Chapter 2

Food Allergy and Seed Storage Proteins

An overview
Allergy is a symptomatic reaction to a normally innocuous environmental antigen (Murphy et al., 2007). It results from the interaction between the antigen and the antibody (or primed T cells) produced by earlier exposure to the same antigen. According to revised nomenclature for allergy by European Academy for Allergy and Clinical Immunology, hypersensitivity reactions are of two types: allergic hypersensitivities, when an immunological mechanism is evident or strongly suspected and non allergic hypersensitivities, when no immunological mechanism is known (e.g. intolerances and adverse food reactions) (Johansson et al., 2001). Hence, allergy is a hypersensitivity reaction initiated by immunological mechanisms. Allergy therefore describes a symptomatic reaction to a normally innocuous environmental antigen that may result from the induction of an immune response. Such reactions are mainly separated into two types: (i) antibody-mediated reactions occurring within minutes of contact with allergen, predominantly involving antibody-mediated mast cell activation and (ii) delayed hypersensitivity reactions involving cell-mediated responses played by sensitized lymphocytes (Rowntree et al., 1985). In the case of antibody-mediated reactions, the antibody typically responsible for the allergic reaction mostly belongs to the immunoglobulin ε isotype (IgE) and these patients are said to suffer from IgE-mediated allergy (Sampson, 2003). Delayed hypersensitivity on the other hand is mediated by T cells and occurs hours to days after exposure to the allergen. Such cell-mediated immunological mechanisms seem to be important in allergic diseases such as contact dermatitis and celiac disease associated with gluten sensitivity (Ritis et al., 1988).

In non IgE mediated allergies, the antibody may belong to IgG isotype, as in anaphylaxis due to immune complexes containing dextran (Hedin et al., 1976). Both
IgG and IgE are found in allergic bronchopulmonary aspergillosis (Patterson et al., 1986). Allergy can also be cell mediated, in which immunologically sensitized lymphocytes play a major role. Cell mediated immunological mechanisms seem to be important in allergic diseases like contact dermatitis and celiac disease (Ciacci et al., 2004).

IgE mediated allergy is also known as type I hypersensitivity or immediate hypersensitivity. The reaction may involve skin (urticaria and eczema), eyes (conjunctivitis), nasopharynx (rhinorrhea, rhinitis), bronchopulmonary tissues (asthma) and GIT (gastroenteritis) (Gould et al., 2003). The reaction usually takes 15 to 30 minutes from the time of exposure to the antigen. The IgE mediated allergic response comprises of three stages, as seen in typical IgE mediated allergic diseases such as asthma, allergic rhinoconjunctivitis and anaphylaxis. On initial exposure, antigen presenting cells present the allergen peptide to TH cells, which then induce B lymphocytes to produce antigen specific IgE. These IgE antibodies bind to high affinity IgE receptors of mast cells and basophils. Re-exposure to the allergen then induces mast cells and basophils to degranulate, releasing preformed inflammatory mediators (histamine, tryptase and heparin) along with newly synthesized mediators (leukotrienes and cytokines) resulting in smooth muscle constriction, vascular permeability, mucous secretion and itching. Late phase reaction is caused by the induced synthesis and release of mediators including prostaglandins, leukotrienes, chemokines and cytokines from activated mast cells. These recruit other leukocytes including eosinophils and TH2 lymphocytes to the site of inflammation. Late phase reactions are associated with a second phase of smooth muscle contraction mediated.
Food Allergy and Seed Storage Protein

by T cells with sustained oedema and with tissue remodelling such as smooth muscle hypertrophy and hyperplasia (Gould and Sutton, 2008).

Both genetic and environmental factors contribute to the development of IgE mediated allergy (Bonini et al., 1994). Very few people show an exaggerated tendency to mount a strong IgE response against antigens that are known to be safe for vast majority of population. This state is called atopy (Johansson et al., 2001). Atopy has a strong familial basis and influenced by several genetic loci. Atopic phenotype is the genetic predisposition to Th2 type immune response and development of classical IgE allergies. Studies on atopic families have identified regions on chromosome 11 (IgE Fc receptor production) and chromosome 5 (IL-4 and IL-5 induced IgE synthesis) (Rosenwasser, 1996).

The prevalence of atopic allergy and of asthma in particular, is increasing in economically advanced regions of the world (Strachan, 2000). This increase in the prevalence of allergic disease in the developed world parallels the reduction in infectious diseases that has resulted from wide spread vaccination, the use of antibiotics and hygiene standards. These observations lead to the formulation of 'hygiene hypothesis', which suggested that less hygienic environments, specifically environments that predispose to infection in early childhood, help to protect against atopy (Wills-Karp et al., 2001). This implies that Th2 responses predominate over Th1 responses by default in early childhood, and the immune system is reprogrammed toward more Th1 dominated responses by cytokine response to early infections. There are evidences for a bias toward Th2 responses in neonates. There are also evidences that the exposure to childhood infections, with the exception of some respiratory infections, help to protect against the development of atopic allergic
diseases (Prioult and Nagler-Anderson, 2005). However, the biggest opposition to the hygiene hypothesis comes from the strong inverse correlation between helminthic infections and the development of allergy. Helminthic infections strongly drive T_{H2} response and an inverse relationship with atopy is very difficult to reconcile with the idea that polarization of T cell response toward T_{H1} is a general mechanism for the protective effect of infection against the development of atopy. This leads to the modification of hygiene hypothesis. The modified hygiene hypothesis proposes that cell types of infection might protect against the development of atopy by the production of regulatory cytokines such as IL-10 and TGF-β which down regulate both T_{H1} and T_{H2} responses (Wills-Karp et al., 2001). Several independent lines of evidences indicate that the function of naturally occurring regulatory T cells (Tregs) is impaired or altered in patients with allergies compared with normal healthy individuals (Sakaguchi, 2004). Tregs are CD4+, CD25+ and FOXP3+ T cells (Maloy and Powrie, 2001). In vitro Tregs have been shown to inhibit proliferation of naive T cells through cell-cell contact. However, in vivo, these cells can also function through induction of inhibitory cytokines such as TGF-β and IL-10. IL-10 modulates many cells and effector functions that are associated with allergic diseases including T_{H2} activation, mast cell and eosinophil function and IgG to IgE ratios. There is an inverse association between IL-10 levels and the severity of asthma and allergic diseases (Hawrylowicz and O'Garra, 2005).

