



ELSEVIER  
MASSON

Disponible en ligne sur [www.sciencedirect.com](http://www.sciencedirect.com)



ScienceDirect

Revue française d'allergologie et d'immunologie clinique xxx (2007) xxx-xxx

REVUE FRANÇAISE  
D'ALLERGOLOGIE  
ET D'IMMUNOLOGIE CLINIQUE

<http://france.elsevier.com/direct/REVCLI/>

## Can any protein become an allergen?

## Est-ce n'importe quelle protéine peut devenir un allergène ?

H. Breiteneder

*Department of Pathophysiology, Center for Physiology, Pathophysiology and Immunology, Medical University of Vienna, AKH-EBO-3Q,  
Wahringer Guertel 18-20, Vienna, 1090 Austria*

### Abstract

Allergens of plant and animal foods and pollen belong to a highly restricted number of protein families. The AllFam Database (<http://www.meduniwien.ac.at/allergens/allfam/>) provides regularly updated lists of protein families that contain allergens. At present, 2% of the 9318 protein families defined by the Pfam Database (<http://pfam.sanger.ac.uk/>) contain allergens. Related protein families can be grouped into superfamilies placing allergenic proteins in an evolutionary context. With the exception of the prolamin superfamily, allergenic plant proteins are found in few member families of their respective superfamilies. This might indicate that allergenicity emerged rather infrequently in a very limited number of protein families. Moreover, most members of a given protein family seem to be non-allergenic. In contrast to plant allergens, the allergenicity of animal food allergens seems to be dependent on the degree of identity to a human homologue. The closer a potential animal allergen is to a human protein, the less likely it is to act as allergen.

© 2008 Publié par Elsevier Masson SAS.

### Résumé

Les allergènes des aliments d'origine animale et végétale et le pollen appartiennent à un nombre limité de familles protéiques. La base de données AllFam (<http://www.meduniwien.ac.at/allergens/allfam/>) met à jour régulièrement la liste des familles protéiques qui contiennent les allergènes. Actuellement, seulement 2 % des 9318 familles protéiques définies par la base de données Pfam (<http://pfam.sanger.ac.uk/>) sont des allergènes. Les familles protéiques sont classées en superfamille replaçant les protéines allergéniques dans leur contexte phylogénétique. Hormis la superfamille des prolamines, les protéines des plantes allergisantes sont retrouvées dans peu de membres de familles de leur superfamille respective. Cela peut indiquer que l'allergénicité peut provenir peu fréquemment dans un nombre limité de familles protéiques. De plus, la plupart des membres d'une famille protéique semble être non allergisant. À l'inverse des allergènes végétaux, l'allergénicité des allergènes alimentaires d'origine animale semble dépendre du degré d'identité à un homologue humain. Plus l'allergène potentiel se rapproche de la protéine humaine, moins il se comportera comme un allergène.

© 2008 Publié par Elsevier Masson SAS.

**Keywords:** AllFam; Protein families; Evolutionary biology of allergens; Emergence of allergenicity

**Mots clés :** Base de données Allfam ; Familles de protéines ; Évolution biologique des allergènes ; Émergence de l'allergénicité

### 1. Introduction

In the late 1980s, the first allergens were studied on a molecular level. Since then hundreds of allergens have been identified from all types of sources (pollen, mites, molds, animal dander, insect venoms, natural rubber latex and foods) and their sequences determined. A number of databases of allergens were established, such as the Official List of

Allergens issued by the International Union of Immunological Societies Allergen Nomenclature Subcommittee (<http://www.allergen.org>), the Allergome (<http://www.allergome.org>), the Food Allergy Research and Resource Program Allergen Database (<http://www.allergenonline.com>) and the InFormAll Database (<http://foodallergens.ifr.ac.uk/>). Most of these databases classify allergens by source. Due to the growing number of available allergen sequences and structures, it is now possible to establish a classification based on protein families. The classification of pollen allergens [1] and plant [2] and animal food allergens [3] into protein families has clearly

E-mail address: [heimo.breiteneder@meduniwien.ac.at](mailto:heimo.breiteneder@meduniwien.ac.at).

shown that these allergens are restricted to few protein families. Related protein families can be grouped together into superfamilies which show allergenic proteins in their evolutionary context [4]. The emerging research discipline of the evolutionary biology of allergens studies the origin of allergens to elucidate the emergence of allergenicity.

## 2. Protein families of allergens

In-depth studies of protein family membership of allergens have been published on plant [2] and animal food allergens [3] and on pollen allergens [1]. The most striking finding was that nearly 60% of all plant food allergens belonged to one of four protein superfamilies [2]. These families were the prolamin, the cupin and the Bet v 1 superfamilies and the profilins. A similar situation has been discovered for pollen [1] and animal food allergens [3], however the distribution of allergens into the respective families was less biased.

The prolamin superfamily is the highest ranking protein superfamily that contains allergens. Allergenic prolamins include members of the nsLTPs [5], the 2S albumin storage proteins [6] and the cereal alpha-amylase/trypsin inhibitors [7]. The cupins are a superfamily of functionally highly diverse proteins whose evolution can be traced from archaea and bacteria to eukaryotes including plants and animals [8]. It was estimated that there were a minimum of 18 different functional families. However, the majority of cupin allergens were only found among either the vicilin-like or the legumin-like seed storage protein families [9]. Profilins can be divided into two groups, one containing vertebrate and viral profilins, the other comprising profilins from invertebrates, protists, plants and fungi [10]. Allergenic profilins were exclusively found in flowering plants. In recent years, a superfamily of proteins related to the major birch pollen allergen Bet v 1 was established. An extended sequence comparison revealed that the Bet v 1 superfamily comprised at least four subfamilies with low levels of sequence similarity between each other: the PR-10 family, the family of major latex proteins (MLP) and ripening-related proteins (RRP), the norcochlorogenic acid synthases, and the cytokinin-binding proteins from legumes [11]. Allergenicity within the Bet v 1 superfamily is confined to the PR-10 family [4].

Pollen allergens were classified into 29 protein families [1]. The major pollen allergen families are expansins, profilins and calcium-binding proteins. Expansins are ubiquitous plant cell wall glycoproteins that catalyze cell wall loosening during cell growth [12]. Expansins consist of two domains. Several groups of allergens are related to one or both expansin domains. Group 1 grass pollen allergens are beta-expansins. The groups 2 and 3 of grass pollen allergens are also related to beta-expansins, but lack the N-terminal domain. Plant profilins were originally discovered as cross-reactive pollen allergens eliciting immunoglobulin E responses in 10 to 20% of pollen-allergic patients [13]. Later they were also described as allergens in plant foods and *Hevea* latex