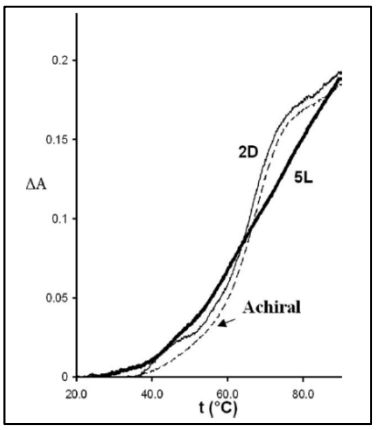


Fig 8. UV melting curves (260 nm) for the DNA-FM with 2D-chiral box PNA (thin line), 5L-chiral box PNA (thick line) and achiral PNA (dashed line) performed in PBS buffer, pH 7; concentration of each strand was 5 μm. Data are expressed as variation of absorbance (ΔA). Adapted from reference 26.

Concerning the binding affinity, the 2D-chiral box PNA-DNA-FM duplex was found to have similar melting temperature as compared to the achiral PNA:DNA-FM, similarly to what observed previously for lysine-based 2D-chiral box PNAs with different sequences,[13] indicating the formation of highly stable duplexes. On the other hand, the melting temperature for 5L-chiral box PNA was significantly higher than in the achiral PNA:DNA and 2D-chiral box PNA-DNA duplexes, thus confirming that this modification is the most suited for obtaining PNA with high affinity for complementary DNA. Besides, the huge steric hindrance, due to the three adjacent arginine monomers in the chiral box, seemed not to interfere in the PNA:DNA-FM binding.[13]

Table 1. UV melting temperature of the duplexes of PNA probes with full match and mismatch DNA. Adapted from reference 26.

PNA	DNA	UV T_m [a] (°C)	Δ T_m [b] (°C)
Achiral PNA	Fullmatch	67	22
	Mismatch	45	
2D-chiral box PNA	Fullmatch	66	26
	Mismatch	40	
5L-chiral box PNA	Fullmatch	75	20
	Mismatch	55	

[a] PNA:DNA melting temperature measured in PBS buffer, pH 7, concentration of 5 μm for each strand
[b] T_m (Fullmatch) – T_m (Mismatch).

The recognition of a single mismatch (A instead of G) was also evaluated by the decrease in the melting temperatures. In this case, the 2D-chiral box PNA-DNA-MS was found to have the lowest melting temperature, with a ΔT_m of 26 °C, greater than that observed for the achiral PNA (22 °C), whereas the 5L-chiral box PNA gave a stable duplex also in the presence of the mismatch, with ΔT_m of only 20 °C, comparable to that of achiral PNA.
The higher recognition properties of the 2D-chiral box model could be due to the different position of the substituent in the PNA backbone, which is attached to a carbon

atom between two rigid amidic groups for the 2D-derivative, whereas in the case of 5L-derivative, the substituent is located on the aminoethyl moiety, which allows a higher conformational freedom (Fig 6). The presence of a mismatch in the central base of the 2D-chiral box could induce a conformational change which then affects the proper conformation of the monomer and of adjacent residues. In this way the overall conformation of the "chiral box" segment would be highly distorted and would give rise to a less stable PNA:DNA duplex. This effect is less pronounced in the case of 5L monomers, due to the possibility for the flanking monomers to better adapt to the distorted conformation, therefore preserving the positive electrostatic interactions which stabilize the duplex.[26] Therefore, while the 5L-chiral box PNA was superior in terms of binding, the 2D-chiral box was found to be the best model for the recognition of single mismatch in solution.

4.3 PNA recognition and binding properties on microarrays

PNA probes recognition properties on solid surface were investigated using the microarray technology.

The surface of glass slides were coated with aminosilane (3-aminopropyltrimethoxysilane + 1,4-phenylenediisothiocyanate). The reaction between the amino ends of PNAs and the surface implies the formation of a very stable thiourea bond, which ensures the anchoring of PNAs to the slide (Fig 9).

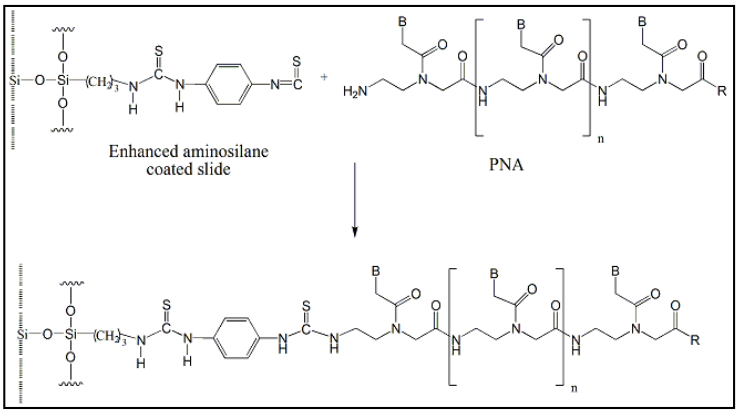


Fig 9. Exemplificative mechanism of PNA deposition on slide array.

PNAs were spotted using solutions of two different concentrations (30 and 50 μM), both producing similar results (Fig. 10) suggesting that the concentration was sufficient in both cases to saturate the active sites of the slide. A set of the achiral, 2D-chiral box, and the 5L-chiral box PNA was spotted in duplicate on the same slide, together with a Cy5-labeled oligonucleotide as a control of deposition. Deposition steps were alternated with SDS and H_2O washes, in order to clean the pin of the array spotter and avoid 'carry-over' contaminations.

Using a multichamber silicone gasket, one set of spots was then hybridized with the full-match oligonucleotide, and the other with the mismatched one, each bearing Cy5-labeling, as described in paragraph 3.3, and using a temperature of hybridization of 40 °C in order to discourage the formation of secondary structure or self- and hetero-dimer among the oligonucleotides, which could decrease hybridization efficiency.

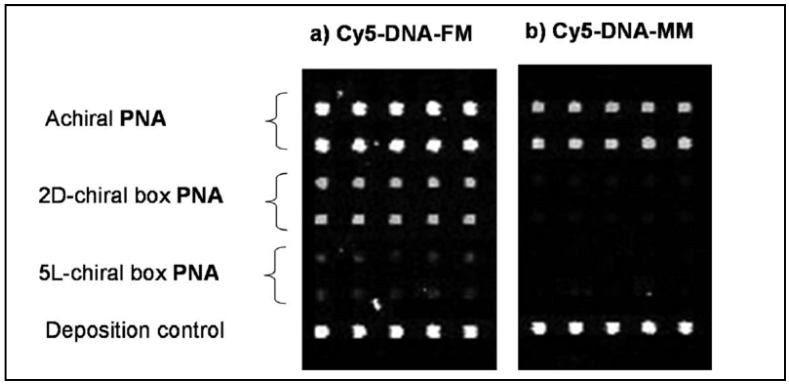


Fig 10. Microarray analysis of Cy-5 labeled DNA oligonucleotides: **a**, full-match DNA-FM and **b**, single-mismatch DNA-MM sequences. Hybridization was performed as described in Experimental part. Adapted from reference 26.

As shown in Fig 10a the achiral PNA gave rise to the more intense signal with the full-match DNA. The 2D-chiral box gave a stronger signal than 5L-chiral box PNA, which showed very weak hybridization. Therefore, the stability of the PNA:DNA-FM duplexes under these conditions were in a different order than those observed in solution, suggesting that electrostatic interactions which stabilize the duplex in solution can be affected by the matrix effect of the activated slide surface and by the additives normally used for hybridization.

The hybridization with the mismatched oligonucleotide gave the results reported in Figure 10b, which were supplemented by a quantitative analysis comparing full match and mismatch spot intensities: these underlined best recognition properties for the 2D-chiral box PNA than with the two other cases (Fig 11).[26] This is in line with the results of sequence selectivity observed in solution. Therefore, for the recognition of a single nucleotide mismatch on the array, the 2D-chiral box model was also in this case found to be the best overall performing in terms of sequence selectivity.

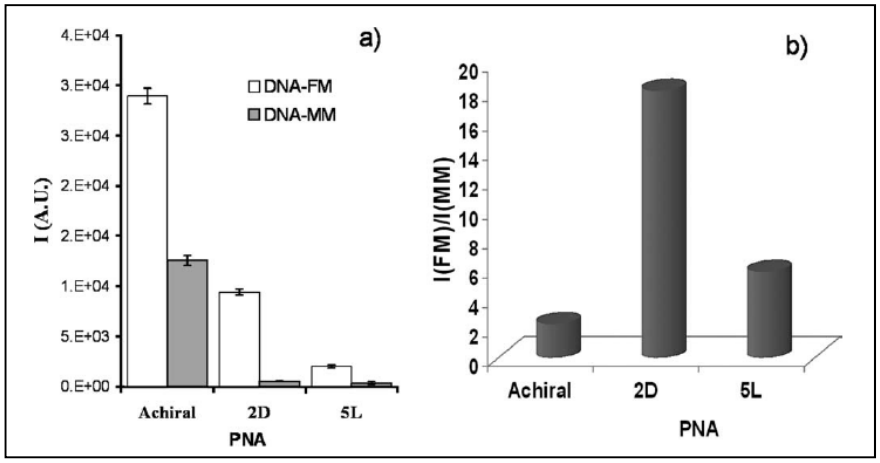


Fig 11. a, Quantitative analysis of microarray signals obtained with achiral PNA, 2D-chiral box PNA and 5L-chiral box PNA with full match (DNA-FM, white bars) and mismatched DNA (DNA-MM, gray bars); vertical bars indicate standard deviations. **b**, Selectivity (I full match/I mismatch) observed in the microarray hybridization for the various PNA. Adapted from reference 26.

4.4 Performance of PNA probes in solution and on surface: a comparison

As shown in described experiments, the behaviour of tested PNA models deeply differs between the binding properties in solution and on surface. 5L-chiral box PNA seemed to be the best model for the DNA binding in solution, followed by fully achiral PNA and 2D-chiral box PNA (Table 1). This situation is diametrically opposed to the PNA behavior on surface. Here, the best binding properties seemed to belong to the achiral PNA while the 5L-chiral box PNA gave very less intense signals (Figg 10a and 11a). On the other side, the high stability of achiral PNA showed on surface was detrimental for the mismatch recognition. Indeed, while in solution all the three PNA structures showed similar good SNP discrimination properties, with 2D-chiral box PNA slightly better performing, the hybridization on array revealed 2D-chiral box PNA as the best structure in terms of selectivity in mismatch recognition, while the achiral PNA still gave a strong signal when the hybridization with mismatch DNA was performed. These results suggest that the surface could play a meaningful role during the hybridization step. It may be speculated that the nature of the solid support and the presence of positive charges on chiral PNA probes are more likely responsible for the reduced binding properties on surface, since they could promote the aggregation of the PNA, lowering the efficiency of the hybridization. Concerning the mismatch recognition properties, it has be noted that there is a major difference between the two types of modifications in chiral PNAs: while the side chains in the 2D model are pointing toward the major groove, in the 5L model they are directed toward the minor groove.[11] The higher sequence selectivity observed in the case of 2D chiral box can be due to the position of the side chain, which is attached to the α-carbon of the more rigid glycine moiety of the PNA backbone, whereas the side chain in the 5L derivative is placed in the more flexible aminoethyl group. Distortion of the former will likely generate a conformation in which the side chains collide with each other, whereas in the case of the 5L derivative, they can be rearranged and the eventual repulsive interactions are compensated by the electrostatic interactions with the negative potential of the minor groove. Since the side chains in the 5L model are closer to the DNA backbone, however, eventual surface-PNA interactions on this side will strongly hamper the interaction with DNA, while the same effect occurring on the major groove side would be better tolerated, thus leading to a better performance on the microarray system.[26]

5. Conclusions

Since several works had showed the importance of chirality of PNAs in affecting in a positive way DNA recognition and enhancing sequence selectivity, in this chapter the performance of three different PNA probes, fully achiral PNA, 2D-arginine-based chiral PNA and 5L- arginine-based chiral PNA, has been investigated, in terms of DNA binding affinity and mismatch recognition, both in solution and on solid surface.

The results showed that the 5L-chiral box PNA was superior in binding affinity in solution, whereas the 2D-chiral box PNA model was superior in performances when recognition of single nucleotides was considered both in solution and in the microarray format. These differences could be mainly addressed to the displacement of the chiral monomers in the PNA-DNA duplex three-dimensional structure and, therefore, to their

interaction with the solid surface. This interfacing, in turns, highly differs according to the chemistry of the stereogenic centres and is enhanced by the presence of several chiral monomers used to form chiral-box PNAs.
The information achieved here can be very precious in the design of a PNA-based microarray for the recognition of single nucleotide polymorphisms in food analysis.

6. References

[1] Nielsen PE, Egholm M, Berg RH, Buchardt O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, 1991, **254**:1497–1500.

[2] Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, Berg RH, Kim SK, Norden B, Nielsen PE. PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature*, 1993, **365**:566–568.

[3] Wittung P, Nielsen PE, Buchardt O, Egholm M, Norden B. DNA like double helix formed by peptide nucleic acid. *Nature*, 1994, **368**:561–563.

[4] Paulasova P, Pellestor F. The peptide nucleic acids (PNAs): a new generation of probes for genetic and cytogenetic analyses. Annales de Génétique, 2004, **47**:349-358.

[5] Yilmaz LS, Okten HE, Noguera DR. Making all parts of the 16s rRNA of Escherichia coli accessible in situ to single dna oligonucleotides. *Applied and Environmental Microbiology*, 2006, **72**:733–744.

[6] Demidov VA, Potaman VN, Frank-Kamenetskil MD, Egholm M, Buchardt O, Sonnichsen SH, Nielsen PE. Stability of peptide nucleic acids in human serum and cellular extracts. *Biochemical Pharmacology*, 1994, **48**:1310–1313.

[7] Corradini R, Sforza S, Tedeschi T, Totsingan F, Marchelli R. Peptide nucleic acids with a structurally biased backbone: effects of conformational constraints and stereochemistry. *Current Topics in Medicinal Chemistry*, 2007, **7**:681-694.

[8] Giesen U, Kleider W, Berding C, Geiger A, Orum H, Nielsen PE. A formula for thermal stability (Tm) prediction of PNA/DNA duplexes. *Nucleic Acids Research,* 1998, **26**:5004–5006.

[9] Ren B, Zhou JM, Komiyama M. Straightforward detection of SNPs in double-stranded DNA by using exonuclease III/nuclease S1/PNA system. *Nucleic Acids Research*, 2004, **32**:e42.

[10] Corradini R, Sforza S, Tedeschi T, Marchelli R. Chirality as a tool in nucleic acid recognition: principles and relevance in biotechnology and in medicinal chemistry. *Chirality*, 2007, **19**:269–294.

[11] Menchise V, De Simone G, Tedeschi T, Corradini R, Sforza S, Marchelli R, Capasso D, Saviano M, Pedone C. Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA-DNA duplex. *Proceedings of the National Academy of Sciences USA*, 2003, **100**:12021–12026.

[12] Sforza S, Tedeschi T, Corradini, Marchelli R. Induction of helical handedness and DNA binding properties of peptide nucleic acids (PNAs) with two stereogenic centres. *European Journal of Organic Chemistry*, 2007, **2007**:5879– 5885.

[13] Sforza S, Corradini R, Ghirardi S, Dossena A, Marchelli R. DNA binding of a D-lysine-based chiral PNA: direction control and mismatch recognition. *European Journal of Organic Chemistry*, 2000, **2000**:2905–2913.

[14] Tedeschi T, Sforza S, Dossena A, Corradini R, Marchelli R. Lysine-based peptide nucleic acids (PNAs) with strong chiral constraint: control of helix handedness and DNA binding by chirality. *Chirality,* 2005, **17***:*S196–S204.
[15] Duheolm KL, Petersen KH, Jensen DK, Egholm M, Nielsen PE, Buchardt O. Peptide nucleic acid (PNA) with a chiral backbone based on alanine. *Bioorganic & Medicinal Chemistry Letters*, 1994, **4:**1077-1080.
[16] Nielsen PE, Haaima G, Lohse A, Buchardt. Peptide nucleic acids (PNAs) containing thymine monomers derived from chiral amino acids: hybridization and solubility properties of D-lysine PNA. *Angewandte Chemie International Edition in English*, 1996, **35**:1939-1942.
[17] Calabretta A, Tedeschi T, Di Cola G, Corradini R, Sforza S, Marchelli R. Arginine-based PNA microarrays for APOE genotyping. *Molecular BioSystem*, 2009, **5**:1323-1330.
[18] Sforza S, Haaima G, Marchelli R, Nielsen PE. Chiral peptide nucleic acids (PNAs): helix handedness and DNA recognition. *European Journal of Organic Chemistry,* 1999, **1**:197-204.
[19] Germini A, Rossi S, Zanetti A, Corradini R, Fogher C, Marchelli R. Development of a peptide nucleic acid array platform for the detection of genetically modified organisms in food. *Journal of Agricultural and Food Chemistry,* 2005, **53**:3958-3962.
[20] Rossi S, Scaravelli E, Germini A, Corradini R, Fogher C, Marchelli R. A PNA-array platform for the detection of hidden allergens in foodstuffs. *European Food Research and Technologies*, 2006, **223**:1-6.
[21] Germini A, Scaravelli E, Lesignoli F, Sforza S, Corradini R, Marchelli R. Polymerase chain reaction coupled with peptide nucleic acid high-performance liquid chromatography for the sensitive detection of traces of potentially allergenic hazelnut in foodstuff. *European Food Research and Technologies*, 2005, **220**:619-624.
[22] Kalendar R, Lee D, Schulman AH. FastPCR software for PCR primer and probe design and repeat search. *Genes, Genomes and Genomics*, 2009, **3**:1-14.
[23] http://www6.appliedbiosystems.com/support/pnadesigner.cfm.
[24] http://eu.idtdna.com/ analyzer/ Applications/OligoAnalyzer
[25] Sforza S, Corradini R, Galaverna G, Dossena A, Marchelli R. The chemistry of peptide nucleic acids. *Minerva Biotecnologica*, 1999, **11**:163-174.
[26] Manicardi A, Calabretta A, Bencivenni M, Tedeschi T, Sforza S, Marchelli R. Affinity and selectivity of C2- and C5-substituted "chiral-box" PNA in solution and on microarrays. *Chirality*, 2010, **22**:E161- E172.

Chapter VII

A PNA microarray for the identification of tomato varieties through single nucleotide polymorphisms (SNPs) detection

1. Introduction

The development of systems able to guarantee the identification of food material from farm to fork is deemed more and more necessary, due to the consumers' demand for high quality products. As any quality ingredient is in danger of being counterfeit, the possibility of univocally identify the ingredients in any step of the food chain represents also a means of protecting both high-quality brand names and consumers. For many cultivated plant species the most frequent fraudulent adulteration consists in the replacement of the declared cultivar with others of lower commercial value. Besides damaging the companies that legitimately employ and exploit a specific brand or product characteristic, the improper use of a denomination can cause confusion among consumers, who can not distinguish in the purchased product its distinctive properties. Nowadays molecular markers represent the most reliable tool for varietal identification in plants. As, in principle, DNA can be found in any food product, application of DNA analyses can be a valuable source of information on the identity of food and feed ingredients.[1] In the past years molecular diversity of plant species has been characterized by the development of a high number of molecular markers, which revealed very useful polymorphisms. Concerning tomato, genetic variability among cultivated varieties showed to be far lower than that observed in other self-compatible species.[2] Besides the strict autogamy of this species, the low rate of polymorphism may be due to the overexploitation of a limited genetic pool in breeding and selection programs. Among the different DNA markers with a high discriminatory power,[3,4,5] single nucleotide polymorphisms (SNPs), which consist in single nucleobase changes within the genome, can be successfully used to identify a specific tomato variety,[6,7] even in processed matrices. A wide range of bioassays for the recognition of SNPs has been developed in the last few years, either in solution or on surface,[8] all aimed at the development of fast, cheap, robust and high-throughput methods. Among these, surface techniques, and in particular microarray-based platforms,[9,10] turned out to be extremely interesting. Such methods usually rely on the recognition of a DNA target by hybridization with a single strand oligonucleotide probe immobilized onto a surface. The quality and efficiency of these methods can be seriously affected by the nature of the probes used to recognize complementary sequences; in particular, oligonucleotide probes may be replaced with modified molecules with an improved affinity and improved specificity in DNA binding, leading to improved recognition properties.[11,12] Peptide nucleic acids (PNAs) are among the most promising probes, showing improved binding properties, if compared to DNA probes, as well as higher chemical and enzymatic resistance. The development of PNA-based surface systems allowed to obtain more efficient assays, in terms of selectivity in the recognition of point mutation, robustness and sensitivity. In food analysis, although still not exploited routinely, several promising applications of PNAs have been published in the last few years. In particular PNA-microarrays were successfully used for the detection of GMOs, hidden allergens and ingredient authentication.[13]

As already described in the previous chapter, PNA-microarrays properties in terms of specificity of recognition, can be further improved by introducing new chemical modifications in the PNA structure, able to allow a straightforward clear-cut discrimination, even in the case of single mismatch recognition (SNP). The synthesis and the binding properties of PNAs containing arginine-based monomers have been

recently reported,[14] demonstrating their enhanced recognition properties, in terms of binding affinity and mismatch recognition, in solution and on microarray platforms.[14,15] Such PNAs might constitute an advancement in SNP recognition as compared to traditional DNA-based systems, characterized by poor selectivity, and therefore could be successfully exploited for gaining information on the identity of raw ingredients in foodstuff.

2. Aim of the work

Using argine-based 2D-chiral box PNA (Arg-PNA), whose enhanced recognition properties in terms of binding affinity and mismatch recognition, both in solution and on microarray platforms, have been demonstrated in the previous chapter, a model Arg-PNA microarray for the simultaneous identification of SNPs characteristic of seven different tomato varieties was designed and developed. Highly selective Arg-PNAs were designed, synthesized and their binding properties in solution were assessed. PNA-microarrays based on these probes were prepared and applied to SNP discrimination, at first in model experiments, using oligonucleotide mixtures simulating the different sequences of the seven tomato varieties, and then by using amplified DNA extracted from real food samples. The strength and the limitations of such a system for SNP recognition will be thoroughly discussed.

3. Experimental part

3.1 PNA design

See paragraph 3.1, Chapter VI – Section II

3.2 Hybridization on microarray

See paragraph 3.3, Chapter VI – Section II

3.3 Plant materials and food matrices

Among the tomato varieties, which in 2007 were the most represented in the Italian seed market, seven were chosen for this study (Table 1). We also analysed seven commercially available tomato products, two chopped, three peeled and two cherry canned tomatoes. They corresponded to the mono-varietal sauces made of the same listed tomato varieties. All these materials were obtained from the tomato food products producer “Annalisa” Lodato S.p.A (Salerno, Italy).

Table 1. Tomato varieties used in this study.

Variety	Fruit Shape	Destination
Leader	Rounded	Chopped tomatoes
Ercole	Elongated	Peeled tomatoes
Talent	Elongated	Peeled tomatoes
PS 1296	Rounded	Chopped tomatoes
Tomito	Cherry	Cherry tomatoes
Minidor	Cherry	Cherry tomatoes
Cirio 3	Elongated	Peeled tomatoes

3.4 DNA extraction from fresh and canned tomato

3.4.1 Chemicals

- D-Sorbitol ($C_6H_{14}O_6$) (Fluka, Sigma-Aldrich, USA)
- Tris base (Sigma-Aldrich, USA)
- Ethylenediaminetetraacetic acid ($C_{10}H_{16}N_2O_8$) (EDTA) (Fluka, Sigma-Aldrich, USA)
- Sodium chloride (NaCl) (AnalaR Normapur®, Italy)
- Cetrimonium bromide ($(C_{16}H_{33})N(CH_3)_3Br$) (CTAB) (Sigma-Aldrich, USA)
- Chloridric acid (HCl) (Carlo Erba, Italy)
- Milli Q H_2O obtained with Millipore Alpha Q system
- Sodium dodecyl sulfate (SDS) (5% solution) (Fluka, Sigma-Aldrich, USA)
- DNA extraction buffer 1:
  - 6.37 g $C_6H_{14}O_6$ (0.35 M)
  - 1.21 g Tris base (0.1 M)
  - 0.14 g EDTA (5 mM)
  - Adjust the pH to 7.0 with HCl solution
  - Add Milli Q H_2O to 100 mL
- DNA extraction buffer 2:
  - 2.42 g Tris base (0.2 M)
  - 1.46 g EDTA (0.05 M)
  - 11.68 g NaCl (2 M)
  - 2 g CTAB (2%)
  - Add Milli Q H_2O to 100 mL
- DNA Microprep Buffer (10 mL for each sample):
  - 2.5 mL DNA extraction buffer 1
  - 2.5 mL DNA extraction buffer 2
  - 1 mL 5% SDS solution
- Chloroform ($CHCl_3$) (AnalaR Normapur®, Italy)
- Isopropanol ($CH_3CH(OH)CH_3$) (Carlo Erba, Italy)
- Ethanol (EtOH) (70% solution in Milli Q H_2O) (Sigma-Aldrich, USA)
- N_2 flux

- Sterile H_2O

3.4.2 Instrumentation

- Digital Scale BCE 62 PT (Orma, Italy)
- Homogenizer Ultraturrax T50 basic (IKA Werke, Germany)
- pH meter 212 (Hanna Instruments, Italy)
- Centrifuge Universal 320 R (Hettich, Germany)
- Hotplate Stirrer F 80 (Falc)
- Autovortex SA6 (Stuart Scientific, UK)
- Qubit fluorometer 1.0 (Invitrogen, UK)

3.4.3 Procedure

40 g of fresh (whole fruit) or canned tomato were weighed in a beaker and homogenized. 10 g of blended sample were transferred to a 50 mL plastic tube and an equal volume of DNA Microprep buffer (10 mL) was added. The solution was mixed and it was incubated at 65 °C for 1 h, under gently shaking. Afterwards, the sample was centrifuged at 9000 r.p.m. for 15' at 4 °C and the supernatant was recovered and moved to another 50 mL plastic tube. An equal volume (~ 10 mL) of chloroform was added and the two phases were mixed by inverting the tube gently. After a centrifugation step at 8000 r.p.m. for 15' at 4 °C, the aqueous phase was recovered and transferred to a clean tube. The separation step with chloroform was repeated whenever the supernatant wasn't clear enough. An equal volume (~ 10 mL) of cold isopropanol (-20 °C) was added to the supernatant and the sample was mixed gently by inverting, in order to allow DNA precipitation. The pellet was separated by centrifugation at 9000 r.p.m. for 45' at 4 °C. The supernatant was discarded and the pellet was washed with 70% EtOH and moved to 1.5 mL eppendorf tube. After another step of 10' in centrifuge, the EtOH was evaporated under N_2 flux and the pellet was re-suspended in 500 µL sterile H_2O. The concentration of extracted DNA was determined using the Qubit® fluorometer (Invitrogen) according to the instructions by the manufacturer. The extracts were frozen at –20 °C until the analysis time.

3.5 DNA purification

3.5.1 Chemicals

- Isopropanol ($CH_3CH(OH)CH_3$) (80% solution) (Sigma-Aldrich, USA)
- Sterile and apyrogenic H_2O pre-warmed at 80 °C (Salf SPA, Italy)
- Wizard DNA Clean Up System (Promega Corporation, USA)
  - Wizard® DNA Clean-Up Resin
  - Wizard® Minicolumns
  - 100 Syringe Barrels (3 cc)

3.5.2 Instrumentation

- Vacuum Manifold Visiprep™ DL (Supelco, Sigma-Aldrich, USA)
- Centrifuge 1-13 (Sigma Laboratory Centrifuges, Germany)
- Hotplate Stirrer F 80 (Falc)

3.5.3 Procedure

One Wizard® Minicolumn was used for each sample. The provided Syringe Barrel was attached to the Luer-Lok® extension of each minicolumn. The tip of the Minicolumn/Syringe Barrel assembly was inserted into the vacuum manifold. The Wizard® DNA Clean-Up resin was thoroughly mixed before removing an aliquot. If crystals or aggregates were present, they were dissolved by warming the resin to 37 °C for 10'. 1 mL of Wizard® DNA Clean-Up Resin was added to a 1.5 mL microcentrifuge tube containing 500 µL of sample and the solution was mixed by inverting several times.

The resin/DNA mix were pipetted into the syringe barrel. The vacuum was applied to draw the solution through the minicolumn and it was broken before the minicolumn dried up. To wash the column, 2 mL of 80% $CH_3CH(OH)CH_3$ were added to the syringe barrel, and the vacuum was re-applied to draw the solution through the minicolumn. The resin was completely dried by continuing to draw a vacuum for 30'' after the solution has been pulled through the column. The syringe barrel was removed and the minicolumn was transferred to a 1.5 mL centrifuge tube. After a centrifuge step at 10000 r.p.m. for 7' to remove any residual $CH_3CH(OH)CH_3$, the minicolumn was transferred to a new centrifuge tube. 50 µL of pre-warmed H_2O were applied to the minicolumn and incubated for 1'. The centrifuge tube was centrifuged for 5' at 10000 r.p.m. to elute the bound DNA. The minicolumn was removed and discarded and the purified DNA was stored in the microcentrifuge tube at –20 °C.

3.6 Oligonucleotide primers design

On the basis of the selected tomato sequences containing SNPs (LeOH 8.4, LeOH 23.1, LeOH 31.3 and LeOH 63, see Results and Discussion section for details), which allowed the discrimination of the tomato analyzed hybrids, different primer pairs were designed to univocally amplify target DNA regions. The primer design was entirely performed using on-line available resources (Primer3 software v 0.9).[16] All the primer pairs that successfully amplified their targets were then tested using the FastPCR software v 6.0,[17] to control their suitability to be used in a multiplex system.

3.7 PCR setting

3.7.1 Chemicals

- Phire® Hot Start II DNA Polymerase 5 U/µL (Finnzymes, Thermo Scientific, Finland)
- Phire Reaction Buffer 5 x containing 7.5 mM $MgCl_2$ (Finnzymes, Thermo Scientific, Finland)
- Deoxynucleotides 100 mM (dATP, dCTP, dGTP and dTTP) (10 mM solution for each dNTPs) (Euroclone, Italy)
- Oligonucleotide primers desalted and lyophilized (20 µM solution in sterile and apyrogenic H_2O) (Thermo Scientific, Italy)
- Sterile and apyrogenic H_2O (Salf SPA, Italy)

3.7.2 Instrumentation

- Aura PCR workstation (BioAir Instruments s.r.l., Italy)
- PCR Sprint Thermal Cycler (Thermo Hybaid, UK)

3.7.3 Procedure

i. Simplex PCR

First of all the efficiency of the primers in amplifying the target sequences were tested by performing PCRs using each primer pair with DNA extracted from both fresh and canned tomatoes.

All PCR procedures were performed in a final volume of 50 µL with the following reagent concentrations: genomic DNA 30 µg, Phire Reaction Buffer 1 x, $MgCl_2$ 1.5 mM, dNTPs 0.2 mM each, primer 0.4 µM each (forward and reverse), Phire® Hot Start II DNA Polymerase 0.15 U/µL. Thermal cycle conditions were as follows: pre-incubation at 95 °C for 5'; 40 cycles consisting of dsDNA denaturation at 95 °C for 40'', primer annealing at 58 °C for 40'', primer extension at 72 °C for 40''; and final elongation at 72 °C for 5'.

ii. Multiplex PCR

In the multiplex PCR a simultaneous amplification of 4 target sequences of tomato genomic DNA was performed, containing all the primer pairs. Considering the primer pairs/target response observed in the simplex PCR, the concentration of each primer pairs was set as follows: 0.5 µM for LeOH 8.4, LeOH 23.1 and LeOH 63 primer pairs, 0,8 µM for LeOH 31.3 primer pair. Besides dNTPs concentration was slightly increased to 0.3 mM each. Thermal cycle conditions were the same as described above.

iii. Unbalanced PCR

In this case a double amplification was performed. The first amplification consisted of a duplex PCR performed as described above. During the second amplification 3 µL of amplified material was used as template for an unbalanced PCR in which the primer of each pair copying the target sequence, that is the forward primer, was labelled with a Cy5 group; the concentration of labelled and unlabelled primers were 0.5 and 0.05 µM respectively. The PCR procedures were performed in a final volume of 100 µL and the same thermal cycle conditions already described were applied.

iv. PCR for Single-Stranded Products

The procedure used was the same above-reported for the double-stranded products except that the PCR primers which wouldn't incorporate in the single DNA strands to be hybridized with the PNA array (reverse primers) were 5'-phosphorylated, while the primer of each pair copying the target sequence was labelled with a Cy5 group. The PCR products obtained were digested with λ-exonuclease (see paragraph 3.9).