2.1 Food allergy

Food is necessary to provide energy and nutrients for sustaining life. However some individuals undergo unfavourable physiological and morphological reactions after ingestion of foods that are known to be safe for vast majority of
Food Allergy and Seed Storage Protein

population. Adverse food reactions (food hypersensitivities) include any abnormal reaction resulting from the ingestion of a food and might be the result of food intolerances (nonallergic food hypersensitivities) or food hypersensitivity/allergy (food allergy) (Johansson et al., 2001; Sampson, 1999). Food intolerances (nonallergic food hypersensitivities) are adverse responses caused by some unique physiologic characteristic of the host, such as metabolic disorders (eg, lactase deficiency) (Miyajima et al., 1993). Food hypersensitivities/allergies are adverse immunologic reactions to proteins present in food that might be due to IgE- or non-IgE-mediated immune mechanisms. Toxic reactions might mimic food hypersensitivities and typically are due to factors inherent in a food, such as toxic contaminants (eg, histamine in scombroid fish poisoning) (Clark et al., 1999) or pharmacologic substances within the food (eg, tyramine in aged cheeses) (D'Andrea et al., 2007), which can affect most healthy individuals when given in appropriate doses. Food aversions also might mimic adverse food reactions but are not reproducible when the patient ingests the food in a blinded fashion (Sampson, 2004).

2.1.1 Epidemiology of food allergies

Atopic disease is increasingly common worldwide, and food allergy seems to be a part of this rise. Food allergy affects up to 8% of children less than 3 years of age and approximately 2% of adults experience food induced allergic disorder (Bock, 1987; Jansen et al., 1994). Fortunately, most children (about 85%) lose their sensitivity to most allergenic foods (egg, milk, wheat, soya) within the first 3–5 years of life (Sampson, 1999). Sensitivity to peanut, tree nuts, and seafood is rarely lost (Burks and Ballmer-Weber, 2006). Food allergy also appears to be the single leading cause of anaphylaxis (Sampson, 2005). There is no specific therapy available for
food allergy, which is particularly troublesome considering the potential of severe reaction. Dietary avoidance is the only method by which food allergic reactions can be avoided but accidental ingestion of food allergens is common because of ubiquitous presence of certain foods.

The foods most commonly causing breathlessness were hazelnut in Norway, Sweden, and Germany, fruits in Iceland, Belgium, Ireland, and Italy, and peanut in the USA (Sicherer, 2002). This variation in causal foods could be due to different eating patterns, cultural and environmental influences, and genetic factors, and to inclusion of non-immunological reactions. Ninety percent of the food related hypersensitivity reactions could be attributed to eight major types of food: milk, eggs, fish and crustaceans, peanuts, tree nuts, soybeans and wheat (Prioult and Nagler-Anderson, 2005). Overall, the typical allergens of infancy and early childhood are egg, milk, peanut, wheat, and soya, whereas allergens responsible for severe reactions in older children and adults are mainly caused by peanut, tree nuts, and seafood (Lee and Burks, 2006). Allergy to fruits and vegetables is common, but usually not severe (Asero, 1999). A number of factors can affect the development of food allergy including diet and culture, route of exposure, processing, cooking, digestion and the allergen itself (Prioult and Nagler-Anderson, 2005).

2.1.2 Immunopathogenesis of food allergies

There are two different ways by which one can get sensitized to the food allergens, either by the ingestion of food (class 1 food allergens) or through inhalation of the air borne allergens that show cross reactivity with the food allergens (class 2 food allergens). In class 1 food allergy the sensitisation process occurs in
Allergens eliciting class I food allergy are stable to treatment with heat, acid and proteases. The most important class I food allergens are milk, eggs and legumes. Class II food allergies develop as a consequence of an allergic sensitisation to inhalant allergens. The immunological basis of these food allergies is IgE cross reactivity. Most of the class II food allergens are heat labile. According to their behaviour during digestion process, they can cause symptoms ranging from oral allergy syndrome to anaphylactic shock (Sicherer and Sampson, 2006; Breiteneder and Ebner, 2000; Egger et al., 2006).

IgE mediated food allergies target many organs: the skin, manifested as urticaria, angioedema, rashes, and/or flushing; the respiratory tract, as nasal congestion and/or bronchospasm; the GIT, as vomiting, abdominal pain and/or diarrhoea and cardiovascular system resulting in hypotension and/or cardiac arrhythmias. An IgE mediated allergy can also result in anaphylactic shock. With non IgE mediated (or T cell mediated) food allergies, patients can have reactions such as atopic dermatitis, enterocolitis and Heiner syndrome. These patients are skin test negative to the causative food despite their clinical symptoms (Sampson, 2004).