3.8 Agarose gel electrophoresis

3.8.1 Chemicals

- TBE buffer 10 X (Euroclone, Italy):
  - 0.89 M Tris pH 8

  - 0.89 M boric acid
  - 0.02 M EDTA
- GellyPhor®LM agarose (Euroclone, Italy)
- Ethidium bromide (EtBr) 5 mg/mL (Euroclone, Italy)
- Milli Q H_2O obtained with Millipore Alpha Q system
- Orange G dye (Sigma-Aldrich, USA)
- Glicerol (50% solution) (Fluka, Sigma-Aldrich, USA)
- Ethylenediaminetetraacetic acid $C_{10}H_{16}N_2O_8$ (EDTA) (0.5 M solution in H_2O, pH 8) (Fluka, Sigma-Aldrich, USA)
- Orange G loading dye 5 x:
  - 0.01 g Orange G Dye
  - 3 mL Glicerol (30%)
  - 0.6 mL EDTA (0.06 M)
  - Add Milli Q H_2O to 5 mL
- EZ Load™ 20 bp Molecular Ruler (Bio-Rad, Germany)

3.8.2 Instrumentation

- HU10 Mini Plus Horizontal Gel Unit (Scie-Plas, UK)
- Photo UV 20 trans-illuminator (Euroclone, Italy)
- MPSU-200/100 Mini Power Supply Unit (Euroclone, Italy)
- Digital camera Coolpix 2000 (Nikon Corporation, Japan)

3.8.3 Procedure

PCR products were analyzed using agarose gel electrophoresis. The gel was prepared with 4% or 2% of agarose in Tris-Borate EDTA 0.5 x with 0.5 μg/mL of EtBr. The samples to be analyzed were prepared as follows: 16 μL of amplified DNA and 4 μL of loading dye. The running conditions were constant voltage at 120 V for 1 h in TBE 0.5 x. The amplified fragments were displayed by placed the gel on a UV trans-illuminator and comparing their molecular weight to the one of molecular standards. Gel image were then acquired by a camera.

3.9 Enzymatic digestion by λ-exonuclease

3.9.1 Chemicals

- λ-exonuclease reaction buffer 10 x (Euroclone, Italy)
- λ-exonuclease enzyme 5 U/μL (Euroclone, Italy)

3.9.2 Instrumentation

- PCR Sprint Thermal Cycler (Thermo Hybaid, UK)

3.9.3 Procedure

The amplified fragments underwent an enzymatic digestion by λ-exonuclease which recognized and selectively digested the 5'-phosporilated DNA strand in a double-stranded DNA. The reaction mix concentrations were as follows: λ-exonuclease reaction buffer 1 x, 10 enzyme units, 500 ng double-stranded PCR product in a final

volume of 100 μL. The digestion was carried out in a thermal cycler for 45' at 37 °C. The reaction was stopped by deactiving the enzyme for 10' at 75 °C.

3.10 PCR purification protocol

3.10.1 Chemicals

- Ethanol (EtOH) (Sigma-Aldrich, USA)
- EUROGOLD Cycle-Pure Kit protocol (Euroclone, Italy)
  - PerfectBind DNA columns
  - 2 mL Collection Tubes
  - SPW wash buffer
  - XP1 buffer
  - Elution buffer (10 mM Tris-HCl, pH 9,0)

3.10.2 Instrumentation

- Centrifuge 1-13 (Sigma Laboratory Centrifuges, Germany)
- Autovortex SA6 (Stuart Scientific, UK)

3.10.3 Procedures

After the reaction volume has been determinated, an equal volume of XP1 buffer was added and the solution was mixed. The sample was applied to a PerfectBind DNA column assembled in a clean 2 mL collection tube and centrifuged at 10000 r.p.m for 1' at room temperature. The liquid was discard and the collection tube was re-used in the following steps. The PerfectBind DNA column was washed by adding 650 μL of SPW wash buffer diluted with absolute ethanol, according to the manufacturer's instructions. A centrifuge step was repeated at 10000 r.p.m. for 1' at room temperature. The liquid was discarded and this step was repeated once. The liquid was discarded and the empty column was centrifuge for 1' at 10000 r.p.m. to dry the column matrix. This step was essential for good DNA yields. The PerfectBind DNA column was placed into a clean 1.5 mL microcentrifuge tube and 30 – 50 μL of elution buffer were added (depending on desired concentration of final product) directly onto the column matrix and centrifuged for 1' at 5000 r.p.m. to elute DNA.
This represented approximately 80 – 90 % of bound DNA. An optional second elution will yield any residual DNA, though at a lower concentration.

4. Results and Discussion

4.1 Choice of tomato DNA sequences containing SNPs

Tomato hybrids used for the analysis have been described in the experimental part. A set of 20 EST-SNPs, as discussed in the literature,[18] and characterized by high probability to be polymorphic in cultivated tomatoes, were selected and assayed as CAPS on genomic DNA of these tomato hybrids, characterized by different fruit shapes. SNP identity was verified by sequencing the PCR-amplified products, as described elsewhere.[3,4] The identified SNPs were illustrated below (Table 2).

Table 2. List of SNPs detected in tomato hybrids, as described in the text.

SNP loci	Tomato hybrids						
	Leader	**Ercole**	**Talent**	**PS1296**	**Tomito**	**Minidor**	**Cirio3**
LEOH2	A*/A*	A*/A*	A*/A*	A*/A*	A*/A*	A*/A*	A*/A*
LEOH8.4	C*/T	C*/T	C*/T	T/T	C*/T	C*/C*	C*/C*
LEOH11.1	T/T	uc/uc	uc/uc	uc/uc	uc/uc	uc/uc	uc/uc
LEOH12.2	T*/T*	T*/T*	T*/T*	T*/T*	T*/T*	T*/T*	T*/T*
LEOH16.2	C*/T	C*/T	C*/C*	C*/C*	C*/C*	C*/C*	C*/C*
LEOH 16.3	T/C*	T/C*	C*/C*	C*/C*	C*/C*	C*/C*	C*/C*
LEOH20.1	G*/G*	G*/G*	G*/G*	G*/G*	G*/G*	G*/G*	G*/G*
LEOH23.1	G*/uc	G*/uc	del/del	G*/uc	G*/uc	G*/G*	G*/G*
LEOH25.1	uc/uc	uc/uc	C*/uc	uc/uc	C*/uc	C*/uc	uc/uc
LEOH28	T/T	uc/uc	uc/uc	uc/uc	uc/uc	uc/uc	uc/uc
LEOH30.1	-	-	-	-	-	-	-
LEOH31.3	A/A	A/A	G*/A	uc/uc	G*/uc	G*/uc	A/A
LEOH31.4	uc/uc	A/A	T*/A	uc/uc	T*/uc	T*/uc	uc/uc
LEOH32.1	T/T	T/T	C*/T	uc/uc	C*/uc	C*/uc	uc/uc
LEOH35.5	uc/uc	uc/uc	A/G*	uc/uc	G*/uc	G*/uc	uc/uc
LEOH39	uc/uc	T/T	uc/uc	uc/uc	uc/uc	uc/uc	uc/uc
LEOH31.6	T/T	T/T	C*/T	uc/uc	C*/uc	C*/uc	uc/uc
LEOH31.7	T*/T*	T*/T*	T*/uc	T*/T*	T*/uc	T*/uc	T*/T*
LEOH31.10	T/T	T/T	G*/uc	T/T	G*/uc	G*/uc	T/T
LEOH 63	T/T	C*/T	C*/T	C*/T	C*/T	T/T	C*/C*

(base)*: observed cut on expected polymorphic nucleotide recognized by CAPS system
uc: no observed cut by CAPS system (no polymorphism); no sequencing
(base): no observed cut by CAPS system (no polymorphism); base confirmed by sequencing
del: nucleotide deletion
-: no observed amplification

Among the detected SNPs, 4 loci were selected, according to their discriminating power among the seven tomato varieties, as reported in Table 3.
As illustrated below, the combination of polymorphisms observed at the loci LeOH 8.4, LeOH 23.1, LeOH 31.1 and LeOH 63 could allow the univocal identification of each tomato variety. For example, at the locus LeOH 8.4 the genetic asset of 'PS1296' is different from all the others at the same locus, allowing to differentiate this variety from all the others. 'Talent' and 'Cirio' present univocal genetic asset at the loci LeOH 23.1 and LeOH 63, respectively. 'Ercole' differentiates from 'Leader' at the locus LeOH 63; from 'Tomito' at the locus LeOH 31.3 and presents a different polymorphism combination from 'Minidor' for all the chosen loci. 'Leader' is discriminated from 'Tomito' at the loci LeOH 31.3 and LeOH 63 and from 'Minidor' at all the other loci, but LeOH 63. 'Tomito' and 'Minidor' could be discriminated at each locus, but LeOH 31.3, where they show the identical G/A genetic asset.

Table 3. Selected SNPs for variety discrimination. 'Cirio3', 'Talent' and 'PS1296' are marked by coloured boxes since they are discriminated from all the others by one locus, directly. All the other varieties are distinguished by a combination of polymorphisms at the selected *loci*, which is indicated by symbols.

Varieties \ Loci	LeOH 8.4	LeOH 23.1	LeOH 31.3	LeOH 63
Ercole ▲	C/T	G/del	A/A	C/T
Cirio3	C/C	G/G	A/A	C/C
Talent	C/T	del/del	G/A	C/T
Leader ●	C/T	G/del	A/A	T/T ▲
Ps1296	T/T	G/del	A/A	C/T
Tomito ♦	C/T	G/del	G/A ● ▲	C/T ●
Minidor	C/C ▲ ● ♦	G/G ▲ ● ♦	G/A ▲ ●	T/T ▲ ♦

The knowledge of these polymorphisms was therefore exploited to design argine-based 2D-chiral box PNAs, able to recognize the single nucleotide mutations during the formation of PNA-DNA duplexes in microarray format, suitable for the discrimination of tomato varieties.

4.2 PNA design

After the preliminary studies presented in the previous chapter had showed that the argine-based 2D-chiral box PNA was superior in performances when recognition of single nucleotides was considered, both in solution and in the microarray format (see chapter VI - section II), the design of other 2D-chiral box PNAs was performed, on the basis of the other three tomato DNA sequences (Table 3 and Fig 3) to allow the development of a microarray for the simultaneous discrimination of tomato varieties.
The PNA sequences were chosen using as basis the 2D-chiral box model, with the polymorphic base to detect in the middle of the strand. Moreover, PNAs were chosen to maximize their hybridization efficiency with their target strand and to minimize any unspecific hybridization with any other non-target sequence. For these reasons, the PNA probes were first checked to minimize any secondary structure or self- and hetero-dimer, which would result in a loss of hybridization efficiency, using the online available program IDT – Integrated DNA Technologies (http://eu.idtdna.com/ analyzer/ Applications/OligoAnalyzer). The sequence specificity of the designed probes was then evaluated to avoid any possible hybridization on other non target regions, by aligning each PNA sequence to the whole genomic sequences by NCBI BLAST.
The characteristics of the selected PNA probes for the simultaneous recognition of tomato SNPs were reported in Table 4.

Table 4. List of the more suitable PNA probes for the simultaneous discrimination of seven tomato varieties. The bold underlined nucleotides are the ones implied in the SNP recognition, red nucleobases represent the 'chiral box'.

Locus Name	LeOH 8.4		LeOH 23.1		LeOH 31.3	LeOH 63	
PNA Name	PNA 4	PNA 1	PNA 2	PNA 7	PNA 5	PNA 6	PNA 3
PNA Sequence (H-NH_2)	AAAGGCCGA	AAAGACCGA	TTTTCCGGTGG	TTTTCGGTGG	CTTATCCGGTGCC	TGTGTCGAAAG	TGTGTCAAAAG
T_m c=[1 µM]	66 °C	60 °C	61 °C	56 °C	65 °C	64 °C	60 °C
Molecular Weight	2494.7	2478.7	3016.3	2765	3487.9	3058.3	3042.3
Length	Residues: 9	Residues: 9	Residues: 11	Residues: 10	Residues: 13	Residues: 11	Residues: 11
Purine %	77.8%	77.8%	36.4%	40.0%	30.8%	63.6%	63.6%
Polymorphic SNP	AAAG**G**CCGA	AAAG**A**CCGA	TTTT**C**CGGTGG	TTTT**–**CGGTGG	CTTATC**C**GGTGCC	TGTGTC**G**AAAG	TGTGTC**A**AAAG
Complementarity	High	Low	High	Low	High	Moderate	Low

H means free N-terminus, NH_2 means carboxamide C-terminus. '-' in PNA 7 indicates a nucleobase deletion.

As shown in Table 4, a PNA probe was designed for each SNP to be detected, except for the locus LeOH 31.3, in order to better exploit the discrimination power of the whole system. Since containing different percentage of purine, PNA probes were designed with different lengths to maintain the same purine/pyrimidine ratio and consequently the same theoretical T_m at the proper concentration. For example, in the case of locus LeOH 8.4, the more suitable length of PNA probes was of 9 residues: due to their high content in purine nucleotides (above 75%) their theoretical T_m were comparable to the ones of longer probes with a lower percentage of purine content (that is PNA 3, PNA 5 and PNA 6).

In the case of PNA 2 and PNA 7, recognizing SNPs at the locus LeOH 23.1, the central position did not actually correspond to a SNP, but to a nucleobase deletion (C is missed in PNA 7).

The PNA synthesis was performed as described before, using the 2D-chiral-box chemistry. Afterwards, synthetic probes were purified by RP-HPLC, characterized by ESI-MS-RP-HPLC and quantified by UV absorbance.[19]

4.3 PNA binding properties in solution

As already performed for the trial PNA in the previous chapter, the PNA probe affinity towards complementary DNA oligonucleotide sequences, mimicking tomato genomic strands, and selectivity for mismatch recognition were initially investigated in solution, by measuring the T_m of both full match and mismatch PNA–DNA duplexes (Table 5).[19]

Experiments were performed in phosphate buffer (10 mM, pH = 7), at a strand concentration of 5 µM. Melting temperatures were evaluated as the first derivatives of the UV absorption curves at 260 nm in a temperature range from 20 to 90 °C.

The values (Table 5) showed that full match oligonucleotides were recognized and bound with a high affinity, with T_m ranging in a quite close range, from 54 to 66 °C. PNAs 1, 3, 4, 5 and 6 were also shown to be able to perform mismatch recognition, with an efficiency which turned out to be quite variable, ranging from good (10 °C drop in T_m for PNA 6) to exceptional (36 °C drop for PNA 1). Very likely,

unpredictable conformational changes due to the sterical hindrance of the arginine side chains and the chiral constrain affected duplex formation, thus affecting the final duplex stability.

Table 5. PNA sequences and T_m of the PNA:DNA full match/mismatch duplexes in PBS buffer 10 mM (pH 7.0) at a 5 μM concentration for each strand. Adapted from reference 19.

PNA (target DNA accession number-SNP nucleobase)	Sequence[a]	T_m/°C PNA–DNA fullmatch[b]	T_m/°C PNA–DNA mismatch[b]
1 (LeOH 8.4-T)	H–(AEEA)$_2$AAAG**A**CCGA–NH_2	54 (with **1′**)	18 (with **4′**)
2 (LeOH 23.1-G)	H–(AEEA)$_2$TTTT**C**CGGTGG-NH_2	56 (with **2′**)	35 (with **8′**)
3 (LeOH 63-T)	H–(AEEA)$_2$TGTGTC**A**AAAG—NH_2	56 (with **3′**)	41 (with **7′**)
4 (LeOH 8.4-C)	H–(AEEA)$_2$AAAG**G**CCGA—NH_2	60 (with **4′**)	45 (with **1′**)
5 (LeOH 31.3-G)	H–(AEEA)$_2$CTTATC**G**GTGCC—NH_2	66 (with **5′**)	40 (with **6′**)
6 (LeOH 63-C)	H–(AEEA)$_2$TGTGTC**G**AAAG—NH_2	61 (with **7′**)	51 (with **3′**)
7 (LeOH 23.1-del)	H–(AEEA)$_2$TTTT**C**GGTGG–NH_2	56 (with **8′**)	54 (with **2′**)

[a] PNA monomers are indicated with the symbol of the corresponding nucleobase; chiral monomers based on D-Arg are reported in bold, the base corresponding to the SNP position is underlined; AEEA = aminoethoxyethoxyacetyl: spacer. [b] Oligonucleotide sequences (SNP base is bold underlined): 1': 5'-TCGG**T**CTTT-3', 2': 5'-CCACC**G**GAAAA-3', 3': 5'-CTTT**T**GACACA-3', 4': 5'-TCGG**C**CTTT-3', 5': 5'-GGCACC**G**GATAAG-3', 6': 5'-GGCACC**A**GATAAG-3', 7': 5'-CTTT**C**GACACA-3',8': 5'-CCACC-GAAAA-3'.

PNA 2 and PNA 7 are to be considered aside, since they were designed to recognize not a mismatch, but a base deletion. PNA 2, the one complementary to the full sequence, showed a consistent drop in affinity towards the oligonucleotide sequence bearing a deletion in the middle (ΔT_m = -21 °C), indicating that a protruding nucleobase in the PNA strand (in this case a C) was poorly tolerated from the duplex. Quite interestingly, PNA 7, complementary to the sequence with the deletion, showed almost the same affinity towards the oligonucleotide sequence bearing an extra nucleobase in the middle, indicating that a protruding nucleobase in the DNA strand (in this case an A) can be very well tolerated, with almost no effect on the duplex stability.

4.4 PNA microarray simulation for tomato SNPs detection

After that PNA binding properties have been tested in solution, the preparation of the PNA microarray platform with the seven synthesized PNAs was then performed. The theoretical signal pattern expected by using the seven designed PNA probes targeted against the corresponding DNA sequences in the microarray is reported in Fig 1.

	PNA LeOH 8.4-T	PNA LeOH 23.1-G	PNA LeOH 63-T	PNA LeOH 8.4– C	PNA LeOH 31.3-G	PNA LeOH 63-C	PNA LeOH 23.1-del
'Ercole' (oblung)	●	●	●	●		●	
'Cirio3' (oblung)		●		●		●	
'Talent' (oblung)	●		●	●	●	●	●
'Leader' (round)	●	●	●	●			
'Ps1296' (round)	●	●	●			●	
'Tomito' (cherry)	●	●	●	●	●	●	
'Minidor' (cherry)		●	●	●	●		

Fig 1. Simulation of the recognition of the seven tomato varieties by the seven designed PNA probes. For each cultivated variety the fruit shape is reported in brackets. Adapted from reference 19.

PNA deposition was carried out by MICROCRIBI Microarray Service-CRIBI (University of Padova, Italy), as previously reported.[14] The devices were tested by hybridizing them with solutions containing Cy5-labelled synthetic oligonucleotides, mimicking the DNA sequences to be recognized (Table 5). The hybridization conditions were optimized for each probe, both with full complementary and mismatched oligonucleotides, in order to define the best parameters allowing simultaneous hybridization of all PNAs and to ensure that no false-positives with no complementary or mismatch sequences would take place. As reported previously,[14] in order to increase the hybridization efficiency and to refine the spot shapes, an incubation–hydration step in SDS 0.1%, saline sodium citrate (SSC) buffer (0.3 M NaCl, 0.03 M sodium citrate, pH = 7), at 40 °C for 30' was introduced before hybridization with the oligonucleotide solutions. This step was particularly required when using the slides several days after PNA spotting. The hybridization solutions were prepared in SSC buffer, 0.1% SDS by mixing Cy5 labelled oligonucleotides at 1 μM concentration. After optimization of the hybridization conditions, the array was tested by simulating the different tomato varieties, each characterized by its own combination of the different SNP-deletion, by mixing the different oligonucleotide sequences, in order to imitate the real samples, as shown in Table 6.

Table 6. Tomato genotypes related to different SNPs. Adapted from reference 19.

Variety	Oligonucleotide sequences simulation[a]
'Ercole'	1′ + 2′ + 3′ + 4′ + 7′
'Cirio 3'	2′ + 4′ + 7′
'Talent'	1′ + 3′ + 4′ + 5′ + 7′ + 8′
'Leader'	1′ + 2′ + 3′ + 4′
'PS1296'	1′ + 2′ + 3′ + 7′ + 8′
'Tomito'	1′ + 2′ + 3′ + 4′ + 5′ + 7′
'Minidor'	2′ + 3′ + 4′ + 5′

[a] Oligonucleotide sequences are reported in Table 5.

The solutions were poured on the slides just after mixing at room temperature, in order to avoid unspecific interactions. Hybridizations were performed in seven independent experiments, one for each variety at room temperature. The results obtained in optimal conditions are shown in Fig 2. By comparing the experimental results reported in Fig 2 with the theoretical ones in Fig 1, it was possible to outline the performance of the PNA probes on the surface. PNA 1 (LeOH 8.4-T), 4 (LeOH 8.4-C), 5 (LeOH 31.3-G), 3 (LeOH 63-T) and 6 (LeOH 63-C) behaved perfectly, showing signals only in the presence of their fullmatch oligonucleotide counterparts, and not giving any signals in the presence of oligonucleotides bearing a single mismatch simulating the SNP. Thus, even in the presence of small differences in stability shown in solution, all these PNA probes were perfectly able to discriminate point mutations on a surface system. PNA 2 (LeOH 23.1-G) and 7 (LeOH 23.1-del), aimed at recognizing the deletion, on the other side, failed to perform a specific recognition, showing positive signals for all oligonucleotides, with or without the deletion. If this result was somehow expected in the case of PNA 7, for which in solution there was practically no difference in T_m between the target and non-target DNA, the result was

somehow surprising for PNA 2, which in solution had shown a quite consistent difference in affinity towards the two DNA sequences.

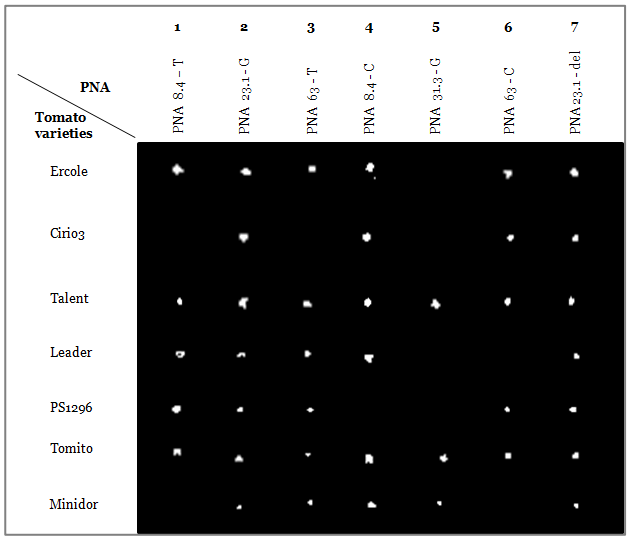


Fig 2. PNA microarray analysis with Cy5-oligonucleotides in different combinations (see Table 6) in order to simulate the seven tomato varieties. Adapted from reference 19.

These results seem to indicate that on the surface, differently than in solution, the recognition of the deletion is somehow difficult, indicating that protruding bases in either strand (PNA or DNA) can be very well tolerated and therefore might not be a good target for PNA probes on a surface system. In any case, the experiment clearly indicated that the seven varieties, except for 'Talent' and 'Tomito', could be discriminated by the microarray here presented.

4.5 DNA extraction from tomato fruits and derivatives

After that the simulation of PNA microarray for tomato variety discrimination has been developed, an investigation of the PNA microarray performance with amplified DNA extracted from real food was undertaken. For this reason, genomic DNA was extracted from both tomato fruits and canned products.

The protocol for DNA extraction described here has been developed elsewhere,[20] and it was previously used for DNA isolation from tomato leaves. It was then applied for genomic DNA extraction from tomato fruits and processed derivatives, in order to check its efficiency and suitability also with these food matrixes.

After extraction, genomic DNA was purified by Wizard® DNA Clean Up System (Promega), in order to purify DNA from digesting enzymes, such as exonucleases and endonucleases, which could degrade DNA after the extraction procedure, or from salts, proteins and organic molecules, which could interfere during the amplification step. Purified DNA was then quantified by Qubit® fluorometer and the concentration for each tomato variety, both for fresh fruits and canned products, was reported in Table 7. It appeared clear that the concentration of DNA extracted from processed tomatoes was drastically reduced. The low amount of detected DNA could obviously ascribed to the heat treatments to which tomatoes were subjected during canned processing and to the prolonged staying in acid conditions during products shelf-life.

Table 7. DNA concentration of tomato fruits and canned products for each variety.

Tomato Varieties	DNA from Tomato Fruit (ng/µL)	DNA from Canned Tomato (ng/µL)
Leader	398	3
Ercole	451	6
Talent	248	10
PS 1296	467	5
Tomito	444	25
Minidor	249	12
Cirio 3	210	8

These rigid handlings can also affect the integrity of DNA molecules, resulting in high-fragmentised regions difficult to be amplified during PCR. For this reason, a trial amplification was performed using a primer pair targeting a tomato-specific gene, Lat52,[21] in order to check the integrity of DNA molecules after processing (Table 8).

Table 8. Primers used for amplificability test for the extracted DNA.

Sequence Target	Primer Pair Sequence (5'-3')	Amplicon Lenght (bp)	T_m (°C)
Lat52	AGACCACGAGAACGATATTTGC-for TTCTTGCCTTTTCATATCCAGACA-rev	92	54.8 54.1

The amplification was carried out on DNA extracted from both tomato fruits and processed products and in all cases a fragment of ~ 92 bp was observed. These results suggested that the DNA isolation and purification systems here applied were appropriate to obtain amplifiable DNA sequences, even from tomato derivatives.

4.6 Development of a multiplex PCR for the simultaneous amplification of tomato polymorphic regions

4.6.1 Primer design

For the development of a multiplex PCR, whose products would allow the discrimination of the seven tomato varieties after PNA array hybridization, a set of four primer pairs was designed, targeting the tomato DNA sequences (Fig 3), chosen among all the others as the high polymorphic ones (Table 3). First of all, several primers were designed and individually tested in silico for their selectivity and efficiency in amplifying their targets. The design of primers and the setting of each single PCR were performed using the Prime3 software,[16] following classical criteria, i.e. length ranging between 18-24 bp, T_m differing of maximum 5 °C among the primer pairs, G-C content lower than 60% and absence of complementary regions, in order to avoid intra- and inter-nonspecific annealing. Since a previous work showed that there is a limit in amplicon size when the DNA template is isolated from processed tomato products,[4] the length of the amplified fragments was also included among the parameters, in order to obtain amplicons smaller than 200 bp. Besides, primer specificity in amplifying their target region was evaluated by aligning each primer to other genomic sequences by NCBI BLAST, in order to avoid their possible

hybridization on any another non-target region among those amplified by the multiplex PCR.

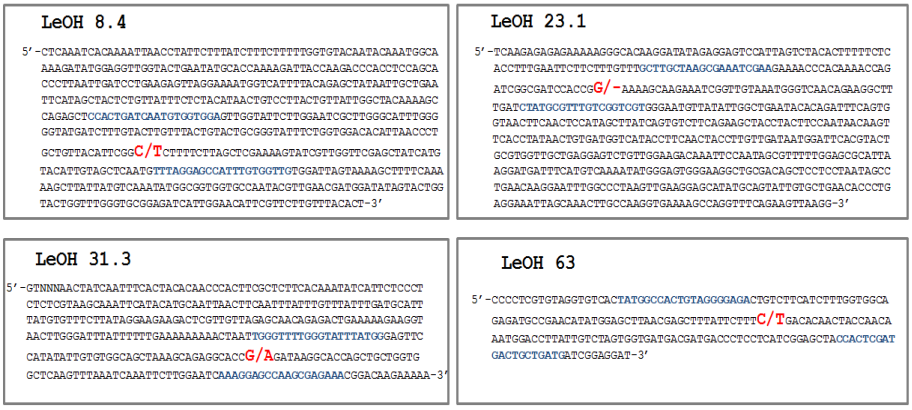


Fig 3. Genomic sequences of the loci LeOH 8.4, LeOH 23.1 and LeOH 63 with the polymorphic bases in bold red. The blue sequences represent primer position. '-' in LeOH 23.1 indicates a nucleobase deletion.

The possibility of combining all the primer pairs together without generating primer-dimers was evaluated by means of the FastPCR Software,[17] and the results were further screened in order to minimize the difference in the melting temperature (T_m), thus allowing the simultaneous amplification of the four target regions. The primers chosen for the multiplex PCR analysis, their sequences and melting temperatures (T_m), and the amplicon lengths are reported in Table 9.

Table 9. Primers used for the Multiplex PCR and their characteristics.

Sequence Target	Primer Pair Sequence (5'-3')	Amplicon Lenght (bp)	T_m (°C)
LeOH 8.4	CCACTGATCAATGTGGTGGA-for CAACCACAAATGGCTCCTAAA-rev	208	60.38 59.98
LeOH 23.1	GCTTGCTAAGCGAAATCGAA-for ACGACCGACAAACGCATAG-rev	118	60.61 59.73
LeOH 31.3	TGGGTTTTGGGTATTTATGG-for TTTCTCGCTTGGCTCCTTT-rev	130	57.23 60.08
LeOH 63	TATGGCCACTGTAGGGGAGA-for CATCAGCAGTCATCGAGTGG-rev	170	60.47 60.43

4.6.2 PCR setting

First of all the efficiency of the primers in amplifying the target sequences were separately tested by performing PCRs using each primer pair corresponding to target genomic DNA (Fig 4): as expected, all the primer pairs successfully amplified the DNA sequences of interest, extracted from tomato leaves, fruits and canned products, giving amplicons with the expected molecular weight, without nonspecific amplified fragments. Particularly, it has to be noticed that the DNA extracted from processed products was still amplified, even when the amplified fragment was bigger than 200 bp (LeOH 8.4), indicating both the quality of the processed tomato DNA and the suitability of the extraction method. It is well-known, indeed, that the proceeding conditions can heavily affect the DNA integrity, making very hard the amplification step.

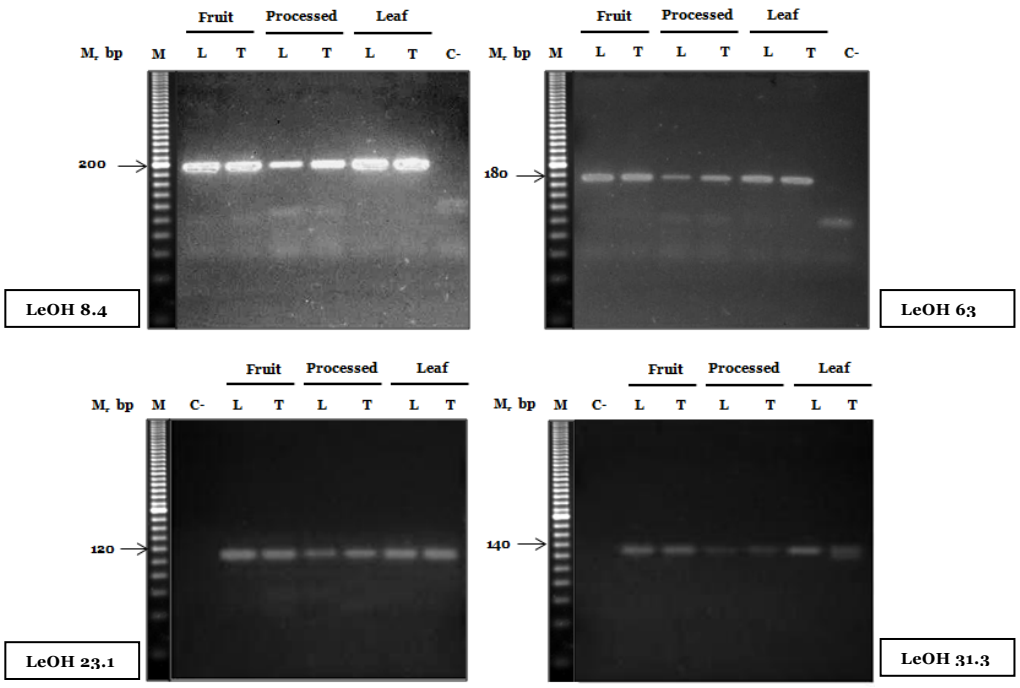


Fig 4. Single PCR amplification of each tomato locus performed on DNA extracted from fruit, leaf and processed tomatoes, using its specific primer pair. 'L': Leader variety; 'T': Tomito variety; 'C-': water negative control; 'M': molecular markers in exact 20 bp increments. The more intense one corresponds to a band at 200 bp.

Of course, differences in the amplification efficiency were shown, indicating the need for an accurate calibration of the primer concentrations in the multiplex, to get comparable amplification response. A weaker amplification efficiency was observed for the PCR targeting the locus LeOH 31.3, probably due to the lower T_m of one of the primers. For this reason, for the multiplex PCR the concentration of the primer pairs amplifying this sequence was increased to 0.8 µM, while the concentration of the other ones was kept at 0.5 µM. The multiplex amplification was carried out amplifying the four tomato DNA sequences simultaneously, by using the four specific primer pairs in the same reaction (Fig 5).

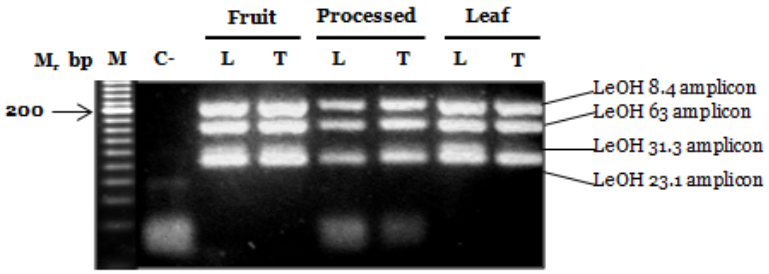


Fig 5. Multiplex PCR of four tomato loci performed on DNA extracted from fruit, leaf and processed tomato, using specific primer pair simultaneously. 'L': Leader variety; 'T': Tomito variety; 'C-': water negative control; 'M': molecular markers in exact 20 bp increments. The more intense one corresponds to a band at 200 bp.

As shown in Fig 5, multiplex PCR gave expected amplicons at each locus for all DNA samples, meaning that the simultaneous amplification wasn't affected by the concurrent presence of all the four primer pairs.

Nevertheless, it appeared quite clear that the strongest amplification signals belonged to tomato leaves and fruits, due to the better quality of extracted DNA compared to the processed one. Indeed, in the lanes corresponding to tomato derivatives a lower band can be observed, due to the primers and dNTPs which were not utilized in the reaction efficiently, because of the scarce availability of the template substrate. The same band was also observed in the water negative control, where no amplification occurred at all. Albeit increased, the concentration of primer pairs targeting the locus LeOH 31.3 didn't improve the amplification efficiency, as shown by the weak band corresponding to the LeOH 31.3 amplicon, whose visualization was further made worse by the slight difference in bp with the LeOH 23.1 amplicon (130 bp and 118 bp, respectively). Although the reduced efficiency of a primer pair, the multiplex PCR procedure here proposed appeared to be a valid method for the simultaneous amplification of the four selected tomato loci, laying the foundations for following hybridization on PNA microarray.

4.7 Amplified sequence hybridization on PNA array

After both simplex and multiplex PCR have been set up and hybridization conditions with oligonucleotides, having the same length as PNAs, have been defined, the performance of hybridization of tomato DNA amplified sequences on PNA-based array was investigated. At first, the LeOH 8.4 amplicon derived from a simplex PCR was used for the hybridization step and, afterwards, analysis was extended to all the other three tomato amplified sequences. The amplification was carried out using the primer copying the target sequence, that is the forward primer, labelled with a Cy5 group and the results were checked by agarose gel electrophoresis. Before the hybridization step on PNA microarray, the solution containing the amplified fragments was denatured at 95 °C for 5', in order to obtain single strands DNA and avoid the formation of secondary structure reducing the hybridization capacity. Hybridization step was performed as described before, both at room temperature and at 40 °C, to reduced self-folding moiety of the single strand DNA, varying the incubation time from 30' to 150'. In all cases, no signals indicating specific PNA–DNA complexation were observed.

In order to facilitate DNA-PNA hybrid formation on the solid surface, a double amplification was performed, at the same locus: the first amplification consisted of a simplex PCR, performed as already described; during the second amplification 3 μL (~ 200 ng) of amplified material were used as template for an unbalanced PCR in which the primer copying the target sequence, that is the forward primer, was labelled with a Cy5 group. The concentrations of labelled and unlabelled primers were 0.5 μM and 0.05 μM and the same thermal cycle conditions already described were applied. Neither with these conditions a specific PNA-DNA hybridization signal was observed. The formation of single strand DNA was also encouraged by increasing the concentration of the forward primer in the reaction mix till 0.7 μM, but no improved results were detected.

In order to selectively obtain only the DNA strand complementary to PNA 1 and PNA 4 (Table 5), dsPCR products, doubly labelled at their 5' termini with a Cy5 dye and a phosphate group, prepared as described in paragraph 3.7.3-*iv*, were digested using λ-exonuclease enzyme which recognized and selectively digested the 5'-phosporilated

DNA strand in a double-stranded DNA. The resulting Cy5-labeled single stranded target amplicon was then used for the hybridization on the array carrying both PNA 1 and PNA 4. Unfortunately, no specific hybridization signal was still obtained.
Different purification steps were then added to the previous PCR systems, in order to remove any possible contaminants, such as polyphenols derived from the DNA extraction step, or excess of primers and dNTPs, which could interfere in the PNA-DNA duplex formation. For this reason all the described PCR experiments were followed by a purification step, performed by EUROGOLD Cycle-Pure Kit protocol, and the residual DNA amount was quantified by Qubit® fluorometer, in order to assure that amplified fragments to be hybridised were still present in solution after purification. Neither this modification allowed us to obtain an improvement in the performance of DNA-PNA duplex formation on array.
Since in all the described hybridization experiments the positive control, that is the PNA on array hybridized with the short oligonucleotide mimicking the complementary genomic sequence, always gave a signal indicating a specific PNA–DNA complexation, we wondered if the absence of any specific signal when using PCR products could be ascribed to a non-properly working amplification reaction or to a problem related to the length of the sequence to be recognized. For this reason an investigation of the microarray performance with long synthetic DNA sequences, mimicking amplified DNA sequences extracted from real food samples, was undertaken.
In particular, the LeOH 23.1 sequence was chosen for testing hybridizations with long DNA tracts. In order to study if and how the different positions of the target sequence in the eventual amplicon would affect the DNA detection, different hybridization experiments by using longer DNA oligonucleotides containing the target sequences in the middle, at the 5' and at the 3' ends, were performed. The different sequences used for hybridization experiments were three 118 bp oligonucleotides differing in the position of the target sequence inside the oligomer, and one shorter variant of one of them (60 bp) (Fig 6).

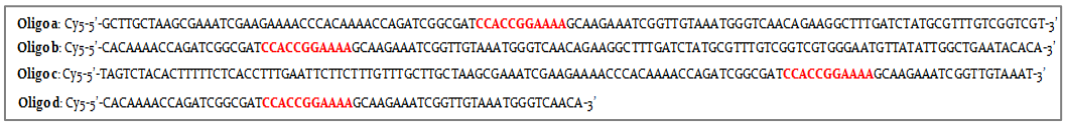


Fig 6. Design of DNA oligonucleotides mimicking the amplified DNA region at the locus LeOH 23.1. The bold red letters indicate the region to be recognized by PNA 2.

Hybridization experiments with the four DNA single strand sequences were performed under the same conditions used for short oligonucleotides. In all cases, no signals indicating specific PNA–DNA complexation were observed, with or without a preliminary incubation step at 95 °C for 5' and/or the hybridization performed at room temperature and at 40 °C. The use of Tween 20 replacing SDS as a non-charged surfactant did not allow us to observe improved results, as well as the use of denaturating conditions (with the addition of small percentages of formamide), in order to make the PCR product more accessible to the PNA probes.
Thus these results seemed to indicate that the chiral box PNA microarray platform, albeit very specific, looses binding efficiency in the presence of longer DNA

sequences, preventing the formation of PNA–DNA duplexes, and this problem, rather than the possibility of non efficient PCR systems, was the cause of the lack of detection obtained with amplicons. Although in previous work it had been shown that on DNA microarray devices the distance between the fluorophore, used to label hybridization, and the target sequence affects the efficiency of detection,[22] in our case this phenomenon seemed not to be responsible for this behaviour, since the target sequence in our experiments had been designed at different distances from the fluorophore (Fig 6).

It may be speculated that the nature of the solid support and the presence of positive charges on PNA probes are more likely responsible for this effect, since they could promote the aggregation of the PNA on the surface, lowering the efficiency of the hybridization. Alternatively, long DNA tracts may aggregate on the positively charged surface, making the target DNA difficult to be bound by the PNA probes, and thus easily rinsed away during the washing steps.

Whatever might be the cause, it has already been demonstrated in the previous chapter that chiral PNA microarrays give less intense signals when hybridized to complementary oligonucleotides, as compared to standard PNAs, showing a higher specificity but a lower sensitivity in DNA detection. Given the promising potentialities of this kind of probe, future studies will have to be implemented in order to better understand the kind of interactions between the positively charged PNA probes, the microarray surface and long DNA tracts, which lowers the PNA binding efficiency.

5. Conclusions

The Arg-PNAs designed and synthesized in this work, tested both in solution and on microarray systems, confirmed the ability to perfectly discriminate sequences containing SNPs and thus their potentiality to be used to genotype tomato varieties by microarray approach. The microarray system here presented was able to simultaneously discriminate tomato varieties, except for 'Talent' and 'Tomito', in simulation experiments using oligonucleotide mixtures. Although the efficiency of DNA binding was somewhat lower than that usually shown by standard PNA arrays, causing a lower sensitivity towards long DNA sequences in this array system, the unsurpassed specificity of the PNA probes makes them very promising for the development of PNA-based genotyping methodologies. Further studies will be needed in order to better clarify the interactions occurring during positively charged PNA probes-long DNA tracts duplex formation on microarray surface in order to develop modified PNA-based systems able to compete in the future with the commercially available oligonucleotide-based recognition systems.

6. References

[1] Martinez I, Aursand M, Erikson U, Singstad TE, Veliyulin E, van der Zwaag C. Destructive and non destructive analytical techniques for authentication and composition analyses of foodstuffs. *Trends in Food Science & Technology*, 2003, **14**:489-498.

[2] Miller JC, Tanksley SD. RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theoretical and Applied Genetics*, 1990,**80**:437-448.
[3] Caramante M, Rao R,. Monti LM, Corrado G. Discrimination of 'San Marzano' accessions: a comparison of minisatellite, CAPS and SSR markers in relation to morphological traits. *Scientia Horticulturae*, 2009, **120**:560-564.
[4] Caramante M, Corrado G, Monti LM, Rao R. Simple sequence repeats are able to trace tomato cultivars in tomato food chains. *Food Control*, 2011, **22**:549-554.
[5] Rao R, Corrado G, Bianchi M, Di Mauro A. (GATA)4 DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. *Plant Breeding*, 2006, **125**:173-176.
[6] Chen J, Wanga H, Shen H, Chai M, Li J, Qi M, Yang W. Genetic variation in tomato populations from four breeding programs revealed by single nucleotide polymorphism and simple sequence repeat markers. *Scientia Horticulturae*, 2009, **122**:6–16.
[7] Jiménez-Gómez JM, Maloof JN. Sequence diversity in three tomato species: SNPs, markers, and molecular evolution. *BMC Plant Biology*, 2009, **9**:85-95.
[8] Ragoussis J. Genotyping technologies for genetic research. *Annual Review of Genomics and Human Genetics*, 2009, **10**:117-133.
[9] Kian-Kon J, Liu WT. Miniaturized platforms for the detection of single-nucleotide polymorphisms. *Analytical and Bioanalytical Chemistry*, 2006, **386**:427-434.
[10] Sassolas A, Leca-Bouvier BD, Blum LJ. DNA biosensors and microarrays. *Chemical Reviews*, 2008, **108**:109-139.
[11] Southern EM. DNA fingerprinting by hybridisation to oligonucleotide arrays. *Electrophoresis*, 1995, **16**:1539-1542.
[12] Teles FRR, Fonseca LP. Trends in DNA biosensors. *Talanta*, 2008, **77**:606-623.
[13] Sforza S, Corradini R, Tedeschi T, Marchelli R. Food analysis and food authentication by peptide nucleic acid (PNA)-based technologies. *Chemical Society Reviews*, 2011, **40**:221-232.
[14] Calabretta A, Tedeschi T, Di Cola G, Corradini R, Sforza S, Marchelli R. Arginine-based PNA microarrays for APOE genotypin. *Molecular BioSystem*, 2009, **5**:1323-1330.
[15] Manicardi A, Calabretta A, Bencivenni M, Tedeschi T, Sforza S, Marchelli R. Affinity and selectivity of C2- and C5-substituted "chiral-box" PNA in solution and on microarrays. *Chirality*, 2010, **22**:E161- E172.
[16] Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S. Bioinformatics methods and protocols: methods in molecular biology. *Humana Press, Totowa, NJ*, 2000, 365-386.
[17] Kalendar R, Lee D, Schulman AH. FastPCR software for PCR primer and probe design and repeat search. *Genes, Genomes and Genomics*, 2009, **3**:1-14.
[18] Yang W, Bai X, Kabelka E, Eaton C, Kamoun S, van der Knaap E, Francis DM. Discovery of single nucleotide polymorphisms in Lycopersicon esculentum by computer aided analysis of expressed sequence tags. *Molecular Breeding*, 2004, **14**:21-34.
[19] Tedeschi T, Calabretta A, Bencivenni M, Manicardi A, Corrado G, Caramante M, Corradini R, Rao R, Sforza S, Marchelli R. A PNA microarray for tomato genotyping. *Molecular BioSystems*, 2011, **7**:1902-1907.

[20] Andreakis N, Giordano I, Pentangelo A, Fogliano V, Graziani G, Monti LM, Rao R. DNA fingerprinting and quality traits of Corbarino cherry-like tomato landraces. *Journal of Agricultural and Food Chemistry*, 2004, **52**:3366-3371.

[21] Yang L, Pan A Jia J, Ding J, Chen J, Cheng H, Zhang C, Zhang D. Validation of a tomato-specific gene, lat52, used as an endogenous reference gene in qualitative and real-time quantitative PCR detection of transgenic tomatoes. *Journal of Agricultural and Food Chemistry*, 2005, **53**:183-190.

[22] Zhang L, Hurek T, Rienhold-Hurek B. Position of the fluorescent label is a crucial factor determining signal intensity in microarray hybridizations. *Nucleic Acids Research*, 2005, **33**:e166.

Chapter VIII

Tomato varieties identification through SNP detection by PNA-Mediated PCR Clamping

1. Introduction

Several techniques for the in vitro analysis of single base mutations in a genome have been published in literature. These include enzymatic,[1,2] or chemical,[3,4] probing of mismatch complexes, gradient gel electrophoresis,[5] use of nucleotide analogues hybridization,[6] with allele specific probes,[7] and oligonucleotide ligation assays.[8] All these methods usually require the target nucleic acid to be amplified by polymerase chain reaction (PCR) in order to enhance their sensitivity. PCR itself has also been used to analyze directly single base mutations by using allele specific oligonucleotides as amplification primers.[9,10] Unfortunately, the general applicability of this approach is limited by the fact that the majority of primer-template mismatches have no significant effect in inhibiting the amplification process.[11] To improve this situation, Ørum and co-workers developed a method that exploits PNAs to block the PCR amplification process in a sequence specific manner, if targeted against one of the PCR primer sites, given the higher thermal stability of a DNA/PNA duplex compared to the corresponding DNA/DNA duplex.[12] The specificity of this approach, called PNA-mediated PCR clamping, has proved to be such that two alleles which differ by only one base pair can be discriminated.[13] PNA clamping has therefore been introduced in the genotyping strategies as a way to elegantly modulate PCRs, for examples, by specific and efficient inhibition of fragments amplification, which differ only by a single base pair mutation.[14]

Different strategies have also been investigated in order to better understand whether the relative position of the PNA and PCR primer target sites could influence the PNA ability to clamp and modulate PCR.[13,15] As shown in Fig 1A, the method can operate by competition for a common target site between a PNA, complementary to the wild-type target sequence, and one of the PCR primers, complementary to the mutant target sequence, or vice versa. When the template contains the wild-type sequence, PNA binding will dominate over primer binding, due to the higher affinity of the matched PNA for the target site. As PNA cannot be extended by the Taq-polymerase the effect of this binding is that the amplification reaction is impaired. When the mutant sequence is present, PCR primer binding will dominate over PNA binding with the resulting generation of amplicons.[12] PCR clamping can also operate by interfering with primer elongation.[13] In one set-up the PNA is located adjacent to one of the PCR primers. Here, clamping is expected to operate by preventing initiation of primer elongation (Fig 1B). In another setup the PNA is located at a distance from the PCR primer. In this case, clamping is expected to operate by elongation arrest (Fig 1C).

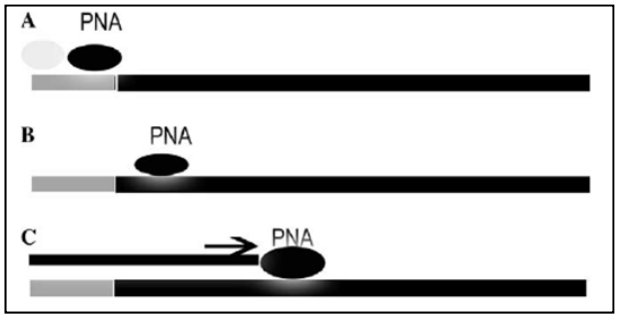


Fig 1. Models for PCR clamping: (A) primer exclusion; (B) inhibition of primer elongation by steric hindrance; (C) elongation arrest. Adapted from reference 15.

Clearly, for such block mechanisms to work the PNA must compete effectively against its cognate PCR primer in binding to their common recognition site or must bind robustly to its specific site in the elongation arrest model.[13] To facilitate this requirement, the normal 3rd step of PCR cycle is usually extended with a distinct PNA annealing step which precedes the PCR primer annealing step and which is at a temperature that allows only the fully complementary PNA to bind to its target sequence (Fig 2).

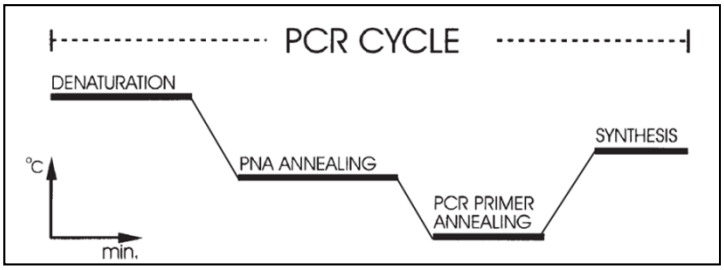


Fig 2. Schematic illustration of the four step PCR cycle profile used in PCR clamping. The PNA annealing step is set at a temperature that allows only the complementary PNA to bind. Adapted from reference 12.

Besides the mechanism, there are fundamental differences in the three described models, which can be exploited at best according to the type of mutations to be detected.[13] For example, PNA clamping protocol based on the primer exclusion principle does not require that the PNA, once bound to its target, is able to prevent read-through by the polymerase, as is the case in the elongation arrest clamping mode. Second, the only variables in the primer exclusion clamping mode are the T_m of the PNA and the PCR primers and these may be tuned to precision simply by changing either the sizes of the PNA and PCR primers, or by altering their exact position on the target DNA. Third, when using the primer exclusion principle there is the further advantage that, in addition to blocking its cognate target site, the PNA will compete with the PCR primer for any cryptic primer sites in the genome, thereby suppressing any occurrence of nonspecific background in the PCR process directed by this primer. This PNA clamping principle has been successfully exploited for the analysis of point mutations both in human,[14] and in microorganism.[16,17]

On the other side, the 'elongation arrest' model has been preferred for the enhancement of PCR amplification of variable number of tandem repeat (VNTR) loci.[18] In this approach the PNA molecules were designed to block the repeat sequences at a given locus during amplification, but at the same time they didn't prevent the polymerase from displacing them and extending the primers to completion. The presence of PNA blocking the repeat units prevented the re-annealing of complementary template strands occurring during the later rounds of amplification, thus reducing preferential amplification and enhancing the overall efficiency.

Despite all the reported examples, the inhibition operated by the primer extension and/or steric hindrance model was shown to be the most efficient condition in determining inhibition of amplification by PNA because it required the lowest concentration of PNA to determine a complete disappearance of bands in the screening of five main varieties of transgenics crops.[15] The use of PNA adjacent to a

primer, specific for the identification of a given event, and sharing few bases with it, was considered the optimal configuration also for the detection of mitochondrial DNA mutations involved in a variety of degenerative desease,[19] and to detect *K-ras* tumor-specific mutations.[20]
Unfortunately, the clamping process is a complex function of affinity and kinetics of dissociation and it is strongly influenced by the PNA concentration and melting temperature, as well as by the sequence to be clamped; for this reason no general rule can be inferred, but each experimentation may require several adjustments in order to reach the best results in terms of SNPs detection.
Despite several studies exploiting the PNA-mediated PCR clamping approaches for SNP detection have been already carried out, no one has been performed with chiral box PNAs, yet. The better performances of these modified PNAs in solution, in binding complementary sequences and in mismatch recognition,[21] could in fact improve the efficiency of the system in discriminating alleles which differ just for one base pair.

2. Aim of the work

In this chapter, the same argine-based 2D-chiral box PNAs (Arg-PNA) developed in the previous chapters were employed in the development of a PNA-mediated PCR clamping method for the recognition of tomato cultivars through their characteristic amplification patterns. Primer pairs were designed in order to compete with cognate PNA probes for binding to their common recognition site, thus blocking DNA amplification of genotypes carrying homozygous mutations. The strength and the limits of this technique will be illustrated.

3. Experimental part

3.1 Plant materials and food matrices

See paragraph 3.3, Chapter VII – Section II

3.2 DNA extraction from fresh and canned tomato

See paragraph 3.4, Chapter VII – Section II

3.3 Primer design

Primer pairs were designed on the tomato genomic sequences by Primer3 software v 0.9,[22] considering the standard features already describes elsewhere. Besides, the T_m for each primer pair was set to be few degrees lower than the T_m of PNA-DNA fullmatch duplex, to avoid PNA loss in binding specificity during primer annealing step. As regards the position, reverse primers were placed on the DNA sequences of interest in order to share two bases at 3' end with the H-terminus of PNA specific for the mutations to be detected (steric hindrance model) and, consequently, the position of forward primers was chosen, among all the ones suggested by the software, in order to minimize the length of the amplicons, making eventually more efficient even the amplification of DNA extracted from processed tomato. Selected PCR primers were then submitted to BLAST search against *Solanum lycopersicum* database in order to

avoid nonspecific annealing that could cause amplification of other non-target sequences besides the input template. The primer couples were further analyzed by using the on line program IDT – Integrated DNA Technologies,[23] to avoid primer pairing (all combinations including forward-reverse primer pair, forward-forward as well as reverse-reverse pairs).

3.4 PNA-mediated PCR clamping

3.4.1 Chemicals

- Phire® Hot Start II DNA Polymerase 5 U/µL (Finnzymes, Thermo Scientific, Finland)
- Phire Reaction Buffer 5 x containing 7.5 mM $MgCl_2$ (Finnzymes, Thermo Scientific, Finland)
- Deoxynucleotides 100 mM (dATP, dCTP, dGTP and dTTP) (10 mM solution for each dNTPs) (Euroclone, Italy)
- Oligonucleotide primers desalted and lyophilized (20 µM solution in sterile and apyrogenic H_2O) (Thermo Scientific, Italy)
- PNA probes diluted in sterile H_2O
- Sterile and apyrogenic H_2O (Salf SPA, Italy)

3.4.2 Instrumentation

- Aura PCR workstation (BioAir Instruments s.r.l., Italy)
- PCR Sprint Thermal Cycler (Thermo Hybaid, UK)

3.4.3 Procedure

All PCR procedures were performed in a final volume of 25 µL with the following reagent concentrations: genomic DNA 30 µg, PNA solution at different concentration diluted in sterile H_2O, Phire Reaction Buffer 1 x, $MgCl_2$ 1.5 mM, dNTPs 0.2 mM each, primer 0.4 µM each (forward and reverse), Phire® Hot Start II DNA Polymerase 0.15 U/µL. Thermal cycle conditions were as follows: pre-incubation at 94 °C for 5'; 30 cycles consisting of dsDNA denaturation at 94 °C for 30'', PNA annealing at specific T_a for 30'', primer annealing at specific T_a for 30'', primer extension at 62 °C for 30''; and final elongation at 62 °C for 5'.

3.5 Agarose gel electrophoresis

See paragraph 3.8, Chapter VII– Section II

4. Result and discussion

4.1 Choice of DNA sequences for tomato discrimination by PNA-mediated PCR clamping

Given the already characterized tomato SNPs, deeply described in the previous chapter, a careful study was carried out in order to select those ones able to discriminate among the seven tomato cultivars with the same shape by a PNA-clamping approach based on the steric hindrance model.

For this purpose, single nucleotide polymorphisms for all the tomato cultivars at each locus were analyzed to find the best combination of sequences whose amplification, in the presence of the specific PNA for the mutations to be detected, would have permitted the identification of tomato varieties by characteristic amplification patterns (Table 1).

Considering that the presence of an heterozygous genotype would not have completely blocked the amplification, since the dsDNA containing the nucleotide not recognized by the PNA probe would have been amplified anyway, three loci were chosen. As shown in Table 1, the amplification of the locus LeOH 8.4 in the presence of PNA specific for the recognition of 'C' mutation would allow to discriminate 'Cirio3', by blocking its amplification, from 'Ercole' and 'Talent', all of them long-shaped. The same for 'Tomito' and 'Minidor' cherry-shaped varieties.

Table 1. Genetic asset concerning SNPs for each tomato varieties and the expected amplification patterns in the presence of the specific PNA probes. '+': expected amplification; '-': non-expected amplification. The coloured boxes enclose varieties which are discriminated at each locus.

Loci \ Varieties	Leader (Rounded)	Ps1296 (Rounded)	Ercole (Elongated)	Talent (Elongated)	Cirio3 (Elongated)	Tomito (cherry)	Minidor (cherry)
PNA LeOH 8.4– C	C/T +	T/T +	C/T +	C/T +	C/C -	C/T +	C/C -
PNA LeOH 63– T	T/T -	C/T +	C/T +	C/T +	C/C +	C/T +	T/T -
PNA LeOH 23-del	G/del +	G/del +	G/del +	del/del -	G/G +	G/del +	G/G +

The full discrimination of 'Ercole' and 'Talent' varieties was expected to be accomplished at the locus LeOH 23, where the PNA specific for the deletion mutation could block the amplification of the homozygous genotype, distinguishing therefore between the two long-shaped varieties. Finally, the round-shaped tomato cultivars 'Leader' and 'PS1296' would present a discerning amplification pattern at the locus LeOH 63, considering that PNA recognizing 'T' mismatch could entirely block the amplification reaction for 'Leader' carrying homozygous mutation. In this way, by performing amplification of these loci in the presence of the specific PNAs for the mutations to be detected, a characteristic electrophoresis pattern for tomato varieties with the same fruit shape could be developed.

4.2 Primer design

Considering that a previous work described the steric hindrance strategy as the one requiring the lowest concentration of PNA to clamp the amplification,[15] and that we aimed to obtain a complete blocking of the PCR reaction, whenever the mutation was recognized by the specific PNA, primer pairs were designed in order to fulfil the requirements of this model. In particular, the reverse primers were positioned on the genomic sequence to be amplified so that it shared the last two bases at the 3' end with the first two of PNA N-terminus.

Besides, for primer design an accurate calibration of the T_m of PNA and relative competitive primer was carried out. The T_m of primer pairs was set to be lower than the one of PNA-DNA duplex fullmatch, observed in solution:[24] in this way, after the PNA hybridization step, the primer annealing step would not risk the PNA to

unbound from its site. Primer pair sequences, the relative T_m and the amplicon length were reported in Table 2 together with the specific PNA sequences to operate the clamping and their T_m.

Table 2. Sequences of the primers used for PCR clamping. The red nucleobases indicate the nucleobase specific for the SNP to be detected.

Locus	Primer Sequence (5'-3')	Primer T_m (°C)	Amplicon Lenght (bp)	PNA sequences H-NH$_2$	PNA T_m fullmatch (°C)
LeOH 8.4	GGGTATGATCTTTGTACTTG-for ATACTTTTCGAGCTAAGAAA-rev	45.28 44.07	98	AAAGGCCGA	60
LeOH 23	ATAGAGGAGTCCATTAGTCT-for TTACAACCGATTTCTTGCTT-rev	45.55 47.68	130	TTTT-CGGTGG	56
LeOH 63	ATAGCCAAGAATGGTCTGTG-for GGTCCATTTGTTGGTAGTTG-rev	49.32 48.78	170	TGTGTCAAAAG	56

Each primer pair was also checked for forward-reverse primer pair, forward-forward as well as reverse-reverse pairs combination and then submitted to BLAST search against *Solanum lycopersicum* database, to reduce the targeting of nonspecific sequences that could reduced the amplification efficiency.

4.3 PNA-mediated PCR clamping

As a start, the efficiency of the primers in amplifying the target DNA, extracted from fresh tomato fruits as previously reported, was tested by performing PCRs using each primer pair without any PNA probe.

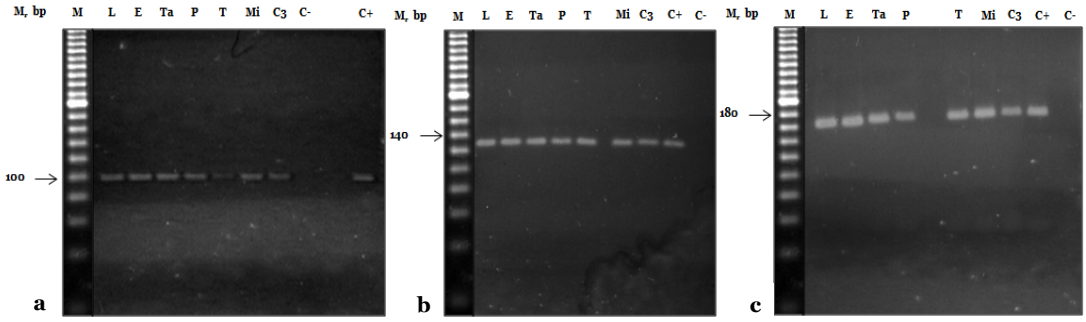


Fig 3. Single PCR amplification of the **a)** locus LeOH 8.4, **b)** locus LeOH 23.1 and **c)** locus LeOH 63, performed on DNA extracted from tomato fruits using specific primer pairs. 'L': Leader variety; 'E': Ercole variety; 'Ta': Talent variety; 'P': PS1296 variety; 'T': Tomito variety; 'Mi': Minidor variety; 'C3': Cirio3 variety; 'C-': water negative control; 'C+': positive control (tomato leaf DNA); 'M': molecular markers in exact 20 bp increments. The more intense one corresponds to a band at 200 bp

As shown in Fig 3, all the primer pairs gave only the expected amplicons, showing both specificity and efficiency in amplifying DNA sequences extracted from fresh tomato fruits. In order to allow the preferential annealing of PNAs to DNA, the T_a (annealing temperature) used for each PNA was meant to be lower of 4-5 °C than their T_m observed in solution in previous experiments (see Chapter VII - Section II). The T_a for each primer pair was therefore set up considering the T_a of PNAs. In general, primers T_a was dropped of about 10 degrees compared to the T_a of cognate PNAs, to avoid the destabilization of PNA-DNA duplex during the primer annealing

step. Considering this, the temperature of Taq DNA polymerase extension step was also lowered from 72 °C to 62 °C. Since PNA molecules are consumed during the amplification course, risking the loss of the efficiency in clamping process, the rounds of the PCRs were also reduced to 30 cycles.
Concerning the amount of PNA, several concentrations were tested to find the one able to clamp the amplification in a specific way.
Concerning the locus LeOH 8.4, 150 nm, 500 nm and 1 μM of specific PNA were used, by applying a PNA T_a of 55 °C and primer pairs T_a of 45 °C. The concentration of 1 μM seemed to work properly (Fig 4).

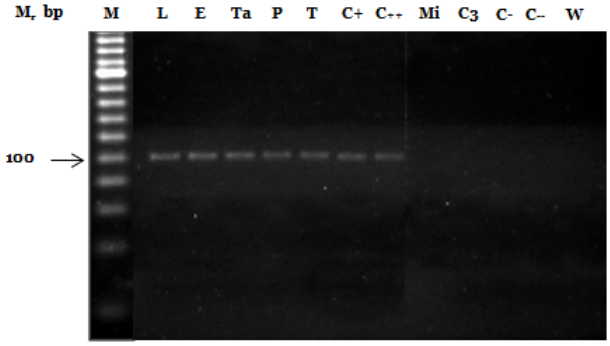


Fig 4. PNA mediated PCR clamping at the locus LeOH 8.4, using the PNA specific for the recognition of 'C' mismatch. 'L': Leader variety; 'E': Ercole variety; 'Ta': Talent variety; 'P': PS1296 variety; 'T': Tomito variety; 'Mi': Minidor variety; 'C3': Cirio3 variety; 'C+': positive control ('Cirio3' DNA fruit amplified without PNA); 'C++': 'Cirio3' DNA fruit amplified in the presence of a nonspecific PNA; 'C-': amplification without DNA with the specific PNA; 'C--': amplification without DNA with nonspecific PNA; 'W': water negative control; 'M': molecular markers in exact 20 bp increments. The more intense one corresponds to a band at 200 bp

Under these conditions, the amplified fragments of expected length were effectively observed only for the heterozygous genotypes and the homozygous ones, carrying the other genetic asset, while the homozygous ones for the specific mutation to be detected completely lacked of amplification, according to the scheme reported in Table 1.
Despite PNAs have been purified by RP-HPLC and their purity has been previously checked by LC-ESI-MS, an amplification test was carried out in the presence of a nonspecific PNA (the one recognizing the mismatch at locus LeOH 31.3, see Table 5, Chapter VII – Section II), in order to demonstrate that the absence of amplified fragments was completely due to the clamping process by the specific PNA and not to random PCR inhibitors derived from the PNA synthesis. As expected, the amplification was observed in the case of the homozygous genotype 'Cirio3' in the presence of the nonspecific PNA, confirming the efficiency and specificity of the clamping system PNA mediated.
According to the expected results showed in Table 1, the high selectivity and specificity in mismatch recognition of argine-based 2D-chiral box PNA fulfilled the modulation of PCR in the discrimination of long- and cherry-shaped tomatoes by characteristic gel electrophoresis patterns.
Unfortunately, the same application wasn't successful for the clamping of the other two tomato loci. In both the cases of the loci LeOH 23.1 and LeOH 63 several

adjustments were made to PNA concentration, primer pairs T_a, PNA T_a and times of annealing steps in order to encourage the PNA-DNA duplex formation and to fine regulate the modulation of PCR. In all the tested cases, the expected PNA-mediated clamping was not obtained, but a nonspecific halt of the amplification for all the genotypes was observed when the PNA concentration was higher than 10 μM. Besides the complex kinetics regulating the PNA-mediated PCR clamping and the influence of the target sequence in the achievement of the clamping process, it's likely that the lower T_m for the PNA-DNA fullmatch duplexes observed in solution for the PNA-LeOH 23.1-del and PNA-LeOH 63-T, compared to the T_m of PNA-LeOH 8.4-C,[24] may destabilize PNA-DNA duplex during the primer annealing step, thus resulting in the amplification of the containing SNP strand. Another trial experiment was also carried out by the 'elongation arrest' strategy, performed at the locus LeOH 63. In this case, the reverse primer was shifted downstream to the PNA position, to let the amplification started until it met the PNA-DNA duplex. Neither with this strategy any clamp of the homozygous genotypes was observed, confirming the difficulties in the setting up the whole system.