The gastrointestinal tract is the largest immunologic organ in the body, possesses the greatest surface area exposed to the outside environment, and is confronted with the largest antigenic load in the form of dietary proteins, commensal organisms and the pathogens (Mayer, 2003). The mucosal immune system of the gut has the extraordinary ability to distinguish between the foreign pathogens and safe nutritional proteins and commensal organisms. Despite the large extent of dietary antigenic exposure, only a small percentage of individuals have food allergy. This is
due to the development of oral tolerance toward the dietary proteins (Sampson, 2004).

Unlike the systemic immune system, the mucosal immune system is highly efficient in inhibiting any immune response against nondangerous antigens (oral tolerance), while at the same time mounting a strong immune response against any dangerous antigen. However, the mucosal immune system is immature in the infants due to the developmental immaturity of the components of the mucosal barriers of the gut, resulting in the reduced efficacy of infant mucosal immunity, probably responsible for high prevalence of food allergies and gastrointestinal infections during the initial few years of life. Despite the evolution of elegant barrier system, about 2% of ingested food antigens are absorbed and transported through the body in an “immunologically intact” form, even through the mature gut (Husby et al., 1987).

Although intact food allergens penetrate the gastrointestinal tract, they generally do not cause clinical symptoms because most individuals acquire tolerance, believed to be the result of T cell anergy or induction of regulatory T cells. Intestinal epithelial cells play a major role in tolerance induction to food antigens. They act as non-professional antigen presenting cells. In addition dendritic cells residing in the non inflammatory environment of the payers patches express IL10 and IL4 which favours generation of tolerance. Finally T regulatory cells, which are potent source of TGF-β are, generated in mucosal lymphoid tissue in response to low dose antigen and mediate bystander tolerance within GIT. The gut flora also plays a significant role in the induction of oral tolerance (Sampson, 2005; Smith and Nagler-Anderson, 2005; Brandtzaeg, 2002).
In healthy individuals, ingestion of innocuous antigens leads to systemic non responsiveness (oral tolerance). In allergic patients, oral tolerance is not induced and instead an exaggerated \( \text{T}_{\text{H}2} \) response to specific dietary antigen(s) ensues. In \( \text{T}_{\text{H}2} \) biased mucosal cytokine environment antigen specific \( \text{T}_{\text{H}2} \) are generated when antigen is presented to naïve T cells by antigen presenting cells, mainly dendritic cells. Activated \( \text{T}_{\text{H}2} \) cells produce IL-4 and IL-13 which promote IgE production by B cells. Upon subsequent ingestion, antigen presentation leads to a rapid T cell activation and secretion of \( \text{T}_{\text{H}2} \) cytokines, triggering mediator release by eosinophil and basophil. At the same time antigen can interact directly with receptor bound IgE on mast cells resulting in cross linking of receptors, which triggers the release of preformed and newly formed chemical mediators, particularly histamine responsible for clinical symptoms (Sampson, 2005).

2.1.3 Molecular properties of food allergens

Food allergens belong to rather limited number of protein families and are also characterized by a number of biochemical and physiochemical properties, many of which are shared by food allergens of plant and animal origins. Some of these properties are:

1. **Ligand binding**: A number of food allergens are able to bind ligand ranging from metal ions to lipids. Certain ligand such as metal ions become integrated into 3 dimensional structure of proteins, often buried deep within the molecule e.g. parvalbumin (Bugajska-Schretter et al., 1998). Some proteins form a cavity into which a ligand fits; this might be a metal ion, steroid or variety of lipid molecules. Other proteins possess a tunnel into which ligand fits, where as yet
other bind ligand through superficial surface interactions (Breiteneder and Mills, 2005). BLg, which belongs to lipocalin superfamily of extracellular lipid binding proteins, it binds a diverse range of molecules, including retinol and its analogues. The ligands bind in the calyx of the characteristic lipocalin β barrel. Unlike BLg, nsLTPs are generally able to bind lipids in tunnel lined with hydrophobic residues running through the protein. The cavity is highly plastic and can bind a wide range of lipophilic molecules including sphingolipids, prostaglandins, amphotericinB and other hydrophobic drugs (Tassin et al., 1998). Ligand binding can have the overall effect of reducing the mobility of polypeptide backbone, increasing both thermal stability and resistance to proteolysis. The ligand may be required for stabilizing the 3 dimensional structure of the protein. The loss of ligand may disrupt the protein folding, with an increase in polypeptide mobility resulting into a partially folded form (Sampson, 2004). E.g. Loss of calcium from parvalbumin triggers a large change in conformation and an associated loss of conformation dependent IgE epitopes (Bugajska-Schretter et al., 1998).

2. Interaction with membrane and other lipid structures: Many plant food allergens are also able to associate with cell membranes and other type of lipid structures formed in foods. Proteins like thionins, thaumatin like proteins (TLPs), 2S albumins and nsLTPs and some defensins interact with the fungal and bacterial cell membranes resulting in depolarization and leakage. These proteins are referred to as pathogen related proteins (Selitrennikoff, 2001). Many allergenic food proteins, like the whey proteins (α-lactalbumin and βLg), caseins, and various seed proteins such as 2S albumins, 11S and 7S seed storage globulins
interact with lipids present in food to form emulsified structures. When proteins adsorb to a lipid layer in an emulsion, they unfold, revealing hydrophobic regions of the molecules that favour interactions with lipid. The proteins also aggregate to form a 2 dimensional gel like layer that has elastic properties necessary to stabilize oil droplets effectively in an emulsion (Breiteneder and Mills, 2005).

3. **Protein stability and mobility:** No single structural motif is associated with the stability of proteins. However, nature has used different types of strategies to develop stable proteins. One of the structural features clearly related to the stability is disulphide bonds. There are many allergens that are highly disulphide bonded, including the members of prolamin superfamily (nsLTPs, 2S albumin, and inhibitors of trypsin and α-amylase found in cereals, together with the TLPs). 2S albumins are highly thermostable and resistant to extreme pH and proteases as indicated by the 2S albumins of mustard seed (Domínguez et al., 1990) and Brazil nut (Murtagh et al., 2003). Their structural homologues, the nsLTPs, also possess 4 disulphide bonds and are resistant to proteolysis, extreme pH, or thermal treatments, refolding to their native structure on cooling (van Ree, 2002). Similarly, the conformation of TLPs is stabilized by 8 disulphide bonds (Batalia et al., 1996). This extensive disulphide cross-linking renders zeamatin, a TLP from corn, highly resistant to proteolysis. However, absence of disulphide bonds does not indicate a lack of stability because the cupin barrel, for example, is highly stable and possess no disulphide bonds. The cupin barrel is found in 7S and 11S seed storage globulins.

4. **Glycosylation:** The immunologic activity of IgE antibodies directed against the carbohydrate portion of glycoallergens has been a matter of debate since the
discovery of N-glycan specific IgE. Carbohydrate specific IgE was reported to induce *in vitro* basophil histamine release for *Lyc e 2*, a glycoallergen of tomato, as well as for *Api g 5* (Bublin et al., 2003), a glycoallergen from celery tuber. However, in both cases the allergen concentrations needed to achieve mediator release was relatively high. In addition to immunologic implications, glycosylation also affects the biological properties of the allergens (Foetisch et al., 2003). N-glycosylation can have a significant stabilizing effect on protein structure and there is evidence that it increases the stability of, for example the 7S globulin of pea and its resistance to chemical denaturation (Pedrosa et al., 2000). The 7S seed storage globulins are frequently glycosylated. The 7S globulin of the peanut, *Ara h1* is known to be glycosylated and has one asparagine linked consensus carbohydrate addition site, the glycan moieties being a heterogenous mixture of N-glycans including Man$_{5-6}$GlcNAc$_2$ and Man$_{3-4}$XylGlcNAc$_2$ (van Ree et al., 2000). The 11S globulins are rarely glycosylated, an exception being that from lupin, in which the major acidic subunit become glycosylated (Duranti et al., 1995).

5. **Repeated structures, aggregates and glycation:** Both the presence of repetitive structures and a propensity to aggregate, either under physiological conditions or as a result of food processing, might affect sensitisation, atleast by enhancing immunogenicity. Food allergens with repeating structures include the tropomyosin allergens of shell fish and seed storage prolamins. The globulin seed storage proteins also share a propensity to form large structures, existing in a salt dependent equilibrium of trimers, hexamers (for 11S globulins), and dodecamers which are held together by noncovalent interactions. A characteristic of seed
Food Allergy and Seed Storage Protein

storage globulins is their ability to form heat-set gel networks after wet thermal processing, such as boiling. Both 11S and 7S globulins, as with the other members of the cupin superfamily are thermostable. 7S globulins have their major thermal transition at around 70°C to 75°C, and 11S globulins unfold at temperatures greater than 94°C. It seems likely that the cupin barrel remains intact, but the unfolding of other regions of polypeptides results in a loss of quaternary structure, with concomitant formation of large aggregates. Unlike soybeans and lentils, peanuts and other nuts are often subjected to thermal processing at low water levels, such as roasting. This will affect protein stability because protein denaturation requires the presence of water. Thus *Ara h* 1 purified from residual soluble protein only becomes unfolded on roasting peanuts to 140°C for 15 minutes (Koppelman et al., 1999), although much of the peanut protein becomes insoluble. This is probably because of covalent modification of proteins through Millard reactions. This involves sugars reacting with free amino groups on proteins to form Amadori compounds, which might rearrange to produce a range of adducts, known as advanced glycation-glycosylation end products. Indications are that glycation reactions might be responsible for the apparent increase in allergenic activity of peanuts after processes such as curing and roasting (Bugajska-Schretter et al., 1998).

2.1.4 Food allergy and cross reactivity

Immunologic cross-reactivity, which is important in many aspects of host defence and immune-mediated diseases, is a prominent feature of allergic disorders (Bonds et al., 2008). A role of cross-reactivity in food allergy was reported as early as in 1942 with the recognition that individuals sensitized to pollen allergens were
more likely to exhibit allergic symptoms upon eating certain fruits (Tuft and Blumenstein, 1942). Several cross-reactivities have been identified, including food-food, pollen-food, and latex-venom associations. The phenomenon of cross-reactivity is based on IgE recognition of structural similarities between the sensitizing aeroallergens and one or more food proteins, which otherwise might not induce sensitization via gastrointestinal exposure (Bonds et al., 2008). This type of cross-reactivity is most easily understood when the sensitizing and cross-reacting proteins belong to the same family of proteins and are easily recognizable by the similarity in their amino acid sequences (e.g. tropomyosins). When the specific cross-reacting aeroallergen and food proteins have not been identified or structurally characterized the molecular basis of cross-reactivity are more difficult to establish. While most cross-reactivity syndromes are thought to be due to IgE antibodies, there is also support for a role of T cells in cross-reactive food allergy as well (Fritsch et al., 1998; Bohle et al., 2003). This T cell mediated cross-reactivity may be independent of IgE-mediated mechanisms, occurring even in the context of denatured proteins and may have downstream effects on other allergic manifestations such as atopic dermatitis (Bohle, 2007; Bohle et al., 2006).