Concerning the behavior of PNA-mediated PCR clamping on DNA extracted from processed tomatoes, preliminary experiments showed that 30 cycles of repeated steps during the PCR were not enough to obtain an amount of dsDNA detectable by gel electrophoresis; on the other hand, the extension of cycle number from 30 to 40 decreased the efficiency of PNA in blocking the amplification reaction since, as already said, the higher the number of cycles, the higher is the PNA consuming and the efficiency loss in PCR clamping.

All these concomitances require further studies in order to better understand and clarify the dynamics occurring during the PNA-mediated PCR clamping for the accomplishment of single nucleotide polymorphism detection.

5. Conclusions

Chiral PNAs have been applied to identify point mutations in a PCR-based assay for the discrimination of tomato cultivars, exploiting their characteristic amplification profiles. The application relies on the increased affinity and specificity of 2D-chiral box PNA in mismatch recognition in conjunction with the fact that PNA molecules are not recognized and as such extended by DNA polymerases. Thus, PNA oligomers act as clamping molecules allowing a precise sequence-specific suppression of replication, when the mutation to be detected is present. The assay as such followed previous clamping protocols,[15,25] in which the 3'-DNA primer was designed to overlap with the sequence of the PNA probe, thus competing for their common binding site.

Despite PNA-mediated PCR clamping is an attractive tool for the detection of SNPs, it suffers from some limitations, such as complex kinetics ruling the whole process which are determined by several reaction parameters, such as the total number of PCR cycles, the primers and PNA annealing temperature and the amount of PNA added, as well as the target sequence.[12] The fine regulation of these parameters may be very time-consuming and not always fruitful; for this reason this technique is not fully widespread as a diagnostic tool, yet. Finally, the development of a PNA mediated PCR clamping assay in real time, measuring fluorescence at the crossing point rather

than evaluating band intensity, may allow a kinetic rather than a static assessment of the clamping reaction, giving a quantitative application of this technique.

6. References

[1] Till BJ, Burtner C, Comai L, Henikoff S. Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Research*, 2004, **8**:2632-2641.

[2] Goldrick MM. RNase cleavage-based methods for mutation/SNP detection, past and present. *Human Mutation*, 2001, **18**:190-204.

[3] Ikeda S, Kubota T, Kino K, Okamoto A. Sequence dependence of fluorescence emission and quenching of doubly thiazole orange labeled DNA: effective design of a hybridization-sensitive probe. *Bioconjugate Chemistry,* 2008, **19**:1719–1725.

[4] Abe H, Wang J, Furukawa K, Oki K, Uda M, Tsuneda S, Ito Y. A reduction-triggered fluorescence probe for sensing nucleic acids. *Bioconjugate Chemistry*, 2008, **19**:1219–1226.

[5] Knapp LA. Single nucleotide polymorphism screening with denaturing gradient gel electrophoresis. *Methods in Molecular Biology*, 2009, **578**:137-151.

[6] Peter M, Toftgaard Nielsen A, Pfundhelle HM, Yoanna C, Lars K, Søren M. Single nucleotide polymorphism genotyping using locked nucleic acid (LNA™). *Expert Review of Molecular Diagnostics*, 2003, **3**:27-38.

[7] Johnson MP, Haupt LM, Griffiths LR. Locked nucleic acid (LNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. *Nucleic Acids Research*, 2004, **32**:e55.

[8] Bruse SE, Moreau MP, Azaro MA, Zimmerman R, Brzustowicz LM. Improvements to bead-based oligonucleotide ligation SNP genotyping assays. *BioTechniques*, 2008, **45**:559-571.

[9] Kanga JH, Yua KH, Parka JY, Ana CM, Junb JC, Leea SJ. Allele-specific PCR genotyping of the *HSP70* gene polymorphism discriminating the green and red color variants sea cucumber (Apostichopus japonicus). *Journal of Genetics and Genomics*, 2011, **38**:351-355.

[10] Hirotsu N, Murakami N, Kashiwagi T, Ujiie K, Ishimaru K. Protocol: a simple gel-free method for SNP genotyping using allele-specific primers in rice and other plant species. *Plant Methods*, 2010, **6**:12.

[11] Kwok S, Kellogg DE, McKinney N, Spasic D, Godal L, Levenson C, Sninsky JJ. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*, 1990, **18**:999–1005.

[12] Ørum H. PCR Clamping. *Current Issues* in *Molecular Biology,* 2000, **2**:27-30.

[13] Ørum H, Nielsen PE, Egholm M, Berg RH, Buchardt O, Stanley C. Single base pair mutation analysis by PNA directed PCR clamping. *Nucleic Acids Research*, 1993, **21**:5332-5336.

[14] Kyger EM, Krevolin MD, Powell MJ. Detection of the hereditary hemochromatosis gene mutation by real-time fluorescence polymerase chain reaction and peptide nucleic acid clamping. *Analytical Biochemistry*, 1998, **260**:142–148.

[15] Peano C, Lesignoli F, Gulli M, Corradini R, Samson MC, Marchelli R, Marmiroli N. Development of a peptide nucleic acid polymerase chain reaction clamping assay for

semiquantitative evaluation of genetically modified organism content in food. *Analytical Biochemistry*, 2005, **344**:174–182.
[16] von Wintzingerode F, Landt O, Ehrlich A, Gobel UB. Peptide nucleic acid-mediated PCR clamping as a useful supplement in the determination of microbial diversity. *Applied And Environmental Microbiology*, 2000, **66**:549–557.
[17] Takiya T, Futo S, Tsuna M, Namimatsu T, Sakano T, Kawai K, Suzuki T. Identification of single base-pair mutation on uidA gene of Escherichia coli O157:H7 by peptide nucleic acids (PNA) mediated PCR clamping. *Bioscience, Biotechnology, and Biochemistry*, 2004, **68**:360-368.
[18] Demers DB, Curry ET, Egholm M, Sozer AC. Enhanced PCR amplification of VNTR locus D1S80 using peptide nucleic acid (PNA). *Nucleic Acids Research*, 1995, **23**:3050-3055.
[19] Murdock DG, Christacos NC, Wallace DC. The age-related accumulation of a mitochondrial DNA control region mutation in muscle, but not brain, detected by a sensitive PNA-directed PCR clamping based method. *Nucleic Acids Research*, 2000, **28**:4350-4355.
[20] Gilje B, Heikkila R, Oltedal S, Tjensvoll K, Nordgård O. High-fidelity DNA polymerase enhances the sensitivity of a peptide nucleic acid clamp PCR assay for K-ras mutations. *Journal of Molecular Diagnostics*, 2008, **10**:325–331.
[21] Manicardi A, Calabretta A, Bencivenni M, Tedeschi T, Sforza S, Marchelli R. Affinity and selectivity of C2- and C5-substituted "chiral-box" PNA in solution and on microarrays. *Chirality*, 2010, **22**:E161- E172.
[22] Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S. Bioinformatics methods and protocols: methods in molecular biology. *Humana Press, Totowa, NJ*, 2000, 365-386.
[23] http://eu.idtdna.com/ analyzer/ Applications/ OligoAnalyzer.
[24] Tedeschi T, Calabretta A, Bencivenni M, Manicardi A, Corrado G, Caramante M, Corradini R, Rao R, Sforza S, Marchelli R. A PNA microarray for tomato genotyping. *Molecular BioSystems*, 2011, **7**:1902-1907.
[25] Behn M, Schuermann M. Simple and reliable factor v genotyping by PNA-mediated PCR clamping. *Journal of Thrombosis and Haemostasis*, 1998, **79**:773–777.

Chapter IX - CONCLUSIONS AND PERSPECTIVES

The present Ph.D. thesis dealt with the use of advanced genomic and proteomic techniques for the assessment of quality and safety of one of the most important traditional and worldwide renowned Italian crop products: tomato.

In the first section, innovative proteomic approaches have been applied to the issue of tomato allergy. The increase of several cases of adverse reaction reported in the last few years, following the intake of both raw and cooked tomatoes,[1,2] emphasized the need to focus on the study of the immunological properties of tomato and on the characterization of its main allergens, by using reliable, specific, sensitive, time- and cost-effective methods.

Thereby in the third chapter, the classical immunological techniques, based on gel electrophoresis for the separation of proteins, along with advanced mass spectrometric systems, have been applied, allowing the detection and the characterization of the main allergens in a set of 12 tomato ecotypes. These experiments identified two allergenic proteins, profilin and suberization-associated anionic peroxidase, already known in the literature as tomato allergens. Moreover, they showed that the immunological response was serum-specific, depending both on the phenotype and on the tested sera, making it really difficult to make general statements on the potential hypoallergenicity of the analysed varieties. The literature is very poor concerning specific and exhaustive studies regarding tomato allergens, like the one described here.[3] Mostly, tomato allergen description has been carried out by coupling the classical immunological revelation methods with Edman degradation and searching public database for alignaments,[4] which limits the fine characterization of proteins. The study reported here, on the other hand, underlined the potential of the high-resolution mass spectrometric systems in the identification of allergenic proteins, being also able to univocally distinguish between isobaric peptides, thus confirming the appropriateness of the exploited method for this purpose.

In the fourth chapter, the characterization of an 'emergent' tomato allergen, namely nonspecific lipid transfer protein (LTP), has been described. Although some interesting attempts have already been carried out in order to characterize this allergen and its isoforms and to define their localization within fruit tissues,[5,6] they were affected by some factors, such as the use of tomato varieties not so common among consumers, which could thwart their generalization. For this reason, the characterization of tomato LTP isoforms, isolated from tomatoes purchased in the local markets, has been achieved by two main approaches. The first one relied on the purification of LTP in tomato peel extracts by using chromatographic systems working at high pressure (fast-protein LC) and ion exchange chromatography, followed by characterization of the obtained protein by LC/MS using a bottom-up approach and analysis of its resistance to gastrointestinal digestion. In this way, a new protein was identified, whose molecular weight didn't correspond to the one of other LTP isoforms already annotated in the database, but whose isoelectric point, secondary and tertiary structures and its resistance to gastrointestinal in vitro digestion, allowed speculation about the nature of this protein as a new LTP isoform. This is the first time, to the best of our knowledge, that a tomato LTP isoform has been so deeply characterized.

Alternatively, the occurrence and the localization of allergenic tomato nsLTP isoforms were investigated by the exploitation of a fast and simple method, based on the fractionation of proteins from peel, pulp and seed extracts, using ultracentrifuge

devices with molecular cut off able to retain proteins with 10 kDa, followed by characterization of retained proteins by LC/MS. This approach revealed the presence of four proteins only in the seed fraction, whose molecular weights didn't fit with any of the proteins already annotated, but whose chromatography patterns and the presence of four disulphide bridges clearly indicate their nature as LTPs.

The presence of so many tomato LTP isoforms has not been completely understood, yet. Since several activities have been detected for these proteins in vitro, it may be speculated that the plant system modulates the expression of such LTPs in order to tailor each of them to the proper function. On the other hand, the non simultaneous detection of these isoforms in the three main tissues of tomato fruits may be due to the sensitivity and effectiveness of the different analytical methods, as well as some intrinsic features of the varieties under examination.

Although further analysis will be necessary to investigate many other aspects related to tomato LTPs (the most important and yet largely unknown being the difference existing among different varieties), such as their immunological properties, the coexistence of so many LTP variants hampers their exploitation as a reliable tool in diagnosis and screening tests.

Despite such limitations, dealing with the nature of the allergen rather than the analysis itself, the high-resolution mass spectrometric techniques, together with classic and innovative purification strategies, have been shown to be a cornerstone for food safety evaluation.

In the second section of this thesis, advanced genomic techniques have been applied to the problem of tomato variety identification. Since one of the most important issues concerning food authenticity deals with the substitution of one species for another with a lower commercial value, or the mixing of one species with similar material from a cheaper species, the ability to identify these ingredients in any steps of the food chain represents the priority in the food traceability systems. Nowadays, DNA molecular marker-based techniques have the potential to greatly simplify methods for authentication. In particular, single nucleotide polymorphisms (SNPs) have showed to be very useful and reliable for detecting diversity among tomato varieties.[7]

Although many techniques have already been developed, in this study SNP genotyping has been carried out by means of specific probes, named Peptide Nucleic Acids (PNAs), which, due to their nature, possess enhanced properties in mismatch recognition compared to oligonucleotide probes.

Since several modifications have been introduced into PNA backbones, in order to improve their performance in terms of specificity and selectivity,[8,9,10] in the sixth chapter the effect of argine based 2D- and 5L- chiral-box PNAs were compared to the performance of the fully 'achiral' PNA for the first time, both in solution and on solid surfaces, in order to define the best PNA model for doing SNP recognition.

The data obtained showed that the 5L-chiral box PNA was superior in binding affinity when in solution, whereas the 2D-chiral box PNA model was superior in performances when recognition of single nucleotides was considered both in solution and in the microarray format. It may be speculated that the nature of this behaviour is likely due to the displacement of the chiral monomers in the PNA-DNA duplex three-dimensional structure and, therefore, to their interaction with the solid surface.

Information achieved here was then exploited in the design of a PNA-based microarray for the recognition of tomato specific SNPs.

So, in the seventh chapter, argine-based 2D-chiral box PNA microarray, for the simultaneous identification of SNPs characteristic of seven different tomato varieties, was designed and developed. The Arg-PNAs designed and synthesized, tested both in solution and on microarray systems, confirmed the ability to perfectly discriminate sequences containing SNPs and thus their potential to be used to genotype tomato varieties by the microarray approach. As already reported before for these modified probes,[11] the system perfectly worked in simulation experiments carried out by using oligonucleotide mixtures, mimicking the different sequences of the seven tomato varieties, and five out seven tomato varieties were univocally discriminated. On the other hand, the efficiency of DNA binding was much lower than that usually shown by achiral PNA arrays, causing a lower sensitivity towards long amplified DNA sequences in this array system. Despite the high potential of this technique mainly due to the outstanding specificity of such probes, further investigations will be necessary in order to elucidate the nature of the interactions occurring during positively charged PNA probes-long DNA tracts duplex formation on microarray surface and to clarify whenever these interactions can affect the efficiency of DNA binding, before it can be seriously considered as the standard method for PNA-based genotyping.

Finally in the eighth chapter, the selectivity and specificity of the same argine-based 2D-chiral box PNAs, in terms of tomato SNP recognition for variety discrimination, were investigated for the first time, to best of our knowledge, in a PNA-mediated PCR clamping system. The clamping of the amplification reaction of genotypes carrying homozygous mutations was achieved by means of primers designed to compete with cognate PNA probes for binding to their common recognition sites. Positive results were obtained for one of the three loci tested, which allowed the discrimination of two cherry- and of two long-shaped varieties. Although very promising for the detection of SNPs, this approach has poorly reliability, since it could be spoiled by many factors, including very complex kinetics, the right number of PCR cycles, the primers and PNA annealing temperature and the amount of PNA added, as well as the target sequences. All these factors hamper its wide diffusion as a diagnostic assay. Moreover, the use of modified PNA, for the first time since this technique has been introduced,[12] may require a finer regulation of the annealing temperatures both of the primers and PNAs, due to the high destabilization exercised by the presence of a mismatch, which could negatively affect the success of the clamping process.

Limitations aside, the data reported here confirmed the Arg-PNA ability to perfectly discriminate sequences containing SNPs and, thus, their potential to be used to genotype tomato varieties. It goes without saying that further investigations will be needed to better tailor the potential of this powerful tool to the different platforms available, in order to generate a reliable, efficient and fast method for variety identification.

In conclusion, it has been shown in the present Ph.D. thesis how the latest advancement in proteomic and genomic techniques can be exploited for assessing safety and quality of food products with unprecedented accuracy.

1. References

[1] Asero R, Mistrello G, Roncarolo D, Amato S, Arcidiacono R, Fortunato D. Detection of a novel allergen in raw tomato. *Journal of Investigational Allergology and Clinical Immunology*, 2008, **18**:397-400.

[2] Primavesi L, Pravettoni V, Brenna O.V, Farioli L, Pastorello EA, Pompei C. Influence of technological processing on the allergenicity of tomato products. *European Food Research and Technology*, 2011, **232**:631–636.

[3] Bässler OY, Weiss J, Wienkoop S, Lehmann K, Scheler C, Dölle S, Schwarz D, Franken P, George E, Worm M, Weckwerth W. Evidence for novel tomato seed allergens: IgE-reactive legumin and vicilin proteins identified by multidimensional protein fractionation–mass spectrometry and in silico epitope modeling. *Journal of Proteome Research*, 2009, **8**:1111–1122.

[4] Kondo Y, Urisu A, Tokuda R. Identification and characterization of the allergens in the tomato fruit. *International Archives of Allergy and Immunology*, 2001, **126**:294–299.

[5] Le LQ, Lorenz Y, Scheurer S, Fötisch K, Enrique E, Bartra J, Biemelt S, Vieths S, Sonnewald U. Design of tomato fruits with reduced allergenicity by dsRNAi-mediated inhibition of ns-LTP (Lyc e 3) expression. *Plant Biotechnology Journal*, 2006, **4**:231–242.

[6] Pravettoni V, Primavesi L, Farioli L, Brenna OV, Pompei C, Conti A, Scibilia J, Piantanida M, Mascheri A, Pastorello EA. Tomato allergy: detection of IgE-binding lipid transfer proteins in tomato derivatives and in fresh tomato peel, pulp, and seeds. *Journal of Agricultural and Food Chemistry*, 2009, **57**:10749–10754.

[7] van Deynze AE, Stoffel K, Buell RC, Kozik A, Liu J, van der Knaap E, Francis D. Diversity in conserved genes in tomato. *BMC Genomics*, 2007, **81**:465-472.

[8] Ganesh KN, Nielsen PE. Peptide nucleic acids analogs and derivatives. *Current Organic Chemistry*, 2000, **4**:931–943.

[9] Sforza S, Corradini R, Ghirardi S, Dossena A, Marchelli R. DNA binding of a D-lysine-based chiral PNA: direction control and mismatch recognition. *European Journal of Organic Chemistry*, 2000, **2000**:2905–2913.

[10] Menchise V, De Simone G, Tedeschi T, Corradini R, Sforza S, Marchelli R, Capasso D, Saviano M, Pedone C. Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA–DNA duplex. *Proceedings of the National Academy of Sciences USA*, 2003, **100**:12021–12026.

[11] Calabretta A, Tedeschi T, Di Cola G, Corradini R, Sforza S, Marchelli R. Arginine-based PNA microarrays for APOE genotyping. *Molecular BioSystem*, 2009, **5**:1323-1330.

[12] Ørum H. PCR Clamping. *Current Issues in Molecular Biology,* 2000, **2**:27-30.

Attachments

Affinity and Selectivity of C2- and C5-Substituted "Chiral-Box" PNA in Solution and on Microarrays

ALEX MANICARDI, ALESSANDRO CALABRETTA, MARIANGELA BENCIVENNI, TULLIA TEDESCHI, STEFANO SFORZA, ROBERTO CORRADINI,* AND ROSANGELA MARCHELLI

Dipartimento di Chimica Organica e Industriale Università di Parma, 43100 Parma, Italy

Contribution to the CD ISBC 2009 Special Issue

ABSTRACT Two peptide nucleic acids (PNAs) containing three adjacent modified chiral monomers (chiral box) were synthesized. The chiral monomers contained either a C2- or a C5-modified backbone, synthesized starting from D- and L-arginine, respectively (2D- and 5L-PNA). The C2-modified chiral PNA was synthesized using a submonomeric strategy to avoid epimerization during solid-phase synthesis, whereas for the C5-derivative, the monomers were first obtained and then used in solid-phase synthesis. The melting temperature of these PNA duplexes formed with the full-match or with single-mismatch DNA were measured both by UV and by CD spectroscopy and compared with the unmodified PNA. The 5L-chiral-box-PNA showed the highest T_m with full-match DNA, whereas the 2D-chiral-box-PNA showed the highest sequence selectivity. The PNA were spotted on microarray slides and then hybridized with Cy5-labeled full match and mismatched oligonucleotides. The results obtained showed a signal intensity in the order achiral >2D-chiral box >5L-chiral box, whereas the full-match/mismatch selectivity was higher for the 2D chiral box PNA. *Chirality 22:E161–E172, 2010.*

KEY WORDS: PNA; DNA recognition; chiral monomers; single nucleotide polymorphisms; microarrays

INTRODUCTION

New DNA binding molecules and new hybridization technologies are essential elements for studying gene-related processes,[1,2] and are currently used in several important applications, including the identification of genetic mutations or single-nucleotide polymorphisms (SNPs), medical diagnostics, gene delivery, assessment of gene expression, and drug discovery.[3–5] Heterogeneous formats for performing such assays have become increasingly available on account of the advancement of DNA microarray technologies,[6–8] which can be used, in combination with multiplex PCR, for the specific detection of specific DNA tracts of interest, such as those used in food analysis.[9–12]

Peptide nucleic acids (PNA, Fig. 1), achiral polyamidic oligonucleotide analogs have high affinity for RNA/DNA[13] and have been a model for designing new molecules mimicking the nucleic acid structures. Stereochemical issues are very important in PNA binding studies, since DNA is a highly asymmetric entity.[14]

The introduction of chirality through the use of amino acidic synthons have been proved to be an efficient tool for controlling the preference of the PNA for a given handedness. Chiral-modified PNA have been used in biological applications either in DNA detection[15,16] or as gene modulators.[17–19]

During the years, we have clarified the role of chirality in these compounds and have discovered some important properties of chiral PNA in terms of DNA affinity and sequence selectivity by using modified PNAs with stereogenic center either in the C2 or in the C5 position of the backbone or in both (Fig. 1).[20–22] The use of a single chiral monomer inserted in the middle of a PNA strand resulted in the modulation of the preferred PNA handedness, which, in turn, affected the preference for the right-handed complementary DNA.[23] In particular, the use of C2-modified monomers derived from D-amino acids and C5-modified monomers derived from L-amino acids were shown to induce a preferential right-handedness, with the latter being a stronger inducer than the former. As a result, C2,C5-modified PNAs were found to bind very effectively to DNA when the stereochemistry was 2D,5L (or 2R,5S).[22]

Abbreviations: AcOEt, ethyl acetate; AEEA, 2-(2-aminoethoxy)ethoxy acetic acid; Boc, *t*-butoxycarbonyl; BSA, *N,O*-Bis(trimethylsilyl)acetamide; DCC, *N,N*-dicyclohexylcarbodiimide; DCM, dichloromethane; DHBtOH, 3-hydroxy-1,2,3-benzotriazin-4(*3H*)-one; DIPEA, *N,N*-diisopropylethylamine; DMF, *N,N*-dimethylformamide; Fmoc, 9-fluorenylmethoxycarbonyl; HBTU, *O*-benzotriazol-1-yl-*N,N,N',N'*-tetramethyluronium hexafluorophosphate; HOBt, 1-hydroxy-1,2,3-benzotriazole; PNA, peptide nucleic acid; SSC, Saline sodium citrate; TFA, trifluoroacetic acid; TLC, thin layer chromatography.

Contract grant sponsor: Italian Ministry of University and Research (MUR) PRIN2007; Contract grant number: 2007F9TWKE_001
Contract grant sponsor: Italian Ministry of Agriculture (MIPAF; Project: GENOPOM)

*Correspondence to: Roberto Corradini, Dipartimento di Chimica Organica e Industriale, Università di Parma, Viale G. P. Usberti 17/A, 43100 Parma, Italy. E-mail: roberto.corradini@unipr.it

Received for publication 4 December 2009; Accepted 12 March 2010
DOI: 10.1002/chir.20865
Published online 14 May 2010 in Wiley Online Library (wileyonlinelibrary.com).

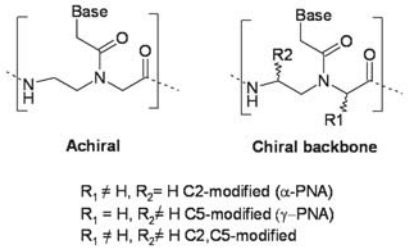


Fig. 1. General scheme of achiral PNA or C2-, C5-, and C2,C5-modified PNA.

High sequence selectivity is one of the properties for which PNAs are considered excellent probes.[24] However, in some applications, namely in the recognition of single base mismatch or single point mutation, the increase in sequence selectivity is highly desirable. Chiral PNAs have been used in combination with different technologies for the improved detection of single point mutations.[15,25]

One of the most efficient model was that obtained by introducing a series of three adjacent C2-modified chiral monomers, which we termed "chiral box": PNAs following this model were shown to have very high direction control (antiparallel vs. parallel binding) and exceptional sequence selectivity.[26–28] In this article, we compare the performances in solution and in microarray technology of modified PNAs bearing a "chiral box" composed of C2- or C5-modified monomers in terms of binding affinity and mismatch recognition.

EXPERIMENTAL

Reagents were purchased from Sigma-Aldrich, Fluka, Applied Biosystems, NovaBiochem and used without purification. DMF was dried over 4 Å molecular sieves and purged with nitrogen to avoid the presence of dimethylamine. THF was dried by distillation. TLC was run on Merck 5554 silica 60 aluminum sheets. Column chromatography was performed as flash chromatography on Merck 9385 silica 60 (0.040–0.063 mm). Reactions were carried out under a nitrogen atmosphere.

N^4-benzoxycarbonyl-N^1-carboxymethylcytosine (**CMC$^{(Z)}$**) and O^6-benzyl-N^9-carboxymethylguanine (**CMG$^{(OBn)}$**) were purchased from ASM (Hannover, Germany). Oligonucleotides, unlabeled or labeled with Cy5 fluorophore (HPLC grade), were purchased from Thermo Scientific and were used as provided.

NMR spectra were obtained on Brucker AC 300 or AVANCE 300 instruments. δ values are in ppm relative to $CDCl_3$ (7.29 ppm for proton and 76.9 ppm for carbon) or DMSO-d_6 (2.50 ppm for proton and 39.5 ppm for carbon). IR spectra were recorded with a FTIR Nicolet 5700. ESI-MS with a Micromass Quattro micro API (QqQ Detector). HR-MS with a Thermo LTQ ORBITRAP XL.

Synthesis

Boc-glycinal Boc-Gly-H (**1**),[23] D-Arg(Tos)-OMe·HCl (**2**),[25] and Boc-(Tos)argininal Boc-Arg(Tos)-H (**6**)[25] were synthesized as described previously.

Boc-2D-Arg(Tos)-PNA-backbone-OMe (3). Boc-Gly-H (**1**) (764.9 mg, 4.81 mmol) and Arg-OMe·HCl (**2**) (1.518 g, 4.01 mmol) were dissolved in MeOH (30 ml) with the addition of DIPEA (662 μl, 4.01 mmol). The solution was stirred for 30 min at room temperature, then was cooled to 0°C with an ice bath and $NaBH_3CN$ (302.1 mg, 4.81 mmol), CH_3COOH (275 μl, 4.81 mmol) were added to the stirred solution. The reaction was allowed to stir for 30 min at 0°C, then at room temperature for 2.5 h and was monitored via TLC. The solvent was evaporated and the residue was redissolved in 200 ml EtOAc and washed with saturated $NaHCO_3$ (200 ml, two times) and saturated aqueous $KHSO_4$ (200 ml). The organic layer was dried over Na_2SO_4, filtered, and evaporated to afford an oil. The oil was purified via column flash chromatography (from EtOAc to AcOEt/MeOH 9:1) to afford 1.32 g (68% yield) of the desired product as colorless foam.

(**3**). **^{1}H NMR** (300 MHz, DMSO-d_6): δ = 7.64 (d, 2H, *J* = 6 Hz, CH tosyl group), 7.29 (d, 2H, *J* = 6 Hz, CH tosyl group), 7.03 (br s, 1H, NH Arg side chain), 6.72 (br s, 1H, NH-Boc), 6.56 (br s, 1H, NH Arg side chain), 3.61 (s, 3H, OCH_3), 3.34 (s, 1H, CH Arg), 2.86–3.23 (m, 6H, CH_2NH), 2.35 (s, 3H, CH_3 tosyl group), 1.61–1.26 (m, 13H, CH_2 Arg side chain and CH_3 Boc group). **^{13}C NMR** (75.4 MHz, DMSO-d_6): δ = 174.1, 156.7, 156.6, 141.7, 141.0, 129.1, 125.9, 79.6, 60.6, 51.9, 47.9, 40.6, 29.6, 29.2, 28.3, 25.5, 21.4. **ESI-MS**: found 486.5 (MH^+); 508.5 (MNa^+); 524.4 (MK^+), calcd 486.6; 508.6; 524.7. **FTIR (KBr)**: ν (cm^{-1}) = 3421.4s, 3350.6s (N—H); 2976.5m (C—H); 1734.8s, 1701.0s (C=O); 1623.8m, 1576.8m (C—C aromatic); 1550.4s (N—H); 1366.5m (S—N); 1254.2s (C—N); 1168.6s (S=O).

Boc-2D-Arg(Tos)-PNA-backbone-OH (4). To a stirred solution of **3** (2.60 g, 5.37 mmol) in THF (50 ml), NaOH (2.14 g, 53.7 mmol in 50 ml of water) was added. The reaction mixture was stirred for 45 min and checked by TLC. The THF was then evaporated and the pH of the solution was lowered to 7.1 with HCl solution to induce the precipitation of the zwitterionic form. The solution was cooled at 4°C for 2 h, then filtered over Buchner and dried under vacuum to afford 2.44 g (97% yield) of the desired product as a white solid. (**4**). m.p.: 199.3°C (dec) **^{1}H NMR** (300 MHz, DMSO-d_6): δ = 7.97 (br s, 1H, NH Arg side chain), 7.64 (d, 2H, *J* = 9 Hz, CH tosyl group), 7.29 (d, 2H, *J* = 6 Hz, CH tosyl group), 6.98 (br s, 1H, NH Arg side chain), 3.29–3.05 (m, 5H, CH_2NH pseudopetide and CH Arg), 2.79–2.69 (m, 3H, CH_2NH Arg side chain), 2.34 (s, 3H, CH_3 tosyl group), 1.73–1.26 (m, 13H, CH_2 Arg side chain and CH_3 Boc group). **^{13}C NMR** (75.4 MHz, DMSO-d_6): δ = 170.8, 156.8, 155.5, 141.6, 140.8, 128.9, 125.4, 77.9, 61.1, 45.9, 37.0, 28.0, 27.2, 24.9, 20.7. **ESI-MS**: found 472.4 (MH^+); 494.4 (MNa^+); 510.4 (MK^+), calcd 472.6; 494.6; 510.7. **FTIR (KBr)**: ν (cm^{-1}) = 3421.4s, 3356.8s (N—H); 2975.9w (C—H); 1710.1s, 1689.5s (C=O); 1609.6s, 1577.7s

(C—C aromatic); 1539.7m (N—H); 1390.4w (S—N); 1261.3s (C—N); 1167.4s (S=O).

Boc-2D-Arg(Tos)-submonomer (5). Compound **4** (2.42 g, 5.00 mmol) was dispersed in DCM (170 ml), then BSA (3.7 ml, 15 mmol) and DIPEA (1.8 ml, 11 mmol) were added. The reaction mixture was stirred at room temperature for 30 min until the solution became nearly clear, then the Fmoc-Cl (1.55 g, 5.99 mmol) was added at 0°C. The mixture was stirred for 10 min at 0°C, then for 2 h at room temperature. The reaction was quenched by addition of MeOH (25 ml). The solvent was evaporated and the residue was redissolved in DCM (150 ml), washed with saturated $KHSO_4$ (150 ml, two times) and brine (150 ml). The organic layer was dried over Na_2SO_4, filtered and evaporated to afford an oil. The oil was purified via column flash chromatography (DCM to DCM/MeOH 9:1) to afford 2.85 mg (82% yield) of **5** as a pale yellow solid. (**5**). m.p. 134.6°C (dec.). **^{1}H NMR** (300 MHz, DMSO-d_6): δ = 7.92–7.82 (m, 2H, aromatic CH Fmoc group), 7.70–7.59 (m, 4H, aromatic CH tosyl and Fmoc group), 7.18–7.45 (m, 6H, CH tosyl and aromatic CH Fmoc group), 4.16–4.33 (m, 3H, *CHCH*$_2$ Fmoc group), 3.38 (br s, 1H, CH Arg), 3.14–2.87 (m, 2H, NH—*CH*$_2$*CH*$_2$—N backbone and *CH*$_2$NH Arg side chain), 2.30 (s, 3H, CH_3 tosyl group), 1.42–1.23 (m, 13H, CH_2 Arg side chain and CH_3 Boc group). **^{13}C NMR** (75.4 MHz, DMSO-d_6): δ = 175.0, 156.7, 155.9, 155.5, 143.8, 141.7, 140.8, 140.5, 128.8, 127.5, 127.1, 126.9, 125.4, 125.0, 77.3, 66.4, 60.8, 48.3, 46.6, 45.2, 29.9, 28.1, 26.6, 20.7. **FTIR (KBr)**: ν (cm^{-1}) = 3425.6s, 3346.8m (N—H); 2974.8w (C—H); 1689.5m, 1684.9m (C=O); 1617.4s (C—C aromatic); 1550.5s (N—H); 1367.1m (S—N); 1252.4m (C—N); 1165.8s (S=O). **HR-MS**: calcd m/z for $C_{35}H_{42}O_8N_5S$ ($M-H^-$): 692.2749, found: 692.2744.