It appears that IgE antibody is the major player, especially in clinical expression of the phenomenon and most efforts have focused on identifying characteristics of molecules that make them targets for IgE binding (Ferreira et al., 2004). Although the particular characteristics of a molecule which make it likely to induce an IgE response have not been fully delineated, it has become clear that most of the currently defined allergens can be grouped into a limited number of protein families, whose members share some structural commonalities. For some groups, the
IgE cross-reactivity is based on highly similar amino acid sequences (e.g. tropomyosin). Members of other cross-reacting groups may have much less sequence identity, but share a discrete three-dimensional structure, or have similar overall conformation which accounts for their cross-reactivity (e.g. Profilins and PR-10 family) (Ferreira et al., 2004). Yet others induce IgE cross-reactivity because of sharing epitopes based on linear sequence homology as well as others which lack sequence identity but have conformational similarity (e.g. PR-5 and seed storage proteins) (Sankian et al., 2005).

The role of cross-reactive carbohydrate determinants (CCDs) in food allergy continues to be debated. CCDs are widely distributed and structurally conserved in many plants and insects and may even play a role in animal allergy (Adedoyin et al., 2007). Plant proteins with glycans that contain xylose and fucose have been shown to induce IgE responses. The diffuse presence of these glycans means that anticarbohydrate IgEs have been thought to be a candidate for inducing clinical cross-reactivity. CCDs have been repeatedly found to cause IgE binding in vitro.

2.1.5 Food allergens of plant origin

Plant food allergens can be classified into families and superfamilies on the basis of their structural and functional properties. Food allergens of plant origin could be broadly classified into two major classes based on the route of their sensitization: food allergens sensitizing through gastrointestinal tract (type 1 food allergens) and food allergens arising from cross-reactive inhalant allergens (type 2 food allergens).
Type 1 food allergens comprise of seed storage proteins and protease inhibitors that have storage and (or) protective functions. These allergens can be broadly classified into two superfamilies: cupin superfamily and prolamine superfamily. Well characterised allergens from this family include allergens from peanut (Ara h 1, Ara h 3 and Ara h 6), soybean, hazelnut (Cor a 8, Cor a 9), wheat (Tri a 19), rye (Sec c 20) etc. This class have been reviewed in detail under the section pertaining to seed storage proteins.

Type 2 food allergens comprise of more diverse range of food allergens together grouped as pathogenesis related proteins (PR) under 14 families and comprises of pollens/fruit latex/fruit allergens mostly related to defence proteins. PRs are defined as proteins that are induced specifically in a plant as a response to infections by pathogens, such as fungi, bacteria, or viruses, or adverse environmental factors. PRs are not a protein superfamily but represent a collection of unrelated protein families that function as part of the plant defence system. The class includes proteins as diverse as class-I chitinases (PR-3 family), thaumatin like proteins (PR-5 family), peroxidises (PR-9 family) etc. Individuals with pollen allergy frequently have allergic symptoms after eating certain plant foods. The majority of these reactions are caused by allergens of Rosaceae fruits (e.g., apple, apricot, and pear) and Apiaceae vegetables (e.g., celery and carrot) that cross-react with allergens that are present in birch pollen, particularly the major birch pollen allergen Bet v 1, and other tree pollens. Well characterized allergens from this family include glucanase (banana), Pers a 1 (avocado), Cas s 5 (chestnut), Bra r 2 (turnip), Pru av 2 (cherry) and Mal d 2 (apple).
2.2 Seed storage proteins

The plant seed is not only an organ of propagation and dispersal but also the major plant tissue harvested by humankind. The amount of protein present in seeds varies from ~10% (in cereals) to ~40% (in certain legumes and oil seeds) of the dry weight forming the major source of dietary protein (Shewry et al., 1995). Although the vast majority of individual proteins present in seeds have either metabolic or structural role, all seeds also contain one or more groups of proteins that are present in high amounts and that serve to provide a store of amino acids for use during germination and seedling growth. These storage proteins are of particular importance because they determine not only the total protein content of the seed but also its quality for various end uses (Shewry et al., 1995). Because of their abundance and economic importance, seed storage proteins were among the earliest of all proteins to be characterised. However, the detailed study of seed storage proteins dates from the turn of century when Osborne classified them into groups on the basis of their extraction and solubility in water (albumins), dilute saline (globulins), alcohol/water mixture (prolamins) and dilute acid/alkali (glutelins) (Osborne and Campbell, 1898). Broadly the seed storage proteins have been classified into cupin super family and prolamine superfamily.