Boc-5L-Arg(Tos)-PNA-backbone-OMe (7). Boc-Arg(Tos)-H (**6**) (686.7 mg, 1.66 mmol) and Gly-OMe·HCl (250.8 mg, 2.00 mmol) were dissolved in MeOH (20 ml). The reaction mixture was cooled to 0°C with an ice bath and $NaBH_3CN$ (132.1 mg, 2.00 mmol), and acetic acid (144 µl, 2.00 mmol) were added to the stirred solution. The reaction was allowed to stir for 30 min at 0°C, then at room temperature for 4 h and was monitored by TLC. The solvent was evaporated and the residue was redissolved in 200 ml EtOAc and washed with saturated $NaHCO_3$ (100 ml, three times) and saturated $KHSO_4$ (100 ml). The organic layer was dried over Na_2SO_4, filtered and evaporated to afford an oil. The oil was purified via column flash chromatography (using gradient elution from EtOAc to AcOEt/MeOH 9:1) to afford 406.7 mg of **7** (50% yield) as a colorless foam.[9] ^{1}H NMR (300 MHz, CDCl3): δ = 7.75 (d, 2H, *J* = 8.2 Hz, CH tosyl group), 7.22 (d, 2H, *J* = 8.2 Hz, CH tosyl group), 6.45 (br s, 3H, N—H), 5.00 (d, 1H, *J* = 8,4 Hz, NH-Boc), 3.72 (s, 3H, CH_3 methyl ester), 3.70–3.55 (m, 1H, CH Arg), 3.44 (d, 1H, *J* = 17.5 Hz, CH_2 Gly), 3.35 (d, 1H, *J* = 17.5 Hz, CH_2 Gly), 3.18 (br s, 2H, *CH*$_2$NH Arg side chain), 2.70–2.50 (m, 2H, CH_2 pseudopeptide moiety), 2.38 (s, 3H, CH_3 tosyl group), 1.78 (br s, 1H, NH amine), 1.61–1.43 (m, 4H, CH_2 Arg side chain),1.41 (s, 9H, CH_3 Boc group). **^{13}C NMR** (75.4 MHz, $CDCl_3$): δ = 172.8, 156.8, 156.2, 141.8, 140.6, 129.1, 125.7, 79.0, 53.5, 52.9, 51.6, 50.3, 40.7, 30.1, 28.2, 25.6, 21.2. **ESI-MS**: found 486.4 (MH^+), calcd 486.6. **FTIR (KBr)**: ν (cm^{-1}) = 3431.1m (N—H); 3342.1m (N—H); 3158.2w (Ar—H); 2976.3m, 2951.6m, 2866.2w (C—H); 1743.0s, 1695.8s, 1623.32s (C=O); 1576.4s, 1549.7s (N—H); 1456.8m, 1436.9m (—CH_2—); 1366.2m (S—N); 1253.6m (C—N); 1168.8s (C—O), 1132.3s (S=O).

Boc-5L-Arg(Tos)-PNA-C$^{(Z)}$-OMe monomer (8-C). N^4-benzoxycarbonyl-N^1-carboxymethylcytosine (**CMC$^{(Z)}$**) (502.9 mg, 1.65 mmol) was dissolved in DMF (10 ml) at 0°C, together with DHBtOH (268.8 mg, 1.65 mmol) and DIPEA (423 µl, 2.56 mmol). EDC·HCl (314.4 mg, 1.64 mmol) was then added, and the solution was stirred for 10 min at 0°C, then for 20 min at room temperature; the backbone **7** (408.2 mg, 0.84 mmol) was then added to the mixture, and the solution was stirred overnight. After completion of the reaction, the DMF was evaporated under vacuum. The residue was redissolved in AcOEt (100 ml) and washed with saturated $KHSO_4$ (50 ml, two times), saturated $NaHCO_3$ (50 ml, two times) and brine (50 ml). The organic layer was dried over Na_2SO_4 and filtered; the solvent was removed and the residue was purified via flash chromatography (AcOEt/MeOH = 95:5) to afford 615.4 mg (95% yield) of the product as a pale yellow foam.

(**8-C**). **^{1}H NMR** (300 MHz, $CDCl_3$): major rotamer δ = 10.81 (br s, 1H, NH Cytosine), 7.69 (d, 2H, *J* = 8.1 Hz, CH tosyl group), 7.55 (br s, 1H, cytosine H(6)), 7.35–7.10 (m, 8H, cytosine H(5), CH tosyl and benzyl group), 6.55 (br s, 2H, NH), 5.56 (br s, 1H, NH-Boc), 5.16 (s, 2H, CH_2 benzyl group), 4.90–4.45 (m, 2H, CO—*CH*$_2$—Cytosine), 4.40–3.85 (m, 2H, CH_2 Gly), 4.00–3.70 (m, 2H, CH_2 pseudo-peptide moiety), 3.59 (s, 3H, OCH_3), 3.41 (br s, 1H, CH Arg), 3.17 (br s, 2H, *CH*$_2$NH Arg side chain), 2.32 (s, 3H, CH_3 tosyl group), 1.80–1.30 (m, 13H, CH_3 Boc group and CH_2 Arg side chain). **^{13}C NMR** (75.4 MHz, $CDCl_3$): major rotamer δ = 169.5, 167.4, 163.3, 156.9, 156.2, 156.0, 152.5, 150.2, 141.8, 140.9, 135.0, 129.1, 128.6, 128.3, 128.2, 125.9, 95.6, 79.7, 77.2, 67.7, 52.8, 52.2, 50.9, 49.2, 48.9, 40.2, 28.8, 28.3, 25.6, 21.4. **ESI-MS**: found 408.3 (MNa_2^{2+}); 771.8 (MH^+); 793.8 (MNa^+); 809.9 (MK^+), calcd 408.4; 771.9; 793.9; 810.0. **FTIR (KBr)**: ν (cm^{-1}) = 3429.0m, 3338.6m (N—H); 3140.0w (C—H aromatic); 2976.0m (C—H); 1750.0s, 1664.9s, 1630.1s (C=O); 1561.9s (N—H); 1453.8m (—CH_2—); 1367.6s (S—N); 1213.8s (C—N); 1132.3m (S=O).

Boc-5L-Arg(Tos)-PNA-C$^{(Z)}$-OH monomer (9-C). To a stirred solution of **8-C** (594.3 g, 0.77 mmol) in THF (40 ml), a solution of $Ba(OH)_2{\cdot}8H_2O$ (363.9 mg, 1.15 mmol) in water (40 ml) was added. The reaction mixture was stirred for 30 min. The THF was then evaporated and the pH of the solution was lowered to 5 with a HCl solution to induce the precipitation of the product. The solution was cooled at 4°C for 2 h, then filtered and dried under vacuum to afford 418.9 mg (72% yield) of the desired product as a white solid.

(**9-C**). m.p. 147.6°C (dec.). **^{1}H NMR** (300 MHz, DMSO-d_6): major rotamer δ = 10.81 (br s, 1H,

Cytosine—*NH*—COO—), 7.81 (d, 1H, J = 7.2 Hz, Cytosine H(6)), 7.64 (d, 2H, J = 8.1 Hz, CH tosyl group), 7.45–7.25 (m, 7H, CH tosyl and benzyl group), 7.02 (d, 1H, J = 7.2 Hz, Cytosine H(5)), 6.90–6.60 (br m, 3H, NH Arg side chain), 5.19 (s, 2H, CH_2 benzyl group), 4.90 (d, 1H, J = 15.9 Hz, CO—*CH_2*—Cytosine), 4.77 (d, 1H, J = 16.2 Hz, CO—*CH_2*—Cytosine), 4.03 (d, 1H, J = 17.1 Hz, CH_2 Gly), 3.87 (d, 1H, J = 17.4 Hz, CH_2 Gly), 3.70–2.90 (m, 5H, —*CH*—*CH_2*—, *CH_2*NH Arg side chain and water), 2.34 (s, 3H, CH_3 tosyl group), 1.60–1.30 (m, 13H, CH_3 Boc group and CH_2 Arg side chain). **^{13}C NMR** (75.4 MHz, DMSO-d_6): major rotamer δ = 170.3, 167.1, 163.0, 156.5, 155.5, 154.9, 153.1, 150.5, 141.6, 141.0, 135.9, 129.0, 128.4, 128.1, 127.9, 125.5, 93.8, 77.9, 66.4, 51.4, 49.4, 48.5, 47.8, 40.2, 28.6, 28.1, 25.4, 20.8. **FTIR (KBr)**: ν (cm^{-1}) = 3447.0m (N—H); 3350.3m (O—H) 2977.8w, 2933.5w (C—H); 1748.3m, 1653.4s, 1635.5s (C=O); 1558.7s (N—H); 1457.1m (—CH_2—); 1368.6m (S—N); 1214.7s (C—N); 1132.3m (S=O). **HR-MS**: calcd m/z for $C_{34}H_{45}O_{10}N_8S$ (MH^+): 757.2974, found: 757.3019; calcd m/z for $C_{34}H_{44}O_8N_{10}NaS$ (MNa^+): 779.2793, found: 779.2831; calcd m/z for $C_{34}H_{43}O_{10}N_8S$ ($M-H^-$): 755.2817, found: 755.2819.

Boc-5L-Arg(Tos)-PNA-G$^{(OBn)}$-OMe monomer (8-G). O^6-benzyl-N^2-carboxymethylguanine (**CMG$^{(OBn)}$**) (501.3 mg, 1.68 mmol) was dissolved in DMF (6 ml) at 0°C, together with DHBtOH (273.2 mg, 1.68 mmol) and DIPEA (416 μl, 2.51 mmol). EDC·HCl (318.2 mg, 1.66 mmol) was then added and the solution was stirred for 10 min at 0°C and for 20 min at room temperature, then a solution of the backbone **7** (406.7 mg, 0.84 mmol) in DMF (6 ml) was added to the mixture. The solution was stirred overnight and then DMF was evaporated under vacuum. The residue was redissolved in AcOEt (200 ml) and washed with saturated $NaHCO_3$ (200 ml, two times), saturated $KHSO_4$ (200 ml, two times) and brine (200 ml). The organic layer was dried over Na_2SO_4 and filtered; the solvent was removed and the residue was purified via flash chromatography (AcOEt/MeOH = 95:5) to afford 511.4 mg (79% yield) of the product as a pale yellow foam.

(**8-G**). **^{1}H NMR** (300 MHz, $CDCl_3$): major rotamer δ = 7.70–7.60 (m, 3H, CH tosyl group and CH guanine), 7.39 (d, 2H, J = 6.0 Hz, CH benzyl group), 7.30–7.20 (m, 3H, CH benzyl group), 7.13 (d, 2H, J = 7.8 Hz, CH tosyl group), 6.50–6.20 (br m, 5H, NH_2 Guanine and NH Arg side chain), 5.73 (br s, 1H, NH-Boc), 5.45 (s, 2H, CH_2 benzyl group), 4.97 (d, 1H, J = 15.9 Hz, CO—CH_2—Guanine), 4.84 (d, 1H, J = 15.9 Hz, CO—CH_2—Guanine), 4.30–4.20 (m, 2H, CH_2 Gly) 3.90–3.70 (m, 1H, CH Arg), 3.61 (s, 3H, OCH_3), 3.40–3.00 (m, 4H, CH_2 pseudopetide moiety and *CH_2*NH Arg side chain), 2.30 (s, 3H, CH_3 tosyl group), 1.40–1.15 (m, 13H, CH_2 Arg side chain and CH_3 Boc group). **^{13}C NMR** (75.4 MHz, $CDCl_3$): major rotamer δ = 169.5, 167.4, 160.9, 159.5, 156.1 142.1, 142.0, 140.6, 140.4, 136.3, 129.2, 128.4, 128.1, 125.9, 114.5, 79.9, 68.1, 52.9, 52.7, 52.4, 51.5, 50.2, 48.8, 40.6, 28.8, 28.4, 25.7, 21.4. **ESI-MS**: found 767.8 (MH^+); 789.7 (MNa^+), calcd 767.9; 789.9. **FTIR (KBr)**: ν (cm^{-1}) = 3351.7m (N—H); 3156.8w (Ar—H); 2975.6m, 2952.4m (C—H); 1745.5m, 1665.7m, 1615.5s (C=O); 1584.4s, 1551.5s (N—H); 1456.6m (—CH_2—); 1365.9m (S—N); 1256.3s (C—N); 1132.3m (S=O).

Boc-5L-Arg(Tos)-PNA-G$^{(OBn)}$-OH monomer (9-G). To a stirred solution of **8-G** (511.4 g, 0.67 mmol) in THF (40 ml), a solution of $Ba(OH)_2{\cdot}8H_2O$ (314.7 mg, 1.00 mmol) in water (40 ml) was added The reaction mixture was stirred for 15 min and checked by TLC. The THF was then evaporated and the pH of the solution was adjusted to 5 with a solution of HCl to induce the precipitation of the product. The solution was cooled at 4°C for 2 h, then filtered and dried under vacuum to afford 461.1 mg (92% yield) of the desired product as a white solid.

(**9-G**). m.p.: 139.4°C (dec.). **^{1}H NMR** (300 MHz, DMSO-d_6): major rotamer δ = 7.67 (s, 1H, C_8H guanine), 7.62 (d, 2H, J = 8.1 Hz, CH tosyl group), 7.50 (d, 2H, J = 7.8 Hz, CH benzyl group), 7.45–7.30 (m, 3H, CH benzyl group), 7.26 (d, 2H, J = 7.8 Hz, CH tosyl group), 6.99 (br s, 3H, NH Arg side chain), 6.73 (br d, 1H, J = 7.5 Hz, NH—Boc), 6.44 (br s, 2H NH_2 Guanine), 5.50 (s, 2H, CH_2 benzyl group), 4.84 (s, 2H, CO—*CH_2*—Guanine), 3.95–3.55 (m, 2H, CH_2 Gly), 3.50–2.90 (m, 5H, —*CHCH*$_2$—, *CH_2*NH Arg side chain and water), 2.31 (s, 3H, CH_3 tosyl group), 1.50–1.15 (m, 13H, CH_2 Arg side chain and CH_3 Boc group). **^{13}C NMR** (75.4 MHz, DMSO-d_6): major rotamer δ = 172.6, 167.5, 159.9, 159.6, 156.8, 154.9, 141.7, 140.9, 140.7, 140.4 136.7, 129.0, 128.3, 127.9, 125.5, 113.0, 77.5, 66.7, 52.9, 51.7, 49.5, 48.5, 43.5, 40.6, 35.7, 28.7, 28.2, 25.3, 20.8. **FTIR (KBr)**: ν (cm^{-1}) = 3348.8m (N—H); 3219.5m (O—H); 2976.2w (C—H); 1721.4m, 1685.4s, 1615.8s (C=O); 1588.0s (N—H); 1456.1m (—CH_2—); 1366.1m (S—N); 1255.3m (C—N); 1166.6m (S=O). **HR-MS**: calcd m/z for $C_{34}H_{45}O_8N_{10}S$ (MH^+): 753.3137, found: 753.3157; calcd m/z for $C_{34}H_{44}O_8N_{10}NaS$ (MNa^+): 775.2957, found: 775.2967; calcd m/z for $C_{34}H_{43}O_8N_{10}S$ ($M-H^-$): 751.2981, found: 751.3008.

PNA synthesis. The synthesis of the achiral **PNA-1** was performed on an ABI 433A peptide synthesizer with a software modified to run the PNA synthetic steps (scale: 5 μmol), using Fmoc-based chemistry and standard protocols, with HBTU/DIPEA coupling and a Rink amide, loaded with Fmoc-PNA-C(Bhoc)-OH as first monomer.

The 2D chiral box **PNA-2** was synthesized using standard Boc-based manual peptide solid-phase synthesis for the introduction of achiral monomers and with submonomeric synthesis for the "chiral box" part, as previously described for 2D-Lys-based PNA.[29]

The 5L chiral box **PNA-3** was synthesized with standard manual Boc-based solid-phase synthesis using MBHA resin and the preformed monomers **9-G** and **9-C**, in addition to commercially available Boc-PNA monomers.

All crude PNA were purified by RP-HPLC with UV detection at 260 nm using a semipreparative column C18 (5 μm, 250 × 10 mm, Jupiter Phenomenex, 300 A) Eluents were as follows: water containing 0.1% TFA (eluent A) and a 50:50 water/acetonitrile mixture containing 0.1% TFA (eluent B); elution gradient: from 100% A to 100% B in 30

min, flow: 4 ml/min. The resulting pure PNA oligomers were characterized by MS-ESI.

Achiral PNA-1 (H-AEEA-AEEA-CTTATCCGGTGCC-NH_2). Yield (after purification): 12%, Calculated MW: 3777.7 ESI-MS: m/z = found 945.1 (calc 945.4, MH_4^{4+}); found 756.2 (calc 756.5; MH_5^{5+}), found 630.4 (calc 630.6; MH_6^{6+}), found 540.6 (calc 540.7; MH_7^{7+}). Residues of compounds lacking of 1 or 2 AEEA spacers were present in the final product.

2D Chiral-Box PNA-2 : H-AEEA-AEEA-CTTATC$_{2DArg}$C$_{2DArg}$G$_{2DArg}$GTGCC-NH_2

Yield (after purification): 13%. Calculated MW: 4075.1. ESI-MS: m/z = found 1019.5 (calc 1019.8, MH_4^{4+}); found 816.0 (calc 816.0; MH_5^{5+}), found 680.1 (calc 680.2; MH_6^{6+}), found 583.1 (calc 583.2; MH_7^{7+}); found 510.3 (calc 510.4; MH_8^{8+}). Residues of compounds lacking of 1 or 2 AEEA spacers were present in the final product.

5L Chiral-Box PNA-3. H-AEEA-AEEA-CTTATC$_{5LArg}$C$_{5LArg}$G$_{5LArg}$GTGCC-NH_2

Yield (after purification): 16%, Calculated MW: 4075.1. ESI-MS: m/z = found 1019.5 (calc 1019.8, MH_4^{4+}); found 815.7 (calc 816.0; MH_5^{5+}), found 680.0 (calc 680.2; MH_6^{6+}), found 583.0 (calc 583.2; MH_7^{7+}); found 510.4 (calc 510.4; MH_8^{8+}). Residues of compounds lacking of 1 or 2 AEEA spacers were present in the final product.

UV and CD analysis. Stock solutions of **PNA-(1-3)**, and of **DNA oligonucleotides (DNA-FM** and **DNA-MM)** were prepared in double distilled water and their actual concentration calculated by UV absorbance using the following ϵ_{260} (M^{-1} cm^{-1}) for the nucleobases: T 8600, C 6600, A 13700, G 11700. By using these concentrations, hybrid solutions containing PNA or PNA:DNA duplexes were prepared. Solutions of 1:1 DNA/PNA were prepared in a buffer consisting of 100 mM NaCl, 10 mM $NaH_2PO_4{\cdot}H_2O$, 0.1 mM EDTA (pH = 7.0). Strand concentrations were 5 μM for each component. All the hybrid samples reported were first incubated at 90°C for 5 min and then slowly cooled to room temperature. Thermal denaturation profiles (Abs vs. T) of the hybrids were measured at 260 nm with an UV/Vis Lambda Bio 20 Spectrophotometer equipped with a Peltier Temperature Programmer PTP6 interfaced to a personal computer. For the temperature range 90 to 18°C, UV absorbance was recorded at 260 nm every 0.1°C. A melting curve was recorded for each duplex. The melting temperature (T_m) was determined from the maximum of the first derivative of the melting curves. CD measurements were carried out on a Jasco (Tokyo, Japan) J715 Spectropolarimeter equipped with a Peltier PTC 348 temperature controller unit.

Array preparation. "Genorama® SAL Slides" (Asper Biotech Ltd.) were used as solid support to which the amino-terminal group of the PNA probes was covalently linked. The deposition of the probes was carried out using a SpotArray™ 24 Microarray Printing System (PerkinElmer™ Life Sciences, Inc). The manufacturer's instructions for the deposition protocol were slightly changed to comply with the special requirement of the chemical structures of PNAs: in particular a 100 mM carbonate buffer (pH 9.0) containing 0.001% sodium dodecyl sulfate (SDS) was used as deposition buffer. Moreover, after every deposition, the pin printing system was washed in three solutions containing 100 mM carbonate buffer (pH 9.0) and SDS at different concentrations (0.2%, 1%, 0.001%), to avoid carry-over of the probes in subsequent depositions. The probes were coupled to the surface by leaving the slides in a humid chamber (relative humidity 75%) at room temperature for 12 h, and the remaining reactive sites were blocked by immersion in a 1% aqueous solution of NH_3 at room temperature for 15 min under gently shaking. The slides were slowly shaken for 15 min with 0.1% SDS buffer prewarmed at 40°C and then washed twice with doubly distilled water at room temperature. Each slide was spin-dried twice in a plastic tube at 1200 rpm for 5 min, to remove the remaining washing solution. Slides were then ready to undergo the hybridization protocol or could be stored under dried atmosphere for future use. Since a fluorescent control probe ($(AC)_{11}$-Cy5)) was deposited to check the efficiency of the deposition step, all the previously described operations were carried out away from direct light in order to prevent degradation of the Cy5 fluorophore.

Hybridization on microarray. The Cy5-labeled **DNA-FM** and **DNA-MM** oligonucleotides were used at a final concentration of 0.1 μM in 2× saline/sodium citrate (SSC) solution and 0.1% SDS buffer. Hybridization was performed by loading 65 μl of each solution to multiwell chambers (Arrayit Corporation, AHC1 × 16), using a silicone gasket (Arrayit Corporation, GAHC1 × 16), to avoid cross contamination, and leaving the slides under slow shaking for 2.5 h at 40°C. After the hybridization step, the slides were washed under slow shaking for 5 min at 40°C with a 2× SSC, 0.1% SDS buffer prewarmed at 40°C, followed by treatment for 1 min with 0.2× SSC and for 1 min with 0.1× SSC at room temperature. The slides were then spin-dried twice at 1200 rpm for 5 min. All posthybridization steps were performed in a dark environment to prevent degradation of the Cy5 fluorophore used to label the oligonucleotides.

Image acquisition. The fluorescent signal deriving from the hybridization was acquired using a ScanArray™ Express Microarray Scanner (PerkinElmer™ Life Sciences, Inc) at λ_{ex} = 646 nm and λ_{em} = 664 nm. To correctly compare the hybridization data, all the images reported were acquired with laser power = 100 and photomultiplier gain = 70. Images were analyzed using the ScanArray program with integration of 170 μm diameter circular area entirely containing the fluorescent spots.

RESULTS AND DISCUSSION

Chiral PNA Design and Synthesis

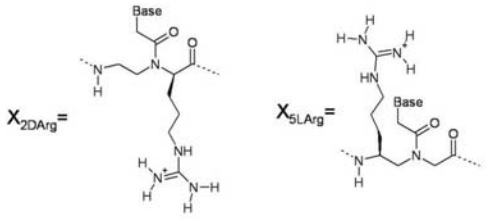
Recent studies carried out in our laboratory have shown that substitution of the achiral *N*-(2-aminoethyl)amino acid backbone with chiral units derived from basic amino acid

H-AEEA-AEEA-CTTATCCGGTGCC-NH_2 Achiral **PNA-1**

H-AEEA-AEEA-CTTAT$C_{2DArg}C_{2DArg}G_{2DArg}$GTGCC-$NH_2$ 2D Chiral box **PNA-2**

H-AEEA-AEEA-CTTAT$C_{5LArg}C_{5LArg}G_{5LArg}$GTGCC-$NH_2$ 5L Chiral box **PNA-3**

"Chiral box"



Full-match DNA-FM: 5'-GGCACCGGATAAG-3'

Mismatch DNA-MM: 5'-GGCACCAGATAAG-3'

Fig. 2. Achiral **PNA-1**, 2D chiral-box **PNA-2**, 5L chiral-box **PNA-3**, and DNA full matched (**DNA-FM**) and mismatched (**DNA-MM**) sequences used in this study. AEEA refers to the 2-(2-aminoetoxy)etoxyacetyl spacer.

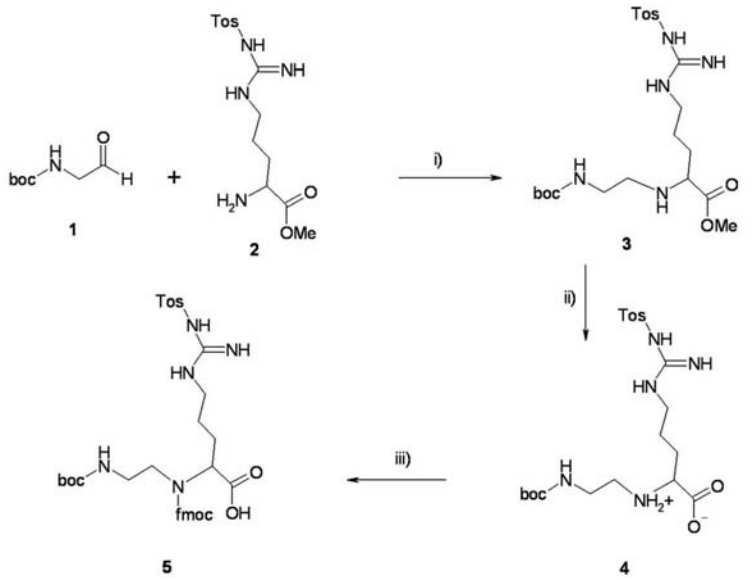


Fig. 3. Synthesis of the chiral 2D-Arg PNA submonomer **5**: (i) $NaBH_3CN$, CH_3COOH in MeOH; η: **68%**; (ii) NaOH in water; η: **97%**; (iii) BSA, DIPEA Fmoc-Cl in DCM; η: **82%**.

Chirality DOI 10.1002/chir

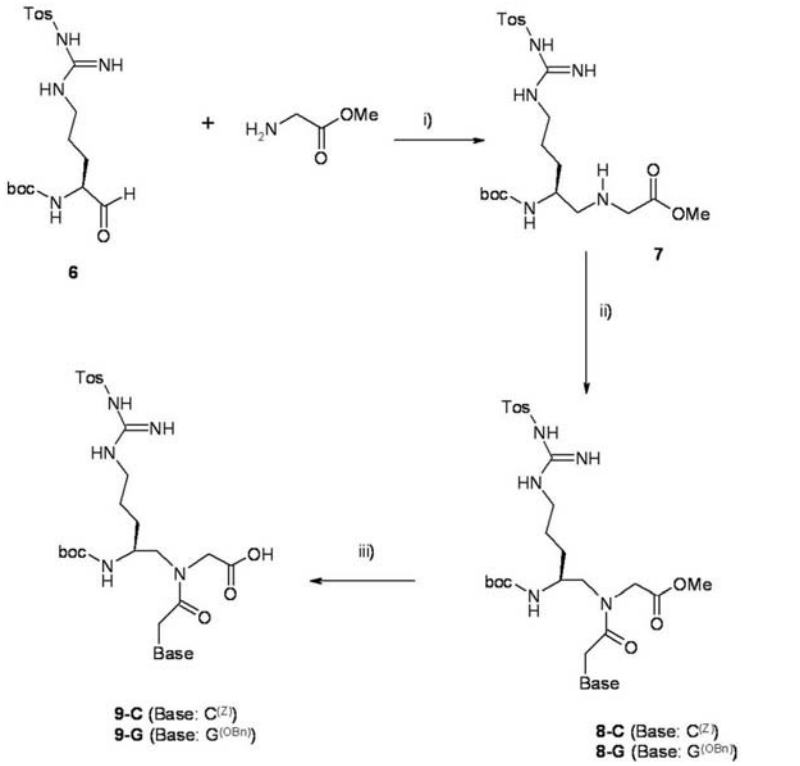


Fig. 4. Synthesis of the 5L-Arg chiral PNA monomers **9**: (i) $NaBH_3CN$, CH_3COOH in MeOH; η: **50%**; (ii) $CMC^{(Z)}$ or $CMG^{(OBn)}$ DCC/DHBtOH in DMF; η: **79–95%**; (iii) $Ba(OH)_2$ in THF/H_2O = 1: 1; η: **72–92%**.

(thus bearing a positive side chain) can positively affect the DNA recognition ability depending on the position and the stereochemistry of the stereogenic centers.[22]

We chose D- or L-arginine as a starting synthon since this unit allows to place a positive charge in the backbone giving rise to electrostatic attractive interactions with the negatively charged DNA. The guanidinium group of Arg was chosen not to compete with terminal amino group of the PNA as a nucleophile in reactions such as those required for the covalent linking of the PNA onto an activated solid surface for microarray fabrication. We chose as a model system the two PNA reported in Figure 2, with the monomers centered on the base to be recognized, composed of either C2- or C5-modified arginine monomers. The full match and the mismatch DNA sequence chosen correspond to target sequences of interest in the recognition of vegetal varieties.

The achiral **PNA-1** was synthesized using solid-phase synthesis on an automatic ABI 433A Synthesizer. The Fmoc strategy was used with HBTU/DIPEA as coupling agent and piperidine/NMP as deprotecting solution.

C-2 substituted chiral PNA monomers are sensitive to racemization,[30,31] and should therefore be synthesized using a special protocol called "submonomeric approach" in which only the submonomer (Fig. 3) lacking of the carboxymethyl base is linked to the peptide chain, followed by deprotection of the α-amino group and coupling with the desired carboxymethyl-nucleobase unit. The synthesis of the C2-modified D-Arg submonomer is depicted in Figure 3. The tosyl protected Arg methyl ester **2** was coupled

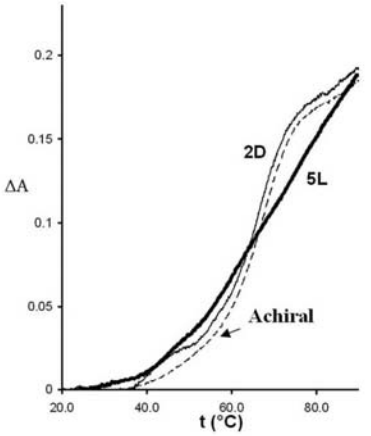


Fig. 5. UV Melting curves (260 nm) for the **DNA-FM** with 2D **PNA-2** (thin line), 5L **PNA-3** (thick line) and achiral **PNA-1** (dashed line), in PBS buffer, pH 7; concentration of each strand was 5 μM. Data are expressed as variation in absorbance (Δ*A*). Cell length was 1 cm.

with Boc-2-aminoacetaldehyde **1** under reductive amination conditions to obtain the C-protected backbone **3**. Deprotection of the methyl ester gave the backbone **4**, and subsequent protection of the α-amino group with Fmoc-chloride in the presence of a temporary protecting silylating agent (BSA) led to the final C2-D-Arg submonomer **5**.

The synthesis of 2D chiral box **PNA-2** (Fig. 2) was carried out using the submonomer synthesis previously described for other chiral-box PNAs.[21,29]

The use of C5-substituted PNA monomers prevents the epimerization problems; therefore, the protected PNA monomer can be preformed and then directly used for the solid-phase synthesis.

Thus, we designed two C5-substituted Boc-PNA monomers by using L-Arg(Tos) synthons, following the synthetic scheme previously described for other amino acids[15]: synthesis of α-substituted Boc-aminoacetaldehyde **6** (through reduction of the corresponding Weinreb amide), which was subsequently used in reductive amination with glycine methyl ester to yield the backbone methyl ester **7**, followed by reaction with the carboxymethylbase to yield the C-protected monomer **8**. Hydrolysis of **8** by 1M $Ba(OH)_2$ in a THF/H_2O solution yielded the chiral monomer **9** (with protected C or G as bases), as reported in Figure 4.

The 5L-chiral box **PNA-3** (Fig. 2) was then synthesized using Boc-chemistry with the two preformed chiral monomers and commercially available Boc-PNA monomers. The crude products were purified by RP-HPLC and characterized by HPLC-MS, as described in the Experimental part.

Recognition Properties in Solution

The UV melting curves of the two chiral PNA are reported in Figure 5, and the corresponding melting temperatures are reported in Table 1 and compared with those of the achiral **PNA-1**. The 2D-chiral box **PNA-2**:DNA duplex was found to have a neat melting transition, with only a slightly lower melting temperature than the achiral PNA:DNA, similarly to what observed previously for lysine-based 2D-chiral box PNAs with different sequences.[26–28] On the contrary, the 5L-chiral box **PNA-3** was found to have a less steep increase in the absorbance upon melting, with a larger transition range. Although under the present conditions it was not possible to determine enthalpy/entropy contribution, it is evident that the thermodynamic properties should be significantly different from the 2D **PNA-2** case. The melting temperature for 5L-chiral box **PNA-3** was significantly higher than in the achiral PNA:DNA duplex, thus confirming that this modification is the most suited for obtaining PNA with high affinity for complementary DNA. This behavior is in line with the recent results of Ly and coworkers, who have used this type of modification for performing strand invasion of duplex DNA.[32,33]

The recognition of a single mismatch (A instead of G) was also evaluated by the decrease in the melting temperatures. In this case, the 2D PNA-**2**: DNA was found to have the lowest melting temperature, with a ΔT_m of 26°C, greater than that observed for the achiral **PNA-1** (22°C), whereas the 5L-chiral box **PNA-3** gave a stable duplex also in the presence of the mismatch, with ΔT_m of 20°C. Therefore, while the 5L **PNA-3** was superior in terms of binding, the 2D **PNA-2** was found to be the best model for the recognition of single mismatch.

The circular dichroism spectra of the two chiral PNA and of the PNA:DNA duplexes are reported in Figure 6. A certain degree of preorganization was present for both, as reported in similar cases for chiral PNA.[34] In particular, single strand 5L-PNA was found to be quite preorganized, with a CD spectrum reminiscent of a PNA arranged in a P-helix form.[34]

The PNA:DNA duplexes gave CD spectra consistent with that of achiral PNA, with a maximum at 265 nm and a minimum at 244 nm, suggesting that the highly

TABLE 1. UV and CD melting temperatures of the duplexes of the PNA with full matched and mismatched DNA duplexes

PNA	DNA	UV T_m[a] (°C)	ΔT_m[b] (°C)	CD T_m[a] (°C)	ΔT_m[b] (°C)
PNA-1	FM	67	22	69	22
PNA-1	MM	45		47	
PNA-2	FM	66	26	67	27
PNA-2	MM	40		40	
PNA-3	FM	75	20	75	19
PNA-3	MM	55		56	

[a]PNA:DNA melting temperature measured in phosphate buffer pH = 7.0, c = 5 μM of each strand.
[b]T_m(Full-match) − T_m(Mismatch).

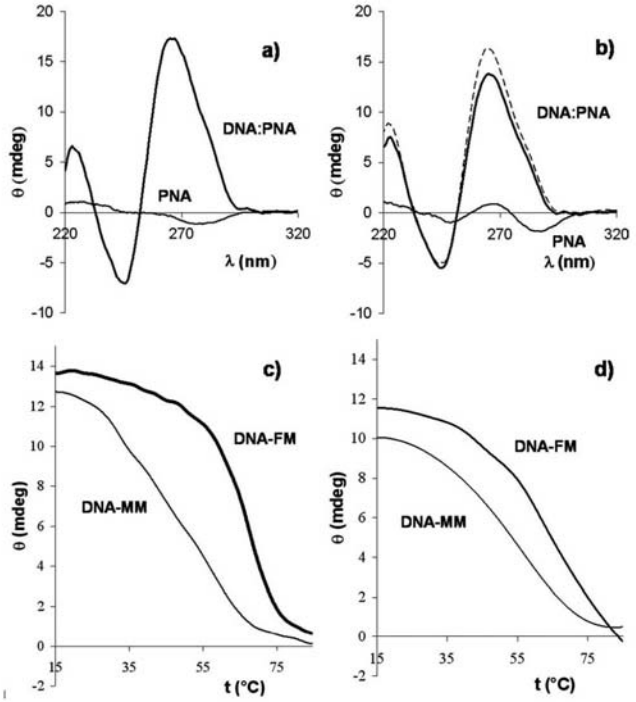


Fig. 6. CD spectra of (**a**) the 2D **PNA-2** and (**b**) **5L PNA-3** alone (thin line) and of the duplex with full complementary DNA (thick line) compared with the spectrum of the achiral **PNA-1: DNA-FM** duplex (dashed line); (**c** and **d**) melting curves at 260 nm for the full-match and single mismatch (G-A) DNA for **PNA-2** (c), and **PNA-3** (d). All measurements were done in PBS buffer, pH 7; concentration of each strand was 5 μM. Cell length was 1 cm.

constrained PNA did not induce a distorted conformation of the duplex, in agreement with our finding about the effect of 2D and 5L stereochemistry of the chiral monomers, which are both compatible with right-handed helices.

The CD melting curves of the duplexes recorder at 260 nm were found to be consistent with those obtained by UV absorption (Figs. 6c–6d), with a significant difference in the recognition of the single mismatch, which, also in this case, was more pronounced for the 2D **PNA-2**.

The higher recognition properties of the 2D chiral box model could be due to the different position of the substituent in the PNA backbone, which is attached to a carbon atom between two rigid amidic groups for the 2D-derivative, whereas in the case of 5L derivative, the substituent is located on the aminoethyl moiety, which allows a higher conformational freedom. The presence of a mismatch in the central base of the 2D chiral box could induce a conformational change which then affects the proper conformation of the monomer and of adjacent residues. If these are also highly constrained, the overall conformation of the "chiral box" segment would be highly distorted and would give rise to a stable PNA:DNA duplex. This effect is less pronounced in the case of 5L monomers, due to the possibility for the flanking monomers to better adapt to the distorted conformation, therefore preserving the positive electrostatic interactions which stabilize the duplex.

DNA Recognition on Microarrays

The recognition properties of the two PNAs were also evaluated on a solid support, using the microarray technology which was used previously for the analysis of multiple sequences, including point mutations.[25]

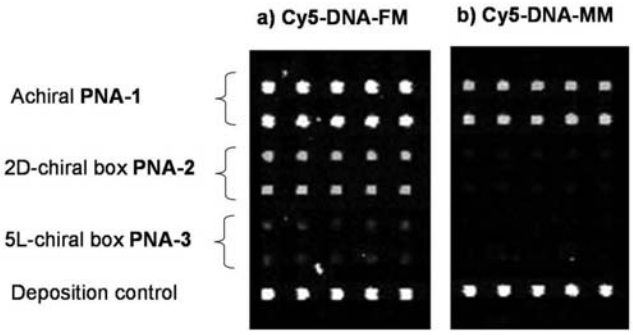


Fig. 7. Microarray analysis of Cy-5 labeled DNA oligonucleotides: (a) full-match **DNA-FM** and (b) single-mismatch **DNA-MM** sequences. Hybridization was performed in SSC buffer (0.1 μM of DNA in 2× SSC solution and 0.1% SDS). Images were obtained using a ScanArray™ Express Microarray Scanner, with λ_{ex} = 646 nm and λ_{em} = 664 nm.

The PNA were spotted on the surface of activated slides using solutions of two different concentrations (30 and 50 μM), both producing similar results (Fig. 7) suggesting that the concentration was sufficient in both cases to saturate the active sites of the slide. A set of the achiral, 2D chiral box, and 5L chiral box PNA was spotted in duplicate on the same slide, together with a Cy5-labeled oligonucleotide as a control of deposition. Using a multichamber silicone gasket, one set of spots was then hybridized with the full-match oligonucleotide, and the other with the mismatched one, each bearing Cy5-labeling.

The results are shown in Figure 7 and show that under the condition used, the achiral PNA gave rise to the more intense signal with the full-match DNA (Fig. 7a). The 2D **PNA-2** gave a stronger signal than 5L **PNA-3**, which showed very weak hybridization. Therefore, the stability of the PNA:DNA duplexes under these conditions were in a different order than those observed in solution, suggesting that electrostatic interactions which stabilize the duplex in solution can be affected by the matrix effect of the activated slide surface and by the additives normally used for hybridization.

Conformational properties are also another feature that can affect the behavior of the probes on the surface either by allowing a tighter interaction with the surface or changing the disposition of the charges toward the analyte. Possible side-chain-nucleobase interactions might occur, in view of the fact that positively charged amino acid side groups can give rise to cation–π interactions in protein–DNA complexes.

The hybridization with the mismatched oligonucleotide gave the results reported in Figure 7b, indicating best rec-

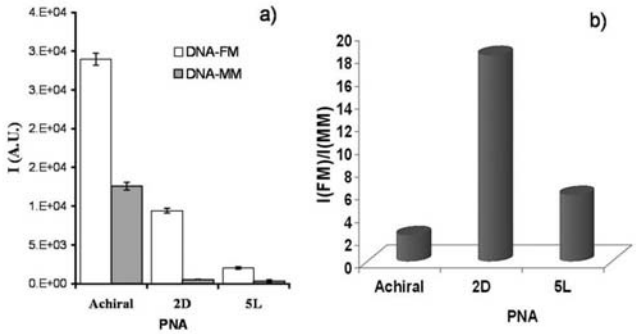


Fig. 8. (a) Quantitative analysis of microarray signals obtained with achiral **PNA-1**, 2D chiral-box **PNA-2**, and 5L chiral-box **PNA-3** with full match (**DNA-FM**, white bars) and mismatched DNA (**DNA-MM**, gray bars); vertical bars indicate standard deviations. (b) Selectivity ($I_{full\ match}/I_{mismatch}$) observed in the microarray hybridization for the various PNA.

Chirality DOI 10.1002/chir

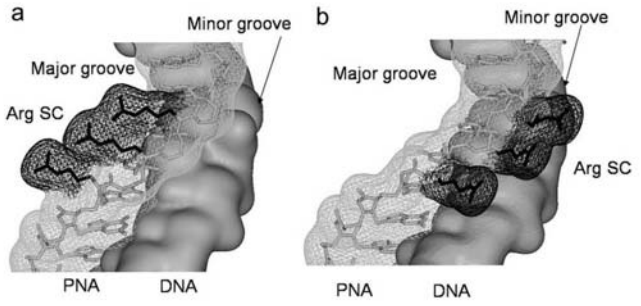


Fig. 9. Model system for the 2D (a) and 5L (b) chiral box Arginine-based PNA, showing the disposition of the Arginine side chains (Arg SC, in black) relative to PNA:DNA major and minor grooves (based on the crystal structure reported in Ref. 21).

ognition properties for the 2D-chiral box **PNA-2** than with the two other cases. This is in line with the results of sequence selectivity observed in solution and is expressed quantitatively in Figure 8.

Therefore, for the recognition of a single nucleotide mismatch on the array, the 2D-chiral box model is also in this case the best overall performing in terms of sequence selectivity.

Origins of Differences in Selectivity

In our recent works, we have established that stereoselectivity in the DNA binding ability was determined mainly by the intramolecular steric interactions between the PNA side chain and PNA backbone, as evaluated by the structural features of PNA:DNA duplexes.[14,22] However, the different behavior of the 2D and 5L-chiral box described in this article requires further discussion. Using the crystal structure of a 2D chiral box PNA-DNA duplex as a scaffold, and introducing arginine side chains either with the 2D or 5L model, the schematic model of the PNA-DNA interaction represented in Figure 9 can be proposed.

It can be noted that there is a major difference between the two types of modifications: while the side chains in the 2D model are pointing toward the major groove, in the 5L model they are directed toward the minor groove. This is in line with more favorable electrostatic interactions with the phosphate groups, due to shorter distance between charges, which generates higher stability of the PNA-DNA duplex.

The higher sequence selectivity observed in the case of 2D chiral box can be due to the position of the side chain, which is attached to the α-carbon of the more rigid glycine moiety of the PNA backbone, whereas the side chain in the 5L derivative is placed in the more flexible aminoethyl group. Distortion of the former will likely generate a conformation in which the side chains collide with each other, whereas in the case of the 5L derivative, they can be rearranged and the eventual repulsive interactions are compensated by the electrostatic interactions with the negative potential of the minor groove.

Since the side chains in the 5L model are closer to the DNA backbone, however, eventual surface-PNA interactions on this side will strongly hamper the interaction with DNA, while the same effect occurring on the major groove side would be better tolerated, thus leading to a better performance on the microarray system.

CONCLUSIONS

In the last few years, we have shown how chirality of PNAs can affect in a positive way DNA recognition and described several examples of increased sequence selectivity.[26,27] In particular, the introduction of a stereogenic center derived from D-amino acid in the C-2 carbon or from L-amino acid at the C-5 carbon of the PNA backbone has turned out to be very effective in inducing higher affinity and increased selectivity. This effect was magnified when several chiral monomers were used to form C-2 substituted chiral-box PNAs, which were used in combination with several techniques for the specific detection of point mutations and single nucleotide polymorphisms.

In this article, we compared for the first time the effect of 2D- and 5L-chiral-modified PNAs in the chiral-box model to increase selectivity in solution and on solid surfaces. The results showed that the 5L **PNA-3** was superior in terms of binding affinity in solution, whereas the 2D **PNA-2** model was superior in performances when recognition of single nucleotides were considered both in solution and in the microarray format. The latter information can be very precious in the design of new devices for the recognition of point mutations or single nucleotide polymorphisms in biomedical or food analysis.

Although the concentration used in the present studies are far from optimal for modern highly sensitive techniques of DNA detection, in the quest of new ultrasensitive methods, the molecular recognition events are the most important starting point. Optimization of techniques and introduction of schemes for amplification of the signal will then be possible, based on the most promising model allowing best binding and best sequence selectivity.

ACKNOWLEDGMENT

The authors thank Dr. Damiano Dotti for his technical assistance in PNA purification.

LITERATURE CITED

1. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. Nat Genet 1999;21:20–24.
2. Wolcott MJ. Advances in nucleic acid-based detection methods. Clin Microbiol Rev 1992;5:370–386.
3. Piatek AS, Tyagi S, Pol AC, Telenti A, Miller LP, Kramer FR, Alland D. Molecular beacon sequence analysis for detecting drug resistance in *Mycobacterium tuberculosis*. Nat Biotechnol 1998;16:359–363.
4. Ranade K, Chang MS, Ting CT, Pei D, Hsiao CF, Olivier M, Pesich R, Hebert J, Chen YD, Dzau VJ, Curb D, Olshen R, Risch N, Cox DR, Botstein D. High-throughput genotyping with single nucleotide polymorphisms. Genome Res 2001;11:1262–1268.
5. Wang J. From DNA biosensors to gene chips. Nucleic Acids Res 2000;28:3011–3016.
6. Service RF. Microchip arrays put DNA on the spot. Science 1998;282:396–399.
7. Southern EM. DNA chips: analyzing sequence by hybridization to oligonucleotides on a large scale. Trends Genet 1996;12:110–115.
8. Epstein CB, Butow RA. Microarray technology: enhanced versatility, persistent challenge. Curr Opin Biotechnol 2000;11:36–41.
9. Blais BW, Phillippe LM, Vary N. Cloth-based hybridization array system for detection of transgenic soy and corn by multiplex polymerase chain reaction. Biotechnol Lett 2002;24:1407–1411.
10. Rudi K, Rud I, Holck A. A novel multiplex quantitative DNA array based PCR (MQDA-PCR) for quantification of transgenic maize in food and feed. Nucleic Acids Res 2003;31:e62/1–e62/8.
11. Germini A, Mezzelani A, Lesignoli F, Corradini R, Marchelli R, Bordoni R, Consolandi C, De Bellis G. Detection of genetically modified soybean using peptide nucleic acids (PNAs) and microarray technology. J Agric Food Chem 2004;52:4535–4540.
12. Germini A, Rossi S, Zanetti A, Corradini R, Fogher C, Marchelli R. Development of a peptide nucleic acid array platform for the detection of genetically modified organisms in food. J Agric Food Chem 2005;53:3958–3962.
13. Nielsen PE, editor. Peptide nucleic acids: protocols and applications, 2nd ed. Norfolk, UK: Horizon Bioscience; 2004.
14. Corradini R, Tedeschi T, Sforza S, Totsingan F, Marchelli R. Peptide nucleic acids with a structurally biased backbone: effect of conformational constraints and stereochemistry. Curr Top Med Chem 2007;7:681–694.
15. Totsingan F, Tedeschi T, Sforza S, Corradini R, Marchelli R. Highly selective single nucleotide polymorphism (SNP) recognition by a chiral (SS) PNA beacon. Chirality 2009;21:245–253.
16. Englund EA, Appella DH. γ-substituted peptide nucleic acids constructed from L-lysine are a versatile scaffold for multifunctional display. Angew Chem Int Ed 2007;46:1414–1418.
17. Ishizuka T, Yoshida J, Yamamoto Y, Sumaoka J, Tedeschi T, Corradini R, Sforza S, Komiyama M. Chiral introduction of positive charges to PNA for double-duplex invasion to versatile sequences. Nucleic Acids Res 2008;36:1464–1471.
18. Dragulescu-Andrasi A, Rapireddy S, He G, Bhattacharya B, Hyldig-Nielsen JJ, Zon G, Ly DH. Cell-permeable peptide nucleic acid designed to bind to the 5′-untranslated region of E-cadherin transcript induces potent and sequence-specific antisense effects. J Am Chem Soc 2006;128:16104–16112.
19. Dragulescu-Andrasi A, Zhou P, He G, Ly DH. Cell-permeable GPNA with appropriate backbone stereochemistry and spacing binds sequence-specifically to RNA. Chem Commun 2005;244–246.
20. Corradini R, Sforza S, Tedeschi T, Marchelli R. Chirality as a tool in nucleic acid recognition: principles and relevance in biotechnology and in medicinal chemistry. Chirality 2007;19:269–294.
21. Menchise V, De Simone G, Tedeschi T, Corradini R, Sforza S, Marchelli R, Capasso D, Saviano M, Pedone C. Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA-DNA duplex. Proc Natl Acad Sci USA 2003;100:12021–12026.
22. Sforza S, Tedeschi T, Corradini, Marchelli R. Induction of helical handedness and DNA binding properties of peptide nucleic acids (PNAs) with two stereogenic centres. Eur J Org Chem 2007;5879–5885
23. Tedeschi T, Sforza S, Dossena A, Corradini R, Marchelli R. Lysine-based peptide nucleic acids (PNAs) with strong chiral constraint: control of helix handedness and DNA binding by chirality. Chirality 2005;17:S196–S204.
24. Nielsen PE. Peptide nucleic acid: a versatile tool in genetic diagnostics and molecular biology. Curr Opin Biotechnol 2001;12:16–20.
25. Calabretta A, Tedeschi T, Di Cola G, Corradini R, Sforza S, Marchelli R. Arginine-based PNA microarrays for APOE genotyping. Mol Bio-Syst 2009;1323–1330.
26. Sforza S, Corradini R, Ghirardi S, Dossena A, Marchelli R. DNA binding of a D-lysine-based chiral PNA: direction control and mismatch recognition. Eur J Org Chem 2000;2905–2913.
27. Corradini R, Feriotto G, Sforza S, Marchelli R, Gambari R. Enhanced recognition of cystic fibrosis W1282X DNA point mutation by chiral peptide nucleic acids probes by a surface plasmon resonance biosensor. J Mol Recognit 2004;17:76–84.
28. Tedeschi T, Chiari M, Galaverna G, Sforza S, Cretich M, Corradini R, Marchelli R. Detection of the R553X DNA single point mutation related to cystic fibrosis by a "chiral box" D-lysine-peptide nucleic acid probe by capillary electrophoresis. Electrophoresis 2005;26:4310–4316.
29. Sforza S, Tedeschi T, Corradini R, Ciavardelli D, Dossena A, Marchelli R. Fast, solid-phase synthesis of chiral peptide nucleic acids with a high optical purity by a submonomeric strategy. Eur J Org Chem 2003;1056–1063.
30. Corradini R, Di Silvestro G, Sforza S, Palla G, Dossena A, Nielsen PE, Marchelli R. Direct enantiomeric separation of *N*-aminoethylamino acids: determination of the enantiomeric excess of chiral peptide nucleic acids (PNAs) by GC. Tetrahedron: Asymmetry 1999;10:2063–2066.
31. Tedeschi T, Corradini R, Marchelli R, Pushl A, Nielsen PE. Racemization of chiral PNAs during solid-phase synthesis: effect of the coupling conditions on enantiomeric purity. Tetrahedron: Asymmetry 2002;13:1629–1636.
32. He G, Rapireddy S, Bahal R, Sahu B, Ly DH. Strand invasion of extended, mixed-sequence B-DNA by γPNAs. J Am Chem Soc 2009;131:12088–12090.
33. Rapireddy S, He G, Roy S, Armitage BA, Ly DH. Strand invasion of mixed-sequence B-DNA by acridine-linked, γ-peptide nucleic acid (γ-PNA). J Am Chem Soc 2007;129:15596–15600.
34. Dragulescu-Andrasi A, Rapireddy S, Frezza BM, Gayathri C, Gil RR, Ly DH. A simple γ-backbone modification preorganizes peptide nucleic acid into a helical structure. J Am Chem Soc 2006;128:10258–10267.

Molecular BioSystems

Dynamic Article Links

Cite this: *Mol. BioSyst.*, 2011, **7**, 1902–1907

www.rsc.org/molecularbiosystems

PAPER

A PNA microarray for tomato genotyping

Tullia Tedeschi,[a] Alessandro Calabretta,[a] Mariangela Bencivenni,[a] Alex Manicardi,[a] Giandomenico Corrado,[b] Martina Caramante,[b] Roberto Corradini,[a] Rosa Rao,[b] Stefano Sforza*[a] and Rosangela Marchelli*[a]

Received 4th February 2011, Accepted 7th March 2011
DOI: 10.1039/c1mb05048f

The design and development of a PNA microarray designed for the simultaneous identification of several SNPs characteristic of seven different tomato varieties is described. Highly selective arginine-based monomer containing PNAs (Arg-PNAs) have been used in order to obtain very selective probes. Seven modified PNA probes were synthesised and their binding properties in solution were studied. PNA-microarrays based on these probes were prepared and applied to SNP discrimination in model experiments using oligonucleotide mixtures simulating the different sequences of the seven tomato varieties. The strength and the limitations of such a system for SNP recognition are thoroughly discussed.

Introduction

In the last few years, the continuous and fast development of biotechnological techniques led to a wide range of selective and sensitive bioassays, aimed at recognizing genetically significant DNA markers with methods at the same time robust and simple,[1] used for the discrimination of an organism among others, or a mutation within the same organism. Although most of the research on DNA markers is related to genetic diseases,[2,3] these techniques are also becoming important in many other fields, such as food safety and authentication.[4,5] In food analysis, DNA recognition can be extremely useful for tracing the origin of a food product or for evidencing undeclared ingredients, even if present in very small amounts. DNA detection is thus increasingly applied as an answer to different needs, such as GMO detection,[6] microbial pathogen determination,[7] assessment of the presence of undeclared allergenic ingredients.[8]

Among the different DNA markers which might be addressed, Single Nucleotide Polymorphisms (SNPs),[9] which consist in single nucleobase changes within the genome, can be used as specific targets to identify a given food or a specific variety. A wide range of bioassays for the recognition of SNPs has been developed in the last few years, either in solution or on surface,[10] all aimed at the development of fast, cheap, robust and high-throughput methods. Among these, surface techniques, and in particular microarray-based platforms,[11,12] turned out to be extremely interesting. Such methods usually rely on the recognition of a DNA target by hybridization with a single strand oligonucleotide probe immobilized onto a surface.

The quality and efficiency of these methods can be seriously affected by the nature of the probes used to recognize complementary sequences; in particular, oligonucleotide probes may be replaced with modified molecules with an improved affinity in DNA binding leading to improved recognition properties.[13,14]

Peptide Nucleic Acids (PNA) (Fig. 1A)[15,16] are among the most promising probes, showing improved binding properties, if compared to DNA probes, as well as higher chemical and enzymatic resistance.

The development of PNA-based surface systems allowed us to obtain more efficient assays, in terms of selectivity in the recognition of point mutation, robustness and sensitivity.[17–19] In food analysis, although still not exploited routinely, several promising applications of PNAs have been published in the last few years. In particular PNA-microarrays were successfully used for the detection of GMOs, hidden allergens and ingredient authentication.[20]

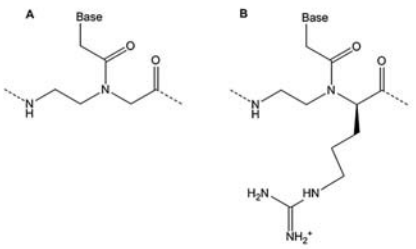


Fig. 1 Chemical structure of a standard PNA (A) and of an Arg-based PNA monomer used for improved SNP recognition (B).

[a] *Dipartimento di Chimica Organica e Industriale, University of Parma, Parco Area delle Scienze 17a, I-43124, Parma, Italy. E-mail: stefano.sforza@unipr.it; Fax: +39-0521-905472*
[b] *Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, University of Napoli, Reggia di Portici-Via Università 100, I-80055, Portici-Napoli, Italy*

Although PNA-microarrays were demonstrated to be extremely efficient, their properties, and in particular the specificity of recognition, can be further improved by introducing new chemical modifications in the PNA structure, able to allow a straightforward clear-cut discrimination, even in the case of single mismatch recognition (SNP). Examples of the modified PNA with improved binding specificity towards point mutations have been reported, obtained by introducing a steric constraint, such as cyclic or linear residues, in the backbone.[21,22]

The synthesis and the binding properties of PNAs containing arginine-based monomers have been recently reported[22,23] (Arg-PNAs, Fig. 1B), demonstrating their enhanced recognition properties, in terms of binding affinity and mismatch recognition, in solution and on microarray platforms.[22,24] Such PNAs might constitute an advancement in SNP recognition as compared to traditional DNA-based systems, characterized by poor selectivity.[14]

Here we show the design and the development of a model Arg-PNA microarray produced for the simultaneous identification of several SNPs characteristic of seven different tomato varieties. The design and synthesis of highly selective arginine-based monomer containing PNAs (Arg-PNAs) are reported together with their binding properties in solution. PNA-microarrays based on these probes were prepared and applied to SNP discrimination in model experiments using oligonucleotide mixtures simulating the different sequences of the seven tomato varieties. The strength and the limitations of such a system for SNP recognition will be thoroughly discussed.

Results and discussion

PNA design and synthesis

A set of 20 EST-SNPs, described in the literature[25] and characterized by high probability to be polymorphic in cultivated tomatoes, were selected and assayed as CAPS on genomic DNA of 7 tomato hybrids, characterized by different fruit shapes. SNPs identity was verified by sequencing the PCR-amplified products. Four SNP loci (LeOH 8.4, LeOH 31.3, LeOH 63, LeOH 23.1) were able to discriminate all the hybrids.

PNA probes were designed on these sequences with the polymorphic SNP placed in the central position of the strand. The PNA sequences were designed by using PNA-Tm calculator software.[26] PNA probes have different lengths to maintain the same purine/pyrimidine ratio and consequently the same T_m. In the case of PNA **2** and PNA **7**, the central position did not actually correspond to a SNP, but to a nucleobase deletion. (C is missed in PNA **7**.)

The theoretical signal pattern expected by using the seven designed PNA probes targeted against the corresponding DNA sequences in the microarray is reported in Fig. 2.

In order to obtain a system able to discriminate the SNPs of interest with enhanced selectivity, modified PNAs have been designed. In particular "chiral box" PNAs, taking their name by the fact that they bear three adjacent modified monomers in the middle of the PNA strand, were chosen for their outstanding ability to discriminate single mismatches, as shown in previous works.[27] Although "chiral box" PNAs had been previously developed using lysine-modified monomers, in this work three 2D-arginine modified monomers (Fig. 3) were introduced in the middle of the sequence, in order to avoid any interference in the binding of the probe to the microarray surface, since the binding is done by reacting the N-terminal group. Their synthesis was performed as reported previously.[23] PNAs were finally purified by RP-HPLC and characterized by ESI-MS-RP-HPLC.

	PNA LeOH 8.4-T	PNA LeOH 23.1-G	PNA LeOH 63-T	PNA LeOH 8.4– C	PNA LeOH 31.3-G	PNA LeOH 63-C	PNA LeOH 23.1-del
'Ercole' (oblung)	●	●	●	●		●	
'Cirio3' (oblung)		●		●		●	
'Talent' (oblung)	●		●	●	●	●	●
'Leader' (round)	●	●	●	●			
'Px1296' (round)	●	●	●			●	
'Tomito' (cherry)	●	●	●	●	●	●	
'Minidor' (cherry)		●	●	●	●	●	

Fig. 2 Simulation of the recognition of the seven tomato varieties by the seven designed PNA probes. For each cultivated variety the fruit shape is reported in brackets.

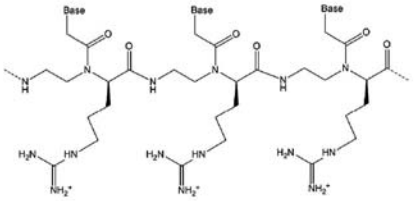


Fig. 3 Chemical structure of "chiral box" PNA: three modified Arg-based PNA monomers are placed consequently in the middle of a standard PNA sequence.

PNA binding properties in solution

The probe affinity towards complementary oligonucleotide sequences and selectivity for mismatch recognition were initially tested in solution by measuring the melting temperatures (T_m) of both fullmatch and mismatch PNA–DNA duplexes (Table 1). Experiments were performed in phosphate buffer (10 mM, pH = 7), at a strand concentration of 5 μM. Melting temperatures were evaluated as the first derivatives of the UV absorption curves at 260 nm in a temperature range from 20 to 90 °C.

The values (Table 1) showed that fullmatch oligonucleotides were recognized and bound with a high affinity, with T_m ranging in a quite close range, from 54 to 66 °C. PNAs **1**, **3**, **4**, **5** and **6** were also shown to be able to perform mismatch recognition, with an efficiency which turned out to be quite

Table 1 PNA sequences and melting temperatures of the PNA : DNA antiparallel fullmatch/mismatch duplexes in phosphate buffer 10 mM (pH = 7) at a 5 μm concentration for each strand

PNA (target DNA accession number-SNP nucleobase)	Sequence[a]	T_m/°C PNA–DNA fullmatch[b]	T_m/°C PNA–DNA mismatch[b]
1 (LeOH 8.4-T)	H–(AEEA)$_2$AAAG**A**CCGA–NH_2	54 (with **1′**)	18 (with **4′**)
2 (LeOH 23.1-G)	H–(AEEA)$_2$TTTT**C**CGGTGG-NH_2	56 (with **2′**)	35 (with **8′**)
3 (LeOH 63-T)	H–(AEEA)$_2$TGTGT**C**AAAG—NH_2	56 (with **3′**)	41 (with **7′**)
4 (LeOH 8.4-C)	H–(AEEA)$_2$AAAG**G**CCGA—NH_2	60 (with **4′**)	45 (with **1′**)
5 (LeOH 31.3-G)	H–(AEEA)$_2$CTTAT**C**C**G**GTGCC—NH_2	66 (with **5′**)	40 (with **6′**)
6 (LeOH 63-C)	H–(AEEA)$_2$TGTGT**C**G**A**AAG—NH_2	61 (with **7′**)	51 (with **3′**)
7 (LeOH 23.1-del)	H–(AEEA)$_2$TTTT**C**GGTGG–NH_2	56 (with **8′**)	54 (with **2′**)

[a] PNA monomers are indicated with the symbol of the corresponding nucleobase; chiral monomers based on 2D-Arg are reported in bold, the base corresponding to the SNP position is underlined; H means free N-terminus, NH_2 means carboxamide C-terminus; AEEA = aminoethoxyethoxy-acetyl: spacer. [b] Oligonucleotide sequences (SNP base is bold underlined): **1′**: 5′-TCGG**T**CTTT-3′, **2′**: 5′-CCACC**G**GAAAA-3′, **3′**: 5′-C**T**TTTGACACA-3′, **4′**: 5′-TCGG**C**CTTT-3′, **5′**: 5′-GGCACC**G**GATAAG-3′, **6′**: 5′-GGCACC**A**GATAAG-3′, **7′**: 5′-CTTT**C**GACACA-3′, **8′**: 5′-CCACC**G**AAAA-3′.

variable, ranging from good (10 °C drop in T_m for PNA **6**) to exceptional (36 °C drop for PNA **1**). Very likely, unpredictable conformational changes due to the sterical hindrance of the arginine side chains and the chiral constrain affected duplex formation, thus affecting the final duplex stability.

PNA **2** and PNA **7** are to be considered aside, since they were designed to recognize not a mismatch, but a base deletion. PNA **2**, the one complementary to the full sequence, showed a consistent drop in affinity towards the oligonucleotide sequence bearing a deletion in the middle (ΔT_m = −21 °C), indicating that a protruding nucleobase in the PNA strand (in this case a C) was poorly tolerated from the duplex. Quite interestingly, PNA **7**, complementary to the sequence with the deletion, showed almost the same affinity towards the oligonucleotide sequence bearing an extra nucleobase in the middle, indicating that a protruding nucleobase in the DNA strand (in this case an A) can be very well tolerated, with almost no effect on the duplex stability.

PNA binding properties on the microarray system

The preparation of the PNA microarray platform with the seven synthesized PNAs was then performed. PNA deposition was carried out by MICROCRIBI Microarray Service-CRIBI (University of Padova, Italy), as previously reported.[22] The devices were tested by hybridizing them with solutions containing Cy5-labelled synthetic oligonucleotides, mimicking the DNA sequences to be recognized. The hybridization conditions were optimized for each probe, both with complementary and mismatched oligonucleotides, in order to define the best parameters allowing simultaneous hybridization of all PNAs.

As reported previously,[22] in order to increase the hybridization efficiency and to refine the spot shapes, an incubation–hydration step in SDS 0.1%, saline sodium citrate (SSC) buffer (0.3 M NaCl, 0.03 M sodium citrate, pH = 7), at 40 °C for 30 min was introduced before hybridization with the oligonucleotide solutions. This step was particularly required when using the slides several days after PNA spotting. The hybridization solutions were prepared in SSC buffer, 0.1% SDS by mixing Cy5 labelled oligonucleotides at 1 μm concentration.

After optimization of the hybridization conditions, the array was tested by simulating the different tomato varieties, each characterized by its own combination of the different SNP-deletion, by mixing the different oligonucleotide sequences, in order to imitate the real samples, as shown in Table 2.

Table 2 Tomato genotypes related to different SNPs

Variety	Oligonucleotide sequences simulation[a]
'Ercole'	1′ + 2′ + 3′ + 4′ + 7′
'Cirio 3'	2′ + 4′ + 7′
'Talent'	1′ + 3′ + 4′ + 5′ + 7′ + 8′
'Leader'	1′ + 2′ + 3′ + 4′
'PS1296'	1′ + 2′ + 3′ + 7′ + 8′
'Tomito'	1′ + 2′ + 3′ + 4′ + 5′ + 7′
'Minidor'	2′ + 3′ + 4′ + 5′

[a] Oligonucleotide sequences are reported in Table 1.

The solutions were poured on the slides just after mixing at room temperature, in order to avoid unspecific interactions. Hybridization was performed in seven independent experiments, one for each variety at room temperature. The results obtained in optimal conditions are shown in Fig. 4.