Cupin superfamily comprises of globular seed storage proteins. These proteins are the most widely distributed groups of storage proteins. They are present not only in dicots but also in monocots (including cereals and palms) and fern spores (Templeman et al., 1987). They can be divided into two groups based on their sedimentation coefficient: the 7S vicilin type globulins and the 11S legumin type globulins. Both groups show considerable variation in their structure, which results
partly from post-translational processing. In addition, both have nutritional significance in that they are deficient in cysteine and methionine, although 11S globulins contain slightly higher levels of these amino acids. The globulin storage proteins have been studied in more detail in legumes notably peas (Konopska, 1982), soybean (*Vicia faba*) (Bailey and Boulter, 1970) and French bean (*Phaseolus vulgaris*) (Lawrence et al., 1994). 11S legumins are the major storage proteins not only in most legumes but also in many other dicots (e.g. brassicas, composites and cucurbits) and some cereals (oats and rice). The mature protein consists of six subunit pairs that interact non-covalently. Each of these subunit pairs in turn consist of an acidic subunit of Mr ~40,000 and a basic subunit of Mr ~20,000 linked by a disulphide bond (Mills et al., 2002). In developing seeds, each subunit is synthesized as a single polypeptide precursor, preproprotein, the signal sequence of which is removed co-translationally. The resultant proprotein assembles into trimer of ~8S in endoplasmic reticulum. The proprotein trimers are transported from endoplasmic reticulum to protein storage vacuoles, where they are cleaved to form acidic and basic polypeptides that are linked by disulphide bond. Finally, the mature protein assembles into hexamer (Shewry et al., 1995). 7S vicilins are typically trimeric proteins of Mr ~150,000 to 190,000 that lack cysteine residues and hence cannot form disulphide bonds. Their detailed subunit composition vary considerably mainly because of differences in the extent of post translational processing (proteolysis and glycosylation) (Mills et al., 2002).

The existence of prolamine superfamily was proposed on the basis of the presence of a conserved skeleton of 8 cysteine residues within the proteins' sequences (Kreis et al., 1984). This superfamily is named after the cereal prolamins,
the major storage proteins of cereal grains (with the exception of oats and rice),
which are characterized by their high contents of proline and glutamine (Shewry et
al., 2002). In addition to the cereal prolamins, the broader definition of the
superfamily now includes several important plant allergen families, 2S albumin seed
storage proteins, nsLTPs, and cereal seed inhibitors of \( \alpha \)-amylase, trypsin, or both.
All of these low-molecular-weight proteins are cysteine rich and have similar 3-
dimensional structures that are rich in \( \alpha \)-helices (Breiteneder and Radauer, 2004).

2.2.1 Synthesis and deposition of seed storage proteins

All of the seed storage proteins are secretory proteins synthesized with a
signal peptide that is cleaved as the protein is translocated into the lumen of the
endoplasmic reticulum (ER). The subsequent events in storage protein processing are
less clearly understood and may vary not only between different species but also
within the same species, depending on the protein type and stage of development.
Secretory proteins assume their folded conformations within the lumen of the ER,
which is also the site of disulfide bond formation assisted by three types of luminal
proteins: molecular chaperones of the HSP70/BiP family, peptidyl-prolyl \textit{cis-trans}
isomerases (PPI), or cyclophilins and protein disulfide isomerise. Molecular
chaperones may facilitate folding by binding transiently to the nascent polypeptides
and may also prevent the formation of incorrect inter- or intramolecular interactions.
BiP-related proteins are present in developing endosperms of cereals such as rice (Li
et al., 1993) and maize (Boston et al., 1991), and they accumulate in higher than
normal levels in high-lysine maize mutants (Boston et al., 1991), possibly due to the
presence of incorrectly folded zeins. Cyclophilins assist protein folding by
accelerating the isomerization of Xaa-Pro bonds, a rate limiting step in protein
folding. The repetitive domains of cereal prolamins contain high levels of proline, and isomerization of Xaa-Pro bonds might therefore be expected to limit their folding. Whether protein disulfide isomerase (PDI) catalyzes disulfide bond formation in storage proteins also remains to be established. 7S and 11S globulin subunits are also assembled in the ER, with the 7S globulins forming the mature trimers.

After protein folding in ER, seed storage proteins are stored in membrane-bound vesicles, referred as protein bodies. Two routes of protein body formation appear to operate in developing cereal endosperms, in one of which the protein body forms from the vacuole and in the other of which it forms from the ER. For example, the major storage proteins in oats and rice are related to the 11s globulins of dicots and appear to be transported from the ER lumen via the Golgi apparatus to the vacuole. The protein bodies then appear to form by fragmentation of the vacuole. In contrast, the prolamins of rice (Krishnan et al., 1986) and maize (Larkins and Hurkman, 1978) appear to be retained within the lumen of the ER, which becomes distended to form protein bodies. Thus, rice endosperm cells contain two populations of protein bodies, some of vacuolar origin (containing glutenins) and others of ER origin (containing prolamins). The situation appears to be more complicated in barley, wheat, and rye, with prolamins present in both ER-derived and vacuolar protein bodies. The mechanisms that determine whether a prolamin is transported to the vacuole or retained in the ER are not known, because neither vacuolar targeting nor ER retention sequences have been identified. The assembly of the 11s globulins appears to be a highly regulated event. The monomeric proteins are initially assembled in the lumen of the ER into trimers that are then transported from the ER.
to the storage vacuole, where they are assembled into their final hexameric form. This assembly process requires specific proteolytic cleavage of the subunits present in the trimers (Dickinson et al., 1989). Uncleaved trimers cannot assemble into hexamers in vitro unless they have been treated with papain. This cleavage does not cause the trimers to disassemble but may result in a conformational change that favours assembly into hexamers. The 11S globulin vacuolar processing protease has been characterized from several species and shown to recognize asparagine processing sites specifically. Scott et al. (1992) purified a soybean protease that cleaves the trimeric 11S globulin proproteins. Hara-Nishimura et al. (1993) also purified an 11S globulin processing peptidase from castor bean that displays similar processing specificity and also appears to be a thiol protease but is unglycosylated. In leguminous plants, the 7s and 11s globulins appear to be in the same protein bodies with no spatial separation. In many other dicots, such as pumpkin, sunflower, brassicas, and castor bean, 2s albumins are stored together with 11s globulins, but how these distinct types of storage protein are organized in the protein bodies is not known. In contrast, there is evidence that the different types of prolamins are spatially separated in the protein bodies of cereal endosperms (Shewry et al., 1995).