By comparing the experimental values reported in Fig. 4 with the theoretical values in Fig. 2, it is possible to outline the

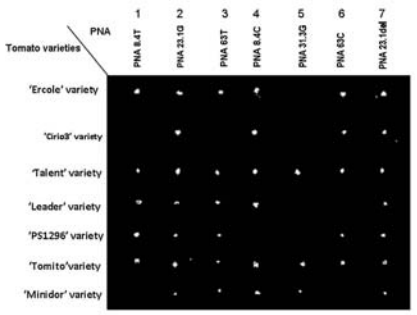


Fig. 4 PNA microarray analysis with Cy5-oligonucleotides in different combinations (see Table 2) in order to simulate the seven tomato varieties.

performance of the PNA probes on the surface. PNA **1** (LeOH 8.4-T), **4** (LeOH 8.4-C), **5** (LeOH 31.3-G), **3** (LeOH 63-T) and **6** (LeOH 63-C) behaved perfectly, showing signals only in the presence of their fullmatch oligonucleotide counterparts, and not giving any signals in the presence of oligonucleotides bearing a single mismatch simulating the SNP. Thus, even in the presence of small differences in stability shown in solution, all these PNA probes were perfectly able to discriminate point mutations on a surface system. PNA **2** (LeOH 23.1-G) and **7** (LeOH 23.1-del), aimed at recognizing the deletion, on the other side, failed to perform a specific recognition, showing positive signals for all oligonucleotides, with or without the deletion. If this result was somehow expected in the case of PNA **7**, for which in solution there was practically no difference in T_m between the target and non-target DNA, the result was somehow surprising for PNA **2**, which in solution had shown a quite consistent difference in affinity towards the two DNA. These results seem to indicate that on the surface, differently than in solution, the recognition of the deletion is somehow difficult, indicating that protruding bases in either strand (PNA or DNA) can be very well tolerated and therefore might not be a good target for PNA probes on a surface system.

In any case, the experiment clearly indicated that the seven varieties, except for 'Talent' and 'Tomito', could be discriminated by the microarray here presented.

Finally, after conditions were set with oligonucleotides having the same length as PNAs, an investigation of the microarray performance with longer DNA sequences, which could better mimic amplified DNA extracted from real food samples, was undertaken.

In particular, the LeOH 23.1 sequence was chosen for testing hybridizations with long DNA tracts. In order to study if and how the different positions of the target sequence in the eventual amplicon would affect the DNA detection, different hybridization experiments by using longer DNA oligonucleotides containing the target sequences in the middle, at the 5′ and at the 3′ ends, were performed. The different sequences used for hybridization experiments were three 118 bp oligonucleotides differing in the position of the target sequence inside the oligomer, and one shorter variant of one of them (60 bp).

Hybridization experiments with the four DNA single strand sequences were performed under the same conditions used for short oligonucleotides. In all cases, no signals indicating specific PNA–DNA complexation were observed, with or without a preliminary incubation step at 95 °C for 5 min and/or the hybridization performed at room temperature and at 40 °C. The use of Tween 20 replacing SDS as a non-charged surfactant did not allow us to observe improved results, as well as the use of denaturating conditions (with the addition of small percentages of formamide), in order to make the PCR product more accessible to the PNA probes.

These preliminary results seem to indicate that the chiral box PNA microarray platform, albeit very specific, loses binding efficiency in the presence of longer DNA sequences, preventing the formation of PNA–DNA duplexes. Moreover, in previous work it had been shown that on DNA microarray devices the distance between the fluorophore used to label hybridization and the target sequence affects the efficiency of detection,[28] in our case this phenomenon seems not to be responsible for this behaviour, since the target sequence in our experiments had different distances from the fluorophore.

It may be speculated that the nature of the solid support and the presence of positive charges on PNA probes are more likely responsible for this effect, since they could promote the aggregation of the PNA on the surface, lowering the efficiency of the hybridization. Alternatively, long DNA tracts may aggregate on the positively charged surface, making the target DNA difficult to be bound by the PNA probes, and thus easily rinsed away during the washing steps.

Whatever might be the cause, it has already been demonstrated in previous experiments that chiral PNA microarrays give less intense signals when hybridized to complementary oligonucleotides, as compared to standard PNAs, showing a higher specificity but a lower sensitivity in DNA detection.[24]

However, given the promising potentialities of this kind of probe, future studies will have to be implemented in order to better understand the kind of interactions between the positively charged PNA probes, the microarray surface and long DNA tracts, which lowers the PNA binding efficiency.

Conclusions

The Arg-PNAs designed and synthesized in this work, tested in solution and on microarray systems, showed the ability to perfectly discriminate SNP containing sequences and thus their potentiality to be used to genotype tomato varieties. The microarray here presented was able to simultaneously discriminate tomato varieties, except for 'Talent' and 'Tomito', in simulation experiments using oligonucleotide mixtures. Although the efficiency of DNA binding was somewhat lower than that usually shown by standard PNA arrays, causing a lower sensitivity towards long DNA sequences in this array system, the unsurpassed specificity of the PNA probes makes them very promising for the development of PNA-based genotyping methodologies. Further studies will be needed in order to develop modified PNA-based systems to a level compatible with the presently available DNA arrays, but they have the potential to compete in the future with the commercially available oligonucleotide-based recognition systems.

Experimental section

Materials and reagents

PNA N-Boc protected standard monomers, *N*-[1*H*-1,2,3-(benzotriazol-1-yl)-(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide (HBTU), *N*-[(dimethylamino)-1*H*-1,2,3-triazol[4,5-*b*]pyridine-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide (HATU), dichloromethane, *N*-methylpyrrolidone (NMP) and *N*,*N*-dimethylformamide (DMF) were purchased from Applied Biosystems (Foster City, CA, USA). Carboxymethyl-*O*-benzyl-guanine was purchased from ASM (Hannover, Germany). Boc-Arg(Tos)-OH, Boc-D-Arg(Tos)-OH and *m*-cresol were from Fluka (Buchs, Switzerland). The (4-methylbenhydryl)amine (MBHA) PS resin was from Novabiochem (Laufelfingen, Switzerland). Diisopropylethylamine (DIPEA), *N*,*O*-bistrimethylsilyl acetamide (BSA),

trifluoromethane sulfonic acid (TFMSA), trifluoroacetic acid (TFA), 3-hydroxy-1,2,3-benzotriazin-4-(3*H*)-one (DhBtOH) and diisopropylcarbodiimide (DIC) were from Sigma-Aldrich (St. Louis, MO, USA). Oligonucleotides (guaranteed oligos grade) were purchased from Thermo Fisher Scientific (Ulm, Germany) and used without further purification.

Definition of target sequences of tomato DNA

DNA isolation, PCR reactions and CAPS analysis were performed as previously described.[29,30] The DNA reaction sequence was prepared with the Big Dye terminator kit (Applied Biosystems) according to manufacturer instructions and analysed on the ABI 3730 sequencer (Applied Biosystems).

PNA synthesis

The synthesis of submonomers and chiral PNAs was performed as described previously for similar PNAs[23] by using manual submonomeric synthesis for the chiral residues and automated Boc-SPPS protocols on a ABI 433A synthesizer, following the procedures provided by the company, for the achiral residues.

HPLC analysis for all PNAs was carried out by LC-MS by using an XTerra analytical C18 column (3 × 250 mm, 5 μm, flow 0.5 ml min ^{1}), gradient elution from 100% H_2O (0.2% HCOOH, eluent A) to 60% H_2O and 40% CH_3CN (0.2% HCOOH, eluent B) in 30 min. The MS detector set in the positive ion mode, capillary voltage 3 kV, cone voltage 30 V, full scan acquisition from 150 to 1500 *m*/*z*. HPLC purification for all PNAs was carried out on a semipreparative Jupiter (Phenomenex) C18 column (10 × 300 mm, 5 μm, flow 4 ml min ^{1}); eluent A: 100% H_2O (0.1% TFA), eluent B: 60% H_2O and 40% CH_3CN (0.1% TFA).

PNA **1**: ESI-MS: calcd *m*/*z*: 771.3 (MH_4^{4+}), 617.2 (MH_5^{5+}), 514.5 (MH_6^{6+}), 441.2 (MH_7^{7+}) found *m*/*z*: 771.3, 617.4, 514.6, 441.2

PNA **2**: ESI-MS calcd *m*/*z*: 767.3 (MH_4^{4+}), 614.0 (MH_5^{5+}), 511.9 (MH_6^{6+}), 438.9 (MH_7^{7+}) found *m*/*z*: 767.5, 614.2, 512.0, 439.1.

PNA **3**: ESI calcd *m*/*z*: 901.6 (MH_4^{4+}), 721.5 (MH_5^{5+}), 601.4 (MH_6^{6+}) found *m*/*z*: 901.7, 721.5, 601.5.

PNA **4**: ESI calcd *m*/*z*: 1085.4 (MH_3^{3+}), 814.3 (MH_4^{4+}), 651.6 (MH_5^{5+}) found *m*/*z*: 1085.3, 814.3, 651.5.

PNA **5**: ESI-MS calcd *m*/*z*: 1019.5 (MH_4^{4+}), 815.8 (MH_5^{5+}), 680.0 (MH_6^{6+}), 583.0 (MH_7^{7+}) found *m*/*z*: 1019.6, 816.2, 680.0, 583.1.

PNA **6**: ESI calcd *m*/*z*: 729.9 (MH_5^{5+}), 608.4 (MH_6^{6+}), 521.7 (MH_7^{7+}) found *m*/*z*: 730.0, 608.6, 522.0.

PNA **7**: ESI-MS calcd *m*/*z*: 726.7 (MH_5^{5+}), 605.8 (MH_6^{6+}), 519.4 (MH_7^{7+}) found *m*/*z*: 726.8, 605.9, 519.4.

Melting temperature measurements

Stock solutions of oligonucleotides were prepared in doubly distilled water; their actual concentration was calculated by UV absorbance using the following ε260 (M ^{1} cm ^{1}) for the nucleobases: T 8600, C 6600, A 13700, G 11700. By using these concentrations, hybrid solutions containing 5 μM PNA–DNA duplexes were prepared. All the hybridization experiments were carried out in 10 mM phosphate buffer, 100 mM NaCl, 0.1 mM EDTA, pH = 7. All the hybrid samples reported were first incubated at 90 °C for 5 min, then slowly cooled to room temperature. The samples were heated (1 °C min ^{1}) and the UV signal variation at 260 nm was recorded. Melting temperatures were measured as the maximum of the first derivatives of the melting curves.

PNA spotting on microarray slides

PNA spotting was performed by MICROCRIBI Microarray Service-CRIBI (University of Padova, Italy) following the procedure reported in a previous paper[22] without further optimization (resulting in some cases in irregular spot shapes) onto commercial slides (Genorama® Microarray Slides, Asper Biotech Ltd., Tartu, Estonia).

Oligonucleotide hybridization on microarray slides

Each Cy5 labeled oligonucleotide (**1′–8′**) was dissolved in a solution containing SSC buffer (0.3 M NaCl, 0.03 M sodium citrate, pH = 7), SDS 0.1% at a concentration of 1 μM. Before hybridization the slides were hydrated with a solution of SSC buffer (0.3 M NaCl, 0.03 M sodium citrate, pH = 7), SDS 0.1% at room temperature for 1 h. Then, oligonucleotide solutions were deposited on the slides by using a hybridization chamber and left at room temperature for 2 h. After hybridization, washing steps were performed at room temperature for 2 min with a solution of SSC buffer (0.3 M NaCl, 0.03 M sodium citrate, pH = 7), SDS 0.1%. Slides were dried by centrifugation at 2500 r.p.m. for 8 min.

Solutions containing longer oligonucleotides in the same conditions described above were incubated at 95 °C for 5 min before pouring them on the slides. The hybridization step was performed at room temperature and at 40 °C for 2 h.

Microarray analysis

Pictures of the PNA microarrays were recorded by using a scanner ScanArray Express 20 (Perkin Elmer, Waltham, USA). Laser was set at an excitation wavelength of 646 nm and emission at 664 nm, for Cy5. Pictures were acquired with laser power at 100% and photomultiplier at 40%.

Acknowledgements

This work was supported by the "Progetto MIUR-Laboratorio pubblico privato GenoPOM", DM17732.

The research was in the context of EUROBIOTECH—European Biotechnology project.

References

1 F. R. R. Teles and L. P. Fonseca, *Talanta*, 2008, **77**, 606.
2 Y. S. Min, J. H. Choi, S. Y. Lee and N. C. Yoo, *J. Microbiol. Biotechnol.*, 2009, **19**(7), 635.
3 Y. Nakamura, *J. Hum. Genet.*, 2009, **54**, 1.
4 A. K. Lockley and R. G. Bardsley, *Trends Food Sci. Technol.*, 2000, **11**, 67.
5 M. Woolfe and S. Primrose, *Trends Biotechnol.*, 2004, **22**, 222.
6 M. Hernandez, D. Rodriguez-Lazaro and A. Ferrando, *Curr. Anal. Chem.*, 2005, **1**, 203.
7 A. Lauri and P. O. Mariani, *Genes Nutr.*, 2009, **4**, 1.

8 R. E. Poms, E. Anklam and M. Kuhn, *J. AOAC Int.*, 2004, **87**, 1391.
9 F. S. Collins, L. D. Brooks and A. Chakravarti, *Genome Res.*, 1998, **8**, 1229.
10 J. Ragoussis, *Annu. Rev. Genomics Hum. Genet.*, 2009, **10**, 117.
11 J. Kian-Kon and W.-T. Liu, *Anal. Bioanal. Chem.*, 2006, **386**, 427.
12 A. Sassolas, B. D. Leca-Bouvier and L. J. Blum, *Chem. Rev.*, 2008, **108**, 109.
13 E. M. Southern, *Electrophoresis*, 1995, **16**, 1539.
14 J. Weiler, H. Gausepohl, N. Hauser, O. N. Jensen and J. D. Hoheisel, *Nucleic Acids Res.*, 1997, **25**, 2792.
15 P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497.
16 P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm and B. Norden, *Nature*, 1994, **368**, 561.
17 O. Brandt and J. D. Hoheisel, *Trends Biotechnol.*, 2004, **22**, 617.
18 F. R. Raymond, H. Hoang-Ann, R. Peytavi, L. Bissonnette, M. Boissinot, F. J. Picard, M. Leclerc and M. Bergeon, *BMC Biotechnol.*, 2005, **5**, 10.
19 A. Germini, S. Rossi and A. Zanetti, *et al.*, *J. Agric. Food Chem.*, 2005, **53**, 3958.
20 S. Sforza, R. Corradini, T. Tedeschi and R. Marchelli, *Chem. Soc. Rev.*, 2011, **40**, 221.
21 J. K. Pokorski, J. M. Nam, R. A. Vega, C. A. Mirkin and D. H. Appella, *Chem. Commun.*, 2005, 2101.
22 A. Calabretta and T. Tedeschi, *et al.*, *Mol. BioSyst.*, 2009, **5**, 1323.
23 P. Zhoua, A. Dragulescu-Andrasia, B. Bhattacharyab, H. O'Keefeb, P. Vattab, J. J. Hyldig-Nielsen and D. H. Ly, *Bioorg. Med. Chem. Lett.*, 2006, **16**(18), 4931.
24 A. Manicardi, A. Calabretta, M. Bencivenni, T. Tedeschi, S. Sforza and R. Marchelli, *Chirality*, 2010, **22**, 161.
25 W. Yang, X. Bai, E. Kabelka, C. Eaton, S. Kamoun, E. van der Knaap and D. M. Francis, *Molecular Breeding*, 2004, **14**, 21.
26 http://www6.appliedbiosystems.com/support/pnadesigner.cfm.
27 S. Sforza, R. Corradini, S. Ghirardi, A. Dossena and R. Marchelli, *Eur. J. Org. Chem.*, 2000, 2905.
28 L. Zhang, T. Hurek and B. Rienhold-Hurek, *Nucleic Acids Res.*, 2005, **33**, e165.
29 M. Caramante, R. Rao, L. M. Monti and G. Corrado, *Sci. Hortic.*, 2009, **120**, 560.
30 M. Caramante, G. Corrado, L. M. Monti and R. Rao, *Food Control*, 2011, **22**, 549.

Eur Food Res Technol
DOI 10.1007/s00217-011-1640-4

ORIGINAL PAPER

Assessing allergenicity of different tomato ecotypes by using pooled sera of allergic subjects: identification of the main allergens

Mariangela Bencivenni · Andrea Faccini · Chiara Bottesini · Rosa Rao · Aikaterini Detoraki · Erminia Ridolo · Gianni Marone · Pier Paolo Dall'Aglio · Arnaldo Dossena · Rosangela Marchelli · Stefano Sforza

Received: 28 May 2011 / Revised: 22 November 2011 / Accepted: 28 November 2011

Abstract An evaluation of the potential allergenicity of different tomato ecotypes is reported. Twelve tomato ecotypes were assessed through a proteomic approach, using pools of sera of allergic patients from two different regions (Emilia Romagna in Northern Italy and Campania in Southern Italy), in order to identify the major allergens and evaluate differences in IgE binding properties of the tomato cultivars. Pooled sera of allergic people from Emilia Romagna showed as the main allergen a suberization-associated anionic peroxidase, whereas pooled sera of allergic patients from Campania were mostly reactive to profilin. The two proteins were identified through a proteomic approach based on the use of high-resolution mass spectrometric techniques. Quite interestingly, in some cases, several ecotypes showed a less reactivity toward patients' sera than other, potentially indicating the possibility to identify ipoallergenic varieties. Anyway, the allergenic pattern response to tomatoes was serum-specific, indicating that the allergenic properties of different tomato ecotypes are defined by the specific proteins to which the patient is sensitized, a strong indication that ipoallergenicity of the different ecotypes is possible, but mostly related to the individual susceptibility.

Keywords Tomato allergens · Proteomic · Mass spectrometry · Tomato local varieties

M. Bencivenni · C. Bottesini · A. Dossena · R. Marchelli · S. Sforza (✉)
Dipartimento di Chimica Organica e Industriale, University of Parma, Parco Area delle Scienze 17/A, 43125 Parma, Italy
e-mail: stefano.sforza@unipr.it

A. Faccini
Centro Interdipartimentale Misure 'Giuseppe Casnati', University of Parma, Parco Area delle Scienze 23/A, 43125 Parma, Italy

R. Rao
Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, University of Naples, Reggia di Portici—Via Università 100, 80055 Portici-Naples, Italy

A. Detoraki · G. Marone
Dipartimento di Medicina Clinica e Scienze Cardiovascolari e Immunologiche, University of Naples, Naples, Italy

E. Ridolo
Dipartimento di Scienze Cliniche, University of Parma, Via Gramsci 14, 43100 Parma, Italy

P. P. Dall'Aglio
Dipartimento di Clinica Medica, Nefrologia e Scienze della Prevenzione, Via Gramsci 14, 43100 Parma, Italy

Introduction

Food allergies have become a serious concern in Western countries, in which they affect about 6% of young children and 3–4% of adults, and their prevalence appears to be on the rise [1]. On November 2007, a directive of the Commission of the European Communities established a list of food ingredients that have to be indicated on the labels, since they may cause adverse reactions in susceptible individuals [2]. Fruits, vegetables and their derivatives have generally not been included in this list as a possible cause of allergic reactions, with the exception of cereals, soybean, nuts, celery, mustard, sesame and lupines. Anyway, other allergenic foods of vegetal origin may also represent a serious threat for consumers' safety, as plant food allergies are spreading quickly, probably due to the cross-reaction with other pollen allergens [3–5].

Published online: 18 December 2011

Among cultivated crop plants, tomato (*Solanum lycopersicum* L.) is increasingly consumed due both to its beneficial effects on human health and to the spreading of fresh or processed cooked products. In spite of this increasing consumption (or maybe because of that), tomato allergy is also increasing and this food can be considered as an emerging allergen, which actually affects 1.5–16% of food-allergic population [6], and it has been shown to be related to other allergies, like allergy against grass pollen [7] and latex [8] by cross-reactivity with homologous protein sequences. So far, the International Union of Immunological Society (I.U.I.S) recognized four tomato allergens: *Lyc e 1* (tomato profilin) [6], *Lyc e 2* (tomato *β-fructofuranosidase*) [9], *Lyc e 3* (tomato nonspecific lipid-transfer protein) [10] and *Lyc e 4* (tomato intracellular pathogenesis-related protein) (http://www.allergen.org), but many other reactive proteins to tomato fruit extracts have been reported in the literature. Kondo and co-workers identified a polygalacturonase 2A (PG2A, 46 kDa), a pectinesterase (PE, 14 kDa), a *β*-fructofuranosidase (22 kDa) and a superoxide dismutase (18 kDa) as IgE-binding allergenic proteins in tomato fruit, by N-terminal amino acid sequencing [11]. During studies devoted to elucidate the correlation existing between cooked and fresh tomato assumption and symptoms, an allergen of about 9 kDa, heat-labile and pepsin-resistant was partially characterized by immunoblotting, pepsin digestion and heating [12]. Several studies were also carried out in order to analyze allergenic properties of the different parts of tomato fruits. Using a multidimensional protein fraction strategy and LC–MS/MS, a legumin (47 kDa) and vicilin (65 kDa) proteins were purified from tomato seeds, showing strong IgE reactivity in immunoblots [13]. Pravettoni and co-workers found different LTP allergenic isoforms in fresh tomato peel, pulp and seeds. Although conventional heat treatments employed during the production of tomato-based products usually degrade proteins, thus strongly reducing their IgE binding capacity [14], industrial tomato derivatives have been demonstrated to still contain LTP [15].

Rather than exclude tomato from the diet of allergic subjects, genetic engineering may be applied in order to obtain hypoallergenic tomato fruits. Gene-silencing approaches were successfully used to target *Lyc e 1* and *Lyc e 3* genes, and low-allergenic tomato plants were obtained [10, 16]. Although these strategies have proved to be very efficient in eliminating allergenic proteins, hypoallergenic genetically modified plants are difficult to produce and maintain as stable lines, since, as mentioned above, allergic subjects' IgE cross-react toward different tomato proteins, so several genes corresponding to main tomato allergens should be silenced in transgenic lines. This could damage cell homeostasis because some tomato allergenic proteins play important physiological and structural roles [17]. Last but not least, genetically modified plants are still not accepted among European public opinion. Alternatively, the selection of genetic resources with low expression of allergenic proteins represents a valuable tool both to understand the genetic base of the accumulation of allergenic proteins and to develop new hypoallergenic varieties [18]. The comprehensive study of the protein expression in a given species, often referred as proteomics, can also be applied to the study of the allergenic proteins. Although trends in food allergy research are increasingly focusing on mass spectrometry-based proteomics (sometimes referred as "allergenomics") [19], examples of its application to an in-depth study of tomato allergens [20] are still quite scarce in the literature.

In this study using two pools of sera of allergic people coming from different Italian regions (Campania and Emilia Romagna), twelve tomato ecotypes were screened by a proteomic approach, in order to identify the major allergens involved and to evaluate differences in IgE binding properties of these cultivars.

Materials and methods

Plant material

Twelve tomatoes accessions were used for the analysis: 1: 'Ventura D' (VD), 2: 'Tondino D' (determinate growth habit) (TD), 3: 'Principe Borghese I' (indeterminate growth habit) (PB), 4: 'Sorrento Globoso Rosato' (SGRI), 5: 'San Marzano' (accession SMMu) (SMMU), 6: 'Pisanello' (PS), 7: 'Tondo Liscio' (TLI), 8: 'Tondino I' (indeterminate growth habit) (TI), 9: 'San Marzano' (accession SMC) (SMC), 10: 'Sorrento Rosato' (STLR), 11: 'Principe Borghese D' (determinate growth habit) (PBD), and 12: 'San Marzano' (SMMo) (SMMO). They were grown in a breeding farm in Sarno, Salerno (Southern Italy), over a spring–summer growing cycle. They were harvested during summer 2008 and frozen at −20 °C until the analysis time.

Chemicals

Acetone, diethyl ether, sodium chloride (NaCl), disodium hydrogen phosphate (Na_2HPO_4), glacial acetic acid (AAC), acetonitrile (ACN) and formic acid (FA) were purchased from Carlo Erba Reagents (Rodano, Mi, Italy). Potassium dihydrogen phosphate (KH_2PO_4), potassium chloride (KCl), dithiothreitol (DTT), 2-iodoacetamide (IAA) and modified trypsin from porcine pancreas were purchased from Sigma–Aldrich (Sigma, St. Louis, MO, USA). Ammonium hydrogen carbonate (NH_4HCO_3) was purchased from Fluka (Sigma, St. Louis, MO, USA).

Tomatoes protein extraction

A total of 50 g of fresh tomato (including peels, pulps and seeds) was homogenized by means of Ultraturrax T50 (IKA Werke, Germany) in cold acetone, and the proteins were allowed to precipitate at −20 °C overnight Pellets were washed twice with cold acetone and once with cold acetone/diethyl ether (1:1). Then, the dry powder was extracted in a phosphate-buffered saline solution (PBS, 0.01 M Na_2HPO_4, 0.094 M NaCl, 0.0015 M KH_2PO_4 and 0.0027 M KCl, pH 7.4) for 1 h at 4 °C under continuous stirring, after the neutralization of the pH by an NaOH solution. After centrifugation (8,334*g* three times for 45′), the supernatant liquid was recovered and protein concentrations were determined using the Qubit fluorometer (Invitrogen, UK) according to the instructions by the manufacturer. The extracts were frozen at −20 °C until the analysis time.

Patients' sera

Human sera were provided by the Department of Medical Clinics, Nephrology, and Prevention Sciences of the University of Parma and by the Department of Clinical Medicine and Cardiovascular and Immunological Sciences of the University of Naples "Federico II". They were collected from patients with a clinical history of allergic reactions toward fresh tomato and/or tomato products. All subjects had positive skin prick test for tomato and specific IgE (>0.7 kU/L) detected at ImmunoCAP dosage.

Sera from each group were pooled together, in order to screen IgE binding pattern of major tomato allergens.

Electrophoresis and IgE immunoblotting

Proteins extracts (30 µg or 1 mg in the case of preparative electrophoresis) were separated by SDS-PAGE under reducing conditions by means of Criterion XT precast gel (12% Bis–Tris) in a Criterion Cell, according to the instructions by the manufacturer (Bio-Rad, Munich, Germany).

For the 2D electrophoresis, at first, protein extracts were ultracentrifuged on an Amicon-ultra-4 centrifugal filter device with a nominal molecular cut-off of 5 kDa in order to remove salts. For the first dimension, isoelectric focusing (IEF) was done using IPG precast gel strips (pH 3–10, 7 cm) according to the manufacturer's instructions under reducing conditions, and for the second dimension, SDS-PAGE was done as described above.

After separations, a part of the gel was stained with Coomassie Blue R-250. For immunoblot analysis, proteins were transferred onto 0.2-µm polyvinylidene fluoride (PVDF) membranes (Bio-Rad, Munich, Germany) by tank blotting using the Bio-Rad Trans blot transfer cell for 1 h at 100 V. After blocking in 3% bovine serum albumin, PBS (prepared as described above) and 0.1% Tween 20 for 1 h, the membranes were incubated overnight with single or pooled sera. The membranes were then incubated with rabbit anti-human antibodies and afterward with goat anti-rabbit antibodies linked to HRP. Allergenic protein detection was achieved by incubation of the blotted membranes with the Opti-4-CN substrate (Bio-Rad, Munich, Germany) prepared following manufacturer instructions.

Characterization of tomato allergens by bottom-up approach

Trypsin in-gel digestion

The Coomassie Blue R-250 stained gel spots/bands were excised from gel and destained with 50% ACN/25 mM NH_4HCO_3. Gel slides were shrunk by addition of ACN, and proteins were reduced by addition of 10 mM DDT in 100 mM NH_4HCO_3 for 1 h at 56 °C and alkylated by addition of 55 mM IAA in 100 mM NH_4HCO_3 for 45′ at room temperature. Then, gel slides were swelled and shrunk twice by addition of 100 mM NH_4HCO_3 and 100% ACN, respectively, and then were dried under a nitrogen stream. Trypsin stock solution (100 ng/µL in 1% acetic acid) was diluted 1:10 in 25 mM NH_4HCO_3 and added to the gel slides. After incubation at 37 °C for 16 h, the supernatant was removed and the gel pieces were washed twice with 50% ACN/25 mM NH_4HCO_3 for 15′ and once with ACN and again the supernatant was removed. The combined supernatants were desalting by Zip-TipC_{18} pipette tips (Millipore, USA) according to the manufacturer's instructions and evaporated under a nitrogen stream.

LTQ-orbitrap analysis

Peptide analysis was performed with a Dionex Ultimate 3000 micro HPLC coupled with the LTQ-Orbitrap Thermo Scientific™ mass spectrometer (Thermo Electron Corporation) equipped with a conventional ESI source. The source parameter was configured as follow: spray voltage 3.5 kV, capillary voltage 49 V and tube lens 75 V. For the chromatography separation, a Jupiter 4U Proteo (90 Å, 300 µm × 15 cm) column was used, and the column oven temperature was set to 25 °C; the separation was run for 82 min using a gradient of 99.8/0.2 H_2O/HCOOH (eluent A) and 99.8/0.2 ACN/HCOOH (eluent B) and a flow rate of 5 µL/min. The gradient was run as follows: 0–4 min 95%A and 5%B, then to 50%A at 60 min, and 10% A and 90%B at 62, 62–72 min 10% A, followed by the re-equilibration of the column. For MS1 scans, the Orbitrap resolution was 60.000 and the ion population 5×10^5, with an m/z window from 200 to 1,800. For MS/MS in the LTQ, the population ion was 3×10^4 (isolation width of 3 m/z unit). 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) in all cases (measures by CIM-Parma, Italy) and collected by Xcalibur® software (Thermo Fisher Scientific Inc.).

MALDI mass spectrometric analysis

MALDI MS and MS/MS experiments were carried out on a 4700 Applied Biosystem Proteomics Analyzer. Each spectrum was taken by the following procedure: 1 μL of the desalted sample is spotted on the target plate and, immediately after, 1 μL of matrix solution (α-cyano-4-hydroxycinnamic acid, 10 mg/mL in 50% CH_3CN 50% TFA 0.1%) is spotted over it and 1 μL of sample was spotted again over it. After evaporation has occurred, the target is ready for the analysis. Mass spectrum acquisition was performed in positive ion reflectron mode by accumulating 625 shots/spectrum. The accelerating voltage was 20 kV. External mass calibration was made with mass peptide standards (Sigma). Data were acquired by using 4000 Series Explorer 3.5 v and analyzed by Data Explorer v 4.9.

Results and discussion

Tomato genetic resources

A group of twelve tomato local varieties, both indeterminate and determinate in growth habits, were selected, differing in fruit shape and size and characterized by different final destination (fresh markets or canning) as a representative pool of tomato samples presently spread in local markets. Moreover, among the different cultivated forms of 'San Marzano' grown in the 'San Marzano' PDO area [21], we selected three accessions that are referred to in this study as 'SMMu', 'SMC' and 'SMMo' as they are presently in evaluation for the inclusion in the PDO disciplinary.

Sera of allergic subjects

Human sera used in this work were collected from patients with a clinical history of allergic reactions toward fresh tomato and/or tomato products. Allergic subjects were recruited from two Italian regions, Campania and Emilia Romagna, in order to establish whether environmental conditions could affect sensitization toward different allergenic proteins.

In Parma (North Italy), a total of five human sera were provided by Allergy and Clinical Immunology Center of the University Hospital from allergic subjects with Oral Syndrome after introduction of tomato and/or tomato products. All subjects had positive skin prick test for tomato and specific IgE ranging from 0.70 to 17.49 kU/L at ImmunoCAP dosage. All patients were also sensitized to other food allergens.

In Naples (South Italy), a total of six human serum samples were provided by the Division of Allergy and Clinical Immunology of the University Hospital 'Federico II', from allergic patients with food allergy to tomato. All subjects had positive skin prick test to tomato. Serum concentrations of IgE to tomato antigen ranged from 0.5 to 2 kU/L. Five out of six of these patients were also sensitized to other food allergens (mostly nuts, peanuts, peach, soybean or sesame seeds.)

After some of them had been tested individually, sera belonging to each geographical group were blended together before immunoblotting experiments, thus using them as technological means to detect IgE binding pattern of major tomato allergens.

Immunoblotting of tomato fruit extract with human sera

In order to define which proteins in tomato fruit extracts were recognized by the sera of allergic subjects and whether differences were present among the different tomato ecotypes, protein extracts, obtained and quantified as reported in Materials and methods section, were separated by mono-dimensional SDS-PAGE, by allowing two lanes for every single sample (Fig. 1a). After separation, for every single ecotype, one lane in the gel was stained with Coomassie brilliant blue, whereas the other one was used in order to transfer the separated proteins onto a PVDF membrane. After incubation with pooled sera of patients from Campania region, a protein with a molecular mass of about 14 kDa was detected as the main allergen. No other nonspecific band was detected, even in correspondence with the most intense bands on the gel, indicating a greater expression of those proteins. The immunoblot profiles of tomato ecotypes are shown in Fig. 1b. Quite interestingly, the IgE binding activity of the ecotypes 'Tondino D' (1), 'Ventura D' (2) and 'Pisanello' (6) was less intense, showing a potential reduced allergenicity for these individuals toward this tomato 14-kDa allergenic protein (see next paragraph).

On the other hand, incubation of PVDF membrane with pooled serum sample collected from patients of the Emilia Romagna region did not confirm the potential hypoallergenicity of those ecotypes, since a totally different immunological response was obtained. In this case, a protein with molecular mass of about 45 kDa reacted toward the pooled sera of the allergic subjects, and scarce differences in the IgE binding properties of tomato ecotypes were observed (Fig. 1c). The same results were also obtained using a different pool of sera of allergic patients, selected in the same way as the previous one (data not shown).