2.2.2 Structural studies on seed storage proteins

Napin Bn1b is a representative member of 2S seed storage proteins, which consists of two polypeptide chains of 3.8 kDa and 8.4 kDa linked by two disulphide bridges. The solution structure of Napin Bn1b has been determined by NMR spectroscopy (Pantoja-Uceda et al., 2004). The three dimensional structure of 2S albumins from other seeds e.g. Ric C 3 from castor bean (Ricinus communis)
Food Allergy and Seed Storage Protein

(Pantoja-Uceda et al., 2003) and SFA-8 from sunflower (*Helianthus annuus*) (Pantoja-Uceda et al., 2004), has also been determined using NMR methods.

The global fold of 2S albumin proteins consists of five amphipatic helices. The five helices are arranged in a right handed superhelix. The tertiary structure is stabilised by four disulfide bonds resulting in a very compact structure. Between helix III and helix IV there is a relatively short segment, named the “hypervariable region” of 2S albumins because of high variability in the length and the composition of this segment among the different members of this family (Breiteneder and Radauer, 2004). The 3 dimensional structures of 2S albumins are very similar indeed to those of nsLTPs, α amylase inhibitors and HPS, which might be expected from the close matching of their corresponding disulphide bond patterns. When the sequences of the representative members of the four related protein families were compared, it has been found that the location and extension of the helices approximately coincide in the four families. There is an important difference between the structures of nsLTPs and those of albumins, α amylase inhibitors and HSP. Helices 3 and 4 are almost coplanar and helix 5 is much further back than in the other two families, in which the V shape formed by the homologue helices is more twisted. As a consequence, an internal cavity is present in nsLTPs that is able to host one molecule of lipid, whereas no cavity appears to be present in members of the other families (Breiteneder and Radauer, 2004).

The 3 dimensional structures of several 7S globulins has recently been determined using X-ray crystallography. The structure of jack bean (*Canavalis ensiformis*) Canavalin has been solved at 2.1 Å resolution (Ko et al., 2000) and that of French bean (*Phaseolus vulgaris*) Phaseolin at 2.2 Å (Lawrence et al., 1994).
These two structures have established a common vicilin monomer fold of two subunits, each consisting of a β barrel and an α helical domain. In these trmeric proteins the constituents monomers are arranged around a 3 fold symmetry axis. Each monomer consists of two α+β module related by a pseudo dyad perpendicular to the trimeric 3 fold. Each module contains a β barrel with an elaborated ‘jelly roll’ folding motif, followed by a α helical domain comprising 3 helices, two of which exhibit helix turn helix motif. The N terminal α helical domain is linked to the C terminal β barrel domain via a segment containing a fourth helix followed by a putatively extended portion.

Glycinin are 11S globulins of soybean (*Glycin max*). In general 11S globulins are composed of several kinds of subunits. Five major subunits have been identified from soybean: A1aB1b, A2B1a, A1B2, A3B4 and A5A4B3 (Adachi et al., 2003). Crystal structure of soybean proglycinin A1aB1b homotrimer was determined at 2.8 Å resolution (Adachi et al., 2001). The protomer consist of N and C terminal modules, that each consists of a jelly roll β domain and a helix domain. However the packaging of trimer into hexamer is not known because it is very difficult to obtain the crystal of 11S globulin suitable for X-ray analysis, because of their molecular heterogeneity. The crystal structure of glycinin A3B4 homohexamer has also been reported at 2.1 Å resolution (Adachi et al., 2001).

The main characteristic of the cupin domain is a two motif sequence: G(X)_{3}H_{2}XH(X)_{11}G and G(X)_{3}P_{4}(X)_{3}H_{3}XN, where X is any amino acid. Between these two His containing motifs, there is a region, that varies in length from 15 residues in many of the bacterial enzymes to more than 50 residues in some seed storage proteins. It is now known that the two conserved histidines and glutamate in
the first motif together with a third conserved histidine in the second motif act as the metal binding site. However most of the seed storage proteins either lack any of the conserved His residues or contain a single His residue in motif1. It is presumed that, as a consequence, they have no metal binding site and therefore no enzymatic associated. There is however a massive accumulation of oxalate during early seed development in soybean and presumably in other legumes, and it is thus tempting to speculate on the possibility that this compound act as a substrate for a residual oxalate degrading capacity provided by the storage proteins being produced at that period (Khuri et al., 2001; Woo et al., 2000).