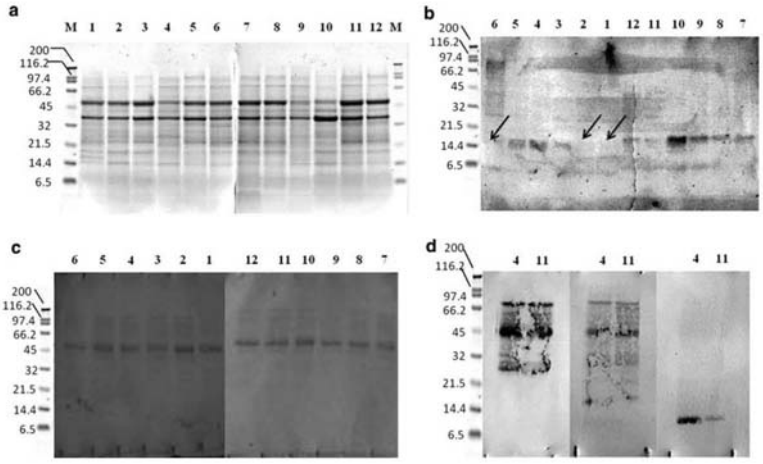


Fig. 1 SDS-PAGE of twelve tomato ecotypes' fruit extracts. Numbers indicate ecotypes as described in "Materials and methods". The lane marked as 'M' represents the molecular weight standards, and their relative molecular masses are reported on the right side. **a** staining with Coomassie brilliant blue; **b** immunoblotting using pooled sera of patients from Campania region; *arrows* show the absence of reactive bands; **c** immunoblotting using pooled sera of patients from Emilia Romagna region; **d** immunoblotting using three sera of patients from Emilia Romagna region, incubated individually with two ecotypes. See text for details

The incubation of the same protein extracts with single serum of patients from Emilia Romagna region only partially confirmed this result. As shown in Fig. 1d, two of these sera, used individually and incubated with the protein extract of two ecotypes, revealed the same 45-kDa reactive protein, as the pool did, while the 14-kDa allergenic protein, observed in previous experiment, was detected using the serum of another patient.

These results showed that the IgE binding patterns of these tomatoes ecotypes were highly serum-specific: Anyway, the allergenic profiles may differ when immunoblottings with single serum are carried out. Since pools of sera contain all the involved IgE, the antigens which are recognized frequently by the single serum are mostly evidenced, quenching the detection of those less recurrent. Anyway, exactly for those reasons, when the most important allergens are to be evidenced, the use of pools of sera might be a very useful technological tool to assess the general allergenicity of a food commodity.

Hence, it could be assumed that the potential hypoallergenicity of tomatoes cannot be generalized for all the allergic subjects, but it should be assessed according to the specific allergen to which every single allergic subject is sensitized.

Identification of the 14-kDa allergenic protein

In order to identify the protein involved in specific IgE binding recognition, and also to possibly gain more evidences on the reason of the low IgE reactivity of some ecotypes, the bands on the gel corresponding to the reactive protein were in-gel-digested by trypsin and the peptide mixture was separated by HPLC and analyzed by LTQ-Orbitrap mass spectrometer. Analysis was carried out on ten ecotypes, including those which showed a lower immunological response. Molecular masses and sequences of the recovered peptides (listed in Table 1) were used to query UniProt database, restricted to *Solanum lycopersicum* L. proteome, and the corresponding protein was identified as profilin (*Lyc e 1*). Among these, diagnostic peptides were also searched to elucidate which of the three profilin isoforms reported in Uniprot database was involved in allergic reactions.

As expected, marker peptides belonging to profilin-1 (SWISS Protein database accession no. Q41344) corresponding to the sequences AEETNIMK (46–54), DFDEPGHLA PTGLFLAGTK (55–73), YMVIQGEPGAVIR (74–86), TAQALIFGVYEEPVTPGQCNMVVEK (99–123) and IGDYLVDQGY (124–133) were not found, since this isoform is expressed in a pollen-specific manner in tomato [22]. Profilin and profilin-2 isoforms (SWISS Protein database accession no. Q8VWR0 and Q93YG7, respectively) were both found in all ecotypes but one, 'Pisanello' (PS) (quite interestingly, one of those showing low IgE reactivity) which showed only peptides belonging to Q8VWR0 isoform (Fig. 2). The low percentage of sequence coverage

Table 1 Identified peptides originated from 14-kDa allergen digestion and LTQ-Orbitrap analysis

SWISS Protein Database Accession Number	Sequence	Position	Calculated MW (Da)[c]	Observed MW (Da)	MS spectra characteristic ions (m/z)
Q8VWR0	HMW peptide	1–71	7,671.6	(discussed in text)	–
	YM*VIQGE**A**G**A**VIR[a]	72–84	1,421.7	1,421.7	1,112.9 (y_{11}), 1,013.3 (y_{10}), 900.4 (y_9), 772.4 (y_8), 586.4 (y_6), 515.3 (y_5)
	GAGGITVK	88–95	701.4	701.4	574.4 (y_6), 517.2 (y_5), 460.2 (y_4), 246.2 (y_2), 147.3 (y_1)
	GAGGITVKK	88–96	829.5	829.5	773.5 (y_8), 588.4 (y_5), 475.3 (y_4), 275 (y_2)
	KGAGGITVK	87–95	829.5	829.5	574.4(y_6), 517.3(y_5), 460.3 (y_4), 246.2 (y_2)
	TNQALIIGIYDEPMTPGQC*NMIVER[b]	97–121	2,862.4	2,862.4	1,660.5 (b_{15}), 1,559.5 (b_{14}), 1,331.3 (b_{12}), 924.2 (b_9), 641.2 (b_6)
	LGDYIIEQGL	122–131	1,119.6	1,119.6	989.0 (b_9), 932.1 (b_8), 804.4 (b_7), 675.3 (b_6), 561.9 (b_5)
Q93YG7	HMW peptide	1–71	7,678.7	(discussed in text)	–
	YMVIQGE**P**EAVIR	72–84	1,503.8	1,503.8	1,210.5 (y_{11}), 1,111.5 (y_{10}), 998.4 (y_9), 870.4 (y_8), 684.3 (y_6)
	GPGGITIK	88–95	741.4	741.4	588.3 (y_6), 530.8 (y_5), 474.4 (y_4)
	GPGGITIKK	88–96	869.5	869.5	724.4 (b_8), 382.2 (b_5)
	TNQALIIGIYDEPMTPGQC*NMIVER[b]	97–121	2,862.4	2,862.4	1,660.5 (b_{15}), 1,559.6 (b_{14}), 1,331.1 (b_{12}), 924.3 (b_9), 754.2 (b_7), 415.2 (b_4)
	LGDYLIEQSL	122–131	1,149.6	1,149.6	1,019.2 (b_9), 932.3 (b_8), 804.3 (b_7), 675.1 (b_6)

Bold letters represent amino acids residues that are different in homologous peptides of profilin isoforms

[a] M* indicates an oxidized methionine

[b] C* indicates a carboxamidomethylcysteine

[c] Monoisotopic MW

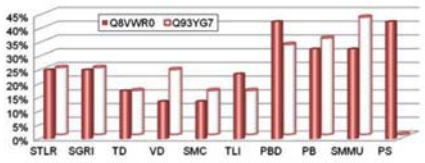


Fig. 2 Percentage of sequence coverage among ecotypes for profilin Q8VWR0 and Q93YG7 isoforms calculated for the analyzed ecotypes, determined by the abundance of peptides identified for each of them

(average among ecotypes was estimated 27% for Q8VWR0 isoform and 23% for Q93YG7 isoform) was likely due to protein's primary structure, poor in positively charged amino acids and thus not easily digested by trypsin. As shown in Fig. 3, the first suitable cleavage site for trypsin on Lys_{71} generates high molecular weight peptides, which are difficult to elute from gel and detected by mass spectrometry. Moreover, in the same peptide, Lys_{43} was not accessible by the enzyme since it is followed by a proline residue, except in the Q41344 isoform in which Ala_{46} and Lys_{54} would allow to generate two additional diagnostic peptides for this isoform discrimination. So, although these two HMW peptides could be useful to discriminate profilin isoforms since they differ for some amino acids and, subsequently, for their molecular masses, they were not found in any ecotypes.

High-resolution tandem mass spectrometry also allowed to identify the sequences of two isobaric peptides of Q8VWR0 isoform, GAGGITVKK and KGAGGITVK (88–96 and 87–95, respectively), which were generated by the cleavage of trypsin on two consecutive lysines at both

a Initiator methionine is cleaved from the mature protein

Fig. 3 Sequence similarity of tomato profilin isoforms: Q8VWR0, Q93YG7 and Q41344. Cutting sites for trypsin are highlighted

Table 2 Identified peptides originated from 45-kDa allergen digestion and LTQ-Orbitrap analysis

Sequence	Protein accession	Calculated MW (Da)[b]	Observed MW (Da)	MS spectra characteristic ions (m/z)
AVVDSAIDAETR	P15003	1,245.6	1,245.6	1,076.5 (y_{10}), 977.4 (y_9), 862.4 (y_8), 775.4 (y_7), 704.4 (y_6), 591.3 (y_5), 476.3 (y_4), 405.3 (y_3), 276.2 (y_2)
LGGQTYSVALGR	P15003	1,220.7	1,220.7	1,108.6 (y_{11}), 1,051.6 (y_{10}), 866.5 (y_8), 765.5 (y_7), 602.5 (y_6), 515.4 (y_5), 416.3 (y_4), 345.2 (y_3), 232.2 (y_2)
VGADMSVINR	Q96577	1,060.5	1,060.6	962.5 (y_8), 905.5 (y_8), 834.4 (y_7), 719.4 (y_6), 588.4 (y_5), 501.4 (y_4), 402.3 (y_3), 289.1 (y_2)
LTSDDDFFTNPM*VK[a]	Q8L5J1	1,644.7	1,644.8	1,431.5 (y_{12}), 1,344.4 (y_{11}), 1,229.4 (y_{10}), 1,114.5 (y_9), 999.4 (y_8), 852.4 (y_7), 705.4 (y_6), 604.4 (y_5), 490.3 (y_4)

[a] M* indicates an oxidized methionine

[b] Monoisotopic MW

N-terminal and C-terminal amino acid sequences, by detecting characteristic ions for each sequence, therefore allowing the unequivocal discrimination of these two peptides. Although the fact that one of the low IgE-responding ecotype did not show the presence of one isoform, this feature was not present in the other two non-IgE binding isoforms, and thus no clear correlation was found between the distribution of the detected profilin isoforms and the weaker immunological response of 'Tondino D', 'Ventura D' and 'Pisanello' ecotypes. Probably, the reduced immunological reactivity can depend not only on the isoform distribution but also on a natural downregulation in the expression of profilin.

Identification of the 45-kDa allergenic protein

The same procedure presented above was applied for the identification of the 45-kDa allergen. Anyway, in this case, after HPLC separation, LTQ-Orbitrap analysis of the tryptic digest of the band with molecular mass of 45 kDa recognized peptides belonging to three different tomato proteins, whose calculated molecular masses were all consistent with the observed one (Table 2). Peptides AVVDSAIDAETR and LGGQTYSVALGR were assigned to tomato suberization-associated anionic peroxidase (SWISS Protein database accession no. P15003). These sequences represent two diagnostic fragments in the discrimination of two known tomato peroxidase isoforms, whose molecular masses are very similar (Protein database accession no. P15003 and P15004): the first differed from the analogous peptide in two alanine residues (Ala_1 and Ala_9, Gly_1 and Asn_9, respectively), while the latter showed a Ser_7 in place of Thr_7.

Peptide VGADMSVINR was found to be shared by three tomato pectinesterase isoforms (SWISS Protein database accession no. P14280, Q96576 and Q96577); according to their calculated molecular masses, the Q96577 protein seemed to fit better the molecular weight observed on gel.

Finally, for peptide LTSDDDFFTNPMVK, the Protein BLAST software only found a perfect match with tomato mannan endo-1,4-beta-mannosidase (SWISS Protein database accession no. Q8L5J1).

Weangsripanaval et al. [23] have already found peroxidase to be the major tomato allergen recognized by the sera of patients suspected to suffer from food allergies and diagnosed to be atopic dermatitis. A different pectinesterase isoform has already been observed as tomato allergens by Kondo et al. [11] (SWISS Protein database accession no. P14280) in patients with OAS after ingestion of fresh tomato fruits. As regards mannan endo-1,4-beta-mannosidase, no

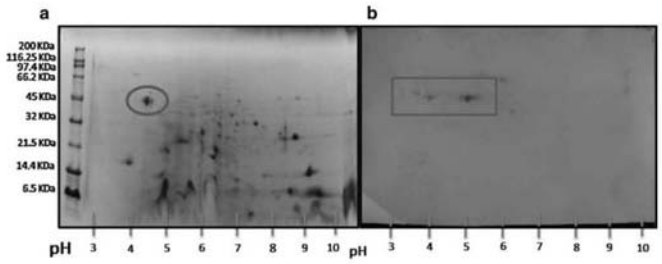


Fig. 4 **a** 2D SDS-PAGE of protein total extract of Sorrento Tondo Liscio Rosato Indeterminato ecotype; **b** corresponding immunoblotting on PVDF membrane using patients' sera from Emilia Romagna region. The *square* and the *circle* indicate the reactive spot and the related protein after electrophoresis, respectively

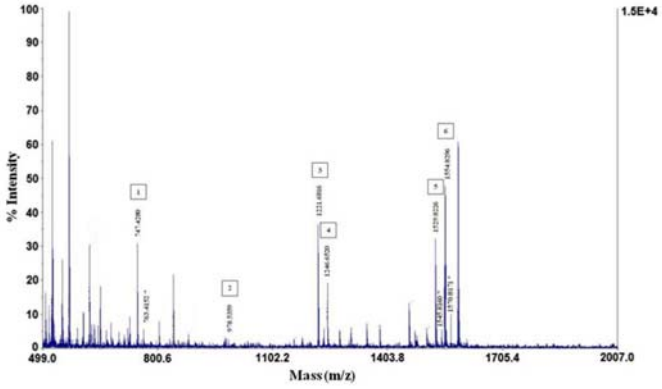


Fig. 5 MALDI-TOF spectrum of 2D reactive spot at 45 kDa, after in-gel trypsin digestion. Numbered peaks were further characterized by MS/MS analysis, in order to confirm the expected sequence suggested by in silico analysis. Ions marked with a star correspond to the previously identified peptide plus an oxidized methionine

evidence of induction of any allergic reaction toward tomato has so far been reported. So, in order to clarify which of the identified proteins was involved in triggering the immunological response, we exploited their different p*I* for a separation of tomato protein total extracts by 2D PAGE (Fig. 4a). Immunoblotting on the bidimensional gel electrophoresis, performed on 'Sorrento Rosato' ecotype, indicated a reactive spot at 45 kDa with a p*I* of about 4 (Fig. 4b), thus excluding the possibility that pectinesterase (whose isoelectric points range from 8.2 to 8.5, depending on the isoforms) [24] and mannan endo-1,4-beta-mannosidase (with an observed isoelectric point of 9) [25] were involved in the allergic reaction.

In order to definitely confirm the identity of the allergenic protein, the corresponding spot was subjected to in-gel trypsin digestion and the tryptic mixture, analyzed by MALDI-TOF spectrometer and revealed peptides belonging to tomato suberization-associated anionic peroxidase (SWISS Protein database accession no. P15003). MALDI-TOF spectrum is shown in Fig. 5. Amino acid sequences of these peptides, derived by in silico digestion of this protein, were further confirmed by MS/MS analysis (Table 3; Fig. 6).

Conclusions

Pooled sera of allergic subjects from two different parts of Italy, Campania and Emilia Romagna regions, were used in order to assess the allergenicity potential of 12 different tomato ecotypes. The ecotypes were assessed by a proteomic approach, performing immunoblotting experiments and identifying the reactive proteins by high-resolution mass spectrometric techniques. Quite interestingly, the two pools showed a totally different immunological response: for the first group, the main allergen was identified as profilin, whereas for the second one the main allergen was identified as suberization-associated anionic peroxidase.

Moreover, experiments by using single sera were performed, further outlining the individual response to different allergenic proteins.

Table 3 Peptides originated from 45-kDa spot digestion identified by MALDI-TOF–TOF analysis

Sequence	Position	Calculated MW (Da)[b]	Observed MW (Da)	MS spectra characteristic ions (m/z)
AVVDSAIDAETR	88–99	1,245.6	1,245.7	862.4 (y_8), 476.3 (y_4), 276.2 (y_2), 175.1 (y_1)
MGASLIR	100–106	746.4	746.4	432.2 (a_5), 460.3 (b_5), 347.2 (b_4), 260.2 (b_3), 271.2 (z_2), 175.1 (y_1)
GYEVIAQAK	145–153	977.5	977.5	978.5 (MH^+)
LGGQTYSVALGR	182–193	1,220.7	1,220.7	765.9 (y_7), 602.4 (y_6), 515.4 (y_5), 416.3 (y_4), 232.2 (y_2), 175.1 (y_1)
EMVALAGAHTVGFAR	231–245	1,528.8	1,528.8	858.5 (y_8), 549.4 (y_5), 450.4 (y_4), 175.1 (y_1)
MGDLPPSAGAQLEIR	336–350	1,553.8	1,553.8	1,251.6 (y_{12}), 1,138.6 (y_{11}), 1,041.5 (y_{10}), 786.5 (y_7), 658.4 (y_5), 288.2 (y_2), 175.1 (y_1)

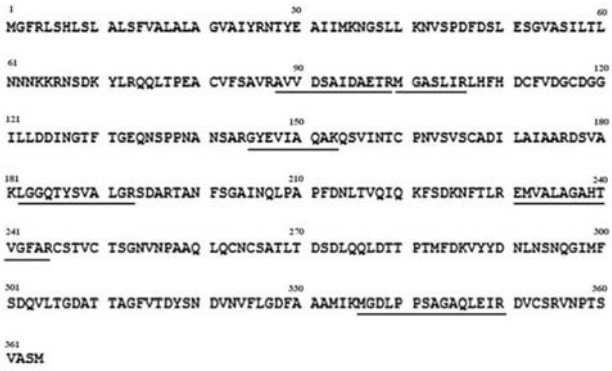


Fig. 6 Amino acid sequence of suberization-associated anionic peroxidase protein, including signal peptide (1–25), identified by MALDI-TOF spectrometer. The *lines* mark peptides identified by MALDI MS/MS analysis

Three different ecotypes showed a low IgE response by using pooled sera of subjects from Campania region, whereas no hypoallergenicity was observed by using pooled sera of subjects from Emilia Romagna region. The hypoallergenicity observed might be due to several reasons: lower expression of the proteins in the different ecotypes, or the presence of protein isoforms having lower allergenicity (although specific experiments aimed at detecting different profilin isoforms outlined that all varieties seemed to have two different profilin isoforms). In any case, it is quite evident that the property of hypoallergenicity for a defined variety, given this scenario, is strictly dependent on which protein is responsible for the allergen response in first instance. For people sensitized to profilin, according to our experiments, some varieties showed a diminished IgE response (and thus might be hypoallergenic), but for people sensitized to suberization-associated anionic peroxidase, the very same varieties were not hypoallergenic at all. These results underline the fact that allergies to a defined food in different subjects can often rely on a completely different immunological response at the molecular level. Although for some subjects hypoallergenic tomato varieties might eventually be found, generally hypoallergenic tomatoes, whose hypoallergenicity is valid for all consumers, seem to be very difficult to obtain, since the characteristic molecular responses of the single allergic subjects are always to be taken into account.

References

1. Sicherer SH, Sampson HA (2009) Food allergy: recent advances in pathophysiology and treatment. Ann Rev Med 60:261–277
2. COMMISSION DIRECTIVE 2007/68/EC of 27 November 2007 amending Annex IIIa to Directive 2000/13/EC of the European Parliament and of the Council as regards certain food ingredients (2007) Official Journal of the European Union. http://eur-lex.europa.eu
3. Osterballe M, Hansen TK, Mortz CG, Bindslev-Jensen C (2005) The clinical relevance of sensitization to pollen-related fruits and vegetables in unselected pollen-sensitized adults. Allergy 60:218–225
4. Vieths S, Scheurer S, Ballmer-Weber B (2002) Current understanding of cross-reactivity of food allergens and pollen. Ann NY Acad Sci 964:47–68

5. Ebner C, Hirschwehr R, Bauer L, Breiteneder H, Valenta R, Ebner H, Kraft D, Scheiner O (1995) Identification of allergens in fruits and vegetables: IgE cross-reactivities with the important birch pollen allergens Bet v 1 and Bet v 2 (birch profilin). J Allergy Clin Immunol 95:962–969
6. Westphal S, Kempf W, Foetisch K, Retzek M, Vieths S, Scheurer S (2004) Tomato profilin *Lyc e 1*: IgE cross-reactivity and allergenic potency. Allergy 59:526–532
7. De Martino M, Novembre E, Cozza G, De Marco A, Bonazza P, Vierucci A (1988) Sensitivity to tomato and peanut allergens in children monosensitized to grass pollen. Allergy 43:206–213
8. Beezhold DH, Sussman GL, Liss GM, Chang NS (1996) Latex allergy can induce clinical reactions to specific foods. Clin Exp Allergy 26:416–422
9. Westphal S, Kolarich D, Foetisch K, Lauer I, Altmann F, Conti A, Crespo JF, Rodriguez J, Enrique E, Vieths S, Scheurer S (2003) Molecular characterization and allergenic activity of Lyc e 2 (b-fructofuranosidase), a glycosylated allergen of tomato. Eur J Biochem 270:1327–1337
10. Le LQ, Lorenz Y, Scheurer S, Fötisch K, Enrique E, Bartra J, Biemelt S, Vieths S, Sonnewald U (2006) Design of tomato fruits with reduced allergenicity by dsRNAi-mediated inhibition of ns-LTP (*Lyc e 3*) expression. Plant Biotechnol J 4:231–242
11. Kondo Y, Urisu A, Tokuda R (2001) Identification and characterization of the allergens in the tomato fruit. Int Arch Allergy Immunol 126:294–299
12. Asero R, Mistrello G, Roncarolo D, Amato S, Arcidiacono R, Fortunato D (2008) Detection of a novel allergen in raw tomato. J Investig Allergol Clin Immunol 18:397–400
13. Bassler OY, Weiss J, Wienkoop S, Lehmann K, Scheler C, Dolle S, Schwarz D, Franken P, George E, Worm M, Weckwerth W (2009) Evidence for novel tomato seed allergens: IgE-reactive legumin and vicilin proteins identified by multidimensional protein fractionation-mass spectrometry and in silico epitope modeling. J Proteome Res 8:1111–1122
14. Germini A, Paschke A, Marchelli R (2007) Preliminary studies on the effect of processing on the IgE reactivity of tomato products. J Sci Food Agric 87:660–667
15. Pravettoni V, Primavesi L, Farioli L, Brenna OV, Pompei C, Conti A, Scibilia J, Piantanida M, Mascheri A, Pastorello EA (2009) Tomato allergy: detection of IgE-binding lipid transfer proteins in tomato derivatives and in fresh tomato peel, pulp, and seeds. J Agric Food Chem 57:10749–10754
16. Le LQ, Mahler V, Lorenz Y, Scheurer S, Biemelt S, Vieths S, Sonnewald U (2006) Reduced allergenicity of tomato fruits harvested from *Lyc e 1*–silenced transgenic tomato plants. J Allergy Clin Immunol 118:1176–1183
17. Breiteneder H, Radauer C (2004) A classification of plant food allergens. J Allergy Clin Immunol 113:821–830
18. Kitagawa M, Moriyama T, Ito H, Ozasa S, Adachi A, Yasuda J, Ookura T, Inakuma T, Kasumi T, Ishiguro Y, Ito Y (2006) Reduction of allergenic proteins by the effect of the *ripening inhibitor* (*rin*) mutant gene in an F1 hybrid of the *rin* mutant tomato. Biosci Biotechnol Biochem 70:1227–1233
19. Monaci L, Visconti A (2009) Mass spectrometry-based proteomics methods for analysis of food allergens. Trends Anal Chem 28:581–591
20. Yagami T, Haishima Y, Tsuchiya T, Tomitaka-Yagami A, Kano H, Matsunaga K (2004) Proteomic analysis of putative latex allergens. Int Arch Allergy Immunol 135:3–11
21. Rao R, Corrado G, Bianchi M, Di Mauro A (2006) (GATA)4 DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. Plant Breed 125:173–176
22. Yu LX, Nasrallah J, Valenta R, Parthasarathy MV (1998) Molecular cloning and mRNA localization of tomato pollen profilin. Plant Mol Biol 36:699–707
23. Weangsripanaval T, Nomura N, Moriyama T, Ohta N, Ogawa T (2003) Identification of suberization-associated anionic peroxidase as a possible allergenic protein from tomato. Biosci Biotechnol Biochem 67:1299–1304
24. Gaffe J, Tieman DM, Handa A (1994) Pectin methylesterase isoforms in tomato. Plant Physiol 105:199–203
25. Carrington CMS, Vendrell M, Domınguez-Puigjaner E (2002) Characterisation of an endo-(1, 4)-b-mannanase (LeMAN4) expressed in ripening tomato fruit. Plant Sci 163:599–606

Curriculum vitae

Mariangela Bencivenni was born in Nocera Inferiore (Salerno, Italy) in 1983. In 2008 she obtained the Master Degree in Biotechnology (cum laude) at the University of Naples Federico II. In the meanwhile, she was accepted as a participant in the training course 'GenoPOM', in the field of genetics applied to plants breeding and certification, at the CNR – Institute of Plants Genetics in Portici (Naples) – where she qualified as Genomics Expert. In order to accomplish this qualification, she carried out an internship at the Department of Organic and Industrial Chemistry at the University of Parma. From 2009 she is attending the Ph.D. course in Food Science and Technology at the same University, where she has also obtained a research grant. Since September 2011 she has been employed at the Tecnopolo 'SITEIA.PARMA – Inderdipartimental center for the safety and the technologies of agro-food industries' in Parma.

The research activity of Mariangela Bencivenni concerned the development of advanced molecular systems for assessing the quality and the safety of plant-derived foods. She studied the allergenic patterns of different tomato cultivars, by SDS-PAGE, Western Blotting and mass spectrometry analysis, in order to compare the different allergenic properties and to identify the major allergens. She also approached the study of gliadin peptides by *in vitro* digestion with gastrointestinal enzymes and mass spectrometry analysis, in order to characterize gliadins' peptidic pattern and to investigate peptides trigging celiac disease. Moreover, she developed PNA-based systems (on surface and in solution) in order to identify tomato varieties by DNA single nucleotide polymorphisms' recognition.

During her academic experience Mariangela Bencivenni gained knowledge in genomics and proteomics field by developing plants and foodstuffs genomic DNA extraction methods, simplex and duplex DNA amplification as well as PNA-mediated amplification methods both in solution (PCR Clamping) and on surface (PNA microarrays); plants proteins extraction methods, mono and bidimensional electrophoresis and Western Blot analysis, in order to detect allergenic proteins. She also improved her skills with with protein purification strategies, by means of Multidimensional Liquid Chromatography and with mass spectrometry, by means of bottom up proteomics approaches, including several types of analyzer (LC/MS, LC/MS/MS, LTQ Orbitrap, Maldi/TOF).

Abroad experience

From February 2011 to July 2011: 'Laboratory of Food Structure and Health' led by Clare Mills at the Institute of Food Research in Norwich (UK).

School and Conferences

– 10th International Symposium on Immunological, Chemical and Clinical Problems of Food Allergens.

Parma, May 26-29, 2008

– Spectrometric identification of Organic Substances in Food – Food Science and Technology course held by Prof. Stefano Sforza.
Parma, 2009

– International Conference Food-Omics, A Science for Nutrition, Health and Wellness in the Post-Genomic Era.
Cesena, May 28-29, 2009

– X scuola sullo studio del Proteoma.
Vitorchiano (Viterbo), December 9-11, 2009

– Fattori Dietetici e Celiachia nel Terzo Millennio; in PastaTrend, il Salone della Pasta.
Bologna, April 26, 2010

– 15th Workshop on the Development in the Italian PhD Research on Food Science Technology and Biotechnology.
Portici (Naples), September 15-17, 2010

– VIII Congresso Nazionale di Chimica degli Alimenti – Qualità e tipicità degli alimenti mediterranei: alimentazione e salute.
Marsala (Trapani), September 20-24, 2010

– 4th International Symposium on Molecular Allergology.
Munich, October 29-31, 2010

– 16th Workshop on the Development in the Italian PhD Research on Food Science Technology and Biotechnology.
Lodi (Milan), September 21-23, 2011

– 2° Workshop MS-J-Day – La spettrometria di Massa nelle Scienze della Vita.
Naples, December 2, 2011

Pubblications in peer-reviewed journals

- A. Manicardi, A. Calabretta, M. Bencivenni, T. Tedeschi, S. Sforza, R. Corradini, R. Marchelli (2010). *Affinity and Selectivity of C2- and C5-Substituted "Chiral-Box" PNA in Solution and on Microarrays*. Chirality **22**:E161-E172.

- T. Tedeschi, A. Calabretta, M. Bencivenni, A. Manicardi, G. Corrado, M. Caramante, R. Corradini, R. Rao, S. Sforza, R. Marchelli (2011). *A PNA Microarray for Tomato Genotyping*. Molecular Biosystem **7:**1902-1907.

- M. Bencivenni, A.Faccini, C. Bottesini, R. Rao, A. Detoraki, E. Ridolo, G. Marone, P.P. Dall'Aglio, A. Dossena, R. M. Marchelli, S. Sforza. (2011). *Assessing allergenicity of different tomato ecotypes by using pooled sera of allergic subjects: identification of the main allergens*. European Food Research and Technology. DOI 10.1007/s00217-011-1640-4.

- B. Prandi, M. Bencivenni, T. Tedeschi, R. Marchelli, A. Dossena, G. Galaverna, S. Stefano. (2012). *Common wheat determination in durum wheat samples through LC/MS analysis of gluten peptides*. Analytical & Bioanalytical Chemistry. Accepted.

Oral Communications

- M. Bencivenni. *Proteomic Tools for the Purification and the Characterization of Tomato Allergen ns-LTP Protein*. Proceedings of 16th Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology. Lodi, September 21-23, 2011; in Proceedings Book: pag 18

- M. Bencivenni, S. Sforza. *Study of tomato allergens by mass spectrometry-based techniques*. Proceedings of 2° Workshop MS-J-Day – La spettrometria di Massa nelle Scienze della Vita. Naples, December 2, 2011; in Proceedings Book: pag 18

Poster Communications

- M. Bencivenni, A. Faccini, S. Sforza, A. Dossena, R. Marchelli. *Proteomic Analysis of Tomato Allergens*. International Conference on FoodOmics: A Science for Nutrition, Health and Wellness in the Post-Genomic Era. Cesena, May 28-29 2009; in Abstract Book: pag 33.

- T. Tedeschi, S. Sforza, A. Calabretta, A. Tonelli, M. Bencivenni, R. Corradini, R. Marchelli. *Dna Analysis By Pna-Based Technologies For Assessing Food Authenticity*. International Conference on FoodOmics: A Science for Nutrition, Health and Wellness in the Post-Genomic Era. Cesena, May 28-29 2009; in Abstract Book: pag 111.

- S. Sforza, T. Tedeschi, A. Calabretta, M. Bencivenni, F. Lambertini, G. Galaverna, A. Dossena, R. Marchelli. *Dna And Peptides As Molecular Markers For Assessing Food*

Authenticity. 4th International Symposium on Recent Advances in Food Analysis. Prague, November 4-6 2009; in Book of abstract: J-5 - pag 447.

- M. Bencivenni. *Allergenicity of Different Tomato Ecotypes*. Proceedings of 15th Workshop on the Development in the Italian PhD Research on Food Science Technology and Biotechnology. Portici (Naples), September 15-17, 2010; in Proceedings Book: pag 213.

- M. Bencivenni, A. Faccini, R. Rao, A. Detoraki, E. Ridolo, G. Marone, P. Dall'Aglio, A. Dossena, S. Sforza, R. Marchelli. *Proprietà Allergeniche di Ecotipi di Pomodoro*. Marsala (Trapani), September 20-24, 2010; in Riassunti: P147 – pag 223.

- M. Bencivenni, T. Tedeschi, A. Calabretta, R. Corradini, G. Corrado, R. Rao, S. Sforza, R. Marchelli. *Tecnologie a PNA per l'Identificazione di Varietà di Pomodoro*. Marsala (Trapani), September 20-24, 2010; in Riassunti: P125 - pag 201.

- M. Bencivenni, A. Faccini, R. Rao, A. Detoraki, E. Ridolo, G. Marone, P. Dall'Aglio, A. Dossena, S. Sforza, R. Marchelli. IgE Binding Properties of Different Tomato Ecotypes. Munich, October 29-31, 2010; in Final Programme and Abstract Book: P73 – pag 61

- B. Prandi, M. Bencivenni, A. Faccini, R. Marchelli, G. Galaverna, S. Sforza. *Peptides Derived from Simulated Gastrointestinal Digestion of the Prolamin Fractions Extracted from Different Wheat Varieties: Implications for the Celiac Disease*. Euro Food Chem XVI- Translating food chemistry into health benefits. Gdansk 6-8 July, 2011; in Conference Proceedings.