2.2.3 Seed storage proteins and allergy

The double-stranded β-helix that comprises the cupin fold appears to be a remarkably stable structural motif, resisting both thermal denaturation and proteolysis. This stability probably plays an important role in permitting sufficient immunologically active fragments to pass down the gastrointestinal tract, and is responsible, in part at least, for the thermostable nature of the allergenic activity of these proteins (Maleki et al., 2000). Such properties, coupled with the abundance of storage globulins in the diet, must contribute to their being able to act as potent allergens. However, while characterized as important allergens in peanut and soybean, the role of cupins in the allergenicity of other nuts and seeds, where they are also abundant components, is less clear. For example, in the Brazil nut, a large proportion of the seed protein is the 11 S globulin, and yet the major allergen in this plant food species is the 2 S albumin (Mills et al., 2002).
Several of the tree nut and seed allergens are 2S albumins. They include *Ber e* 1 from Brazil nut, *Jug r* 1 from from the English walnut, and 2S albumin from cashew nuts. *Ses I* 2 is the clinically most important allergen of sesame seeds. The protein responsible for allergic reactions to mustard were identified as the 2S albumins: *Sin a* 1 from yellow mustard seeds and *Bra j* 1 from oriental mustard seeds. *Ara h* 2, 6, and 7 belong to the conglutin protein family, which is related to the 2S albumin family. *Ara h* 2 was found to act as weak trypsin inhibitor that protects *Ara h* 1 from degradation (Breiteneder and Radauer, 2004). The 2S albumins are highly compact, which confers on them a particular stability to thermal denaturation and to digestion by proteolytic enzymes (Shewry et al., 1995). These properties are related primarily to transport over mucosal barriers so that the proteins can reach their targets intact to mount an allergic response. On epitope mapping of the allergen *Sin a* 1 from yellow mustard (*Sinapsis alba*), which is sequentially and structurally related to napins, two immunodominant regions were defined. One is located very close to the 'hypervariable region' of the 2S albumins, which forms a very flexible loop between helices 3 and 4. Recently, a minimal linear IgE-reactive epitope was identified in the major food allergen, *Jug r* 1, in the 2S albumin of English walnut. This four-residue epitope, Arg-Gly-Glu-Glu, is also located in the hypervariable region, which reinforces the idea that the major epitope of napins and structurally related proteins may reside in this region. More information is needed, in particular on the structure of the complexes of IgE with 2S albumins, in order to understand more closely the atomic interactions between antibody and antigen, as well as the role of regions other than the epitope in the correct orientation of the allergen into the interacting region of the Fab fragment of the antibody (Shewry et al., 1995).
2.2.4 Physiological roles of seed storage proteins

Storage proteins are a group that comprises of proteins generated mainly during seed production and stored in the seed that serve as nitrogen sources for the developing embryo during germination. The cellular mechanisms involved in providing this resource appear to be highly specialized, including dedicated compartments termed protein storage vacuoles (PSVs), conserved storage proteins, and specific post-translational processing of these proteins within the PSV compartment (Muntz, 1998). Although the post-translational polypeptide cleavage of storage proteins upon arrival at the PSV is common, the specific function(s) of that processing has not been elucidated in vivo. An attractive hypothesis for proteolytic processing suggests it triggers protein conformational changes that enable dense packaging and long-term stable storage of reserves within the PSV.

Though, being the storage house of nitrogen, appears to be the primary function of seed storage proteins, many other different functions has been ascribed to this family of the protein. Many of these proteins are known to bind various ligands serving either as transporter or storage. A subclass of 2S albumin, referred as nsLTPs are involved in binding lipid molecules (Cheng et al., 2004). Heme binding and thiamine binding activities have been detected in case of pea albumin 2 protein (PA2). PA2 has also been associated with polyamine biosynthesis (Vigeolas et al., 2008). Soluble thiamine binding protein has been reported from buckwheat (Watanabe et al., 1998). Seed storage proteins have been frequently reported to possess lectin like activity (Langston-Unkefer et al., 1984).

Various antifungal and/or antibacterial proteins such as chitinases, β-glucanases, thionins, ribosome-inactivating proteins and permatins have been
Food Allergy and Seed Storage Protein

detected in seeds. Antimicrobial proteins and peptides have been isolated from seeds of maize (*Zea mays* L.) (Duvick et al., 1992) and radish (*Raphanus sativus* L.). They are believed to play a role in plant defence because of their strong antimicrobial activity (Terras et al., 1995). This belief is supported by their ability to confer resistance (to pathogens) to transgenic plants containing genes that encode them. Other plant-derived proteins have insecticidal properties that can, for example, protect seeds from attack by larvae of various bruchids (Mourey et al., 1998) and inhibit the growth and development of *Helicoverpa punctigera* (Wallengren) larvae (Jennings et al., 2001). Of particular interest are plant-derived proteins called cyclotides (circular proteins in which the N and C termini are linked via a peptide bond), which have antimicrobial and insecticidal properties (Jennings et al., 2001). Ocatin, a protein isolated from the Andean tuber crop oca (*Oxalis tuberosa*), is reported to have antibacterial and antifungal effects (Flores et al., 2002).

**2.3 Genesis of the present study**

Food allergies appear to be on the increase along with other forms of allergic disease. Around 7–10 food items are responsible for the majority of allergies, including several of plant origin. Allergies are usually triggered by the protein components in a food, which are also known as allergens. Therefore, it is important to address the basis of protein allergenicity. While the abundance of a protein in a food is one factor involved in determining its allergenic potential, this does not appear to be a sufficient determinant. Through an analysis of common properties of plant food allergens that trigger a reaction via the gastrointestinal tract it has become evident that food allergens belong almost exclusively to three structurally related protein superfamilies—the prolamins, the cupins and the cysteine proteases. Whilst
we cannot as yet predict the allergenicity of a given protein, an understanding of the structural attributes of proteins that predispose them to become allergenic is important if we need to understand what makes some foods more allergic than others.

The work presented in the thesis is an endeavour to understand many unanswered questions associated with food allergy. It is an attempt to identify the structural attributes common to plant food allergens that will allow us to investigate whether these shared properties play a role in determining allergy potential in the future. The study focuses on role of structural attributes in determining allergenicity. Efforts have also been made to identify the normal physiological roles of the proteins which are otherwise not clearly understood in many instances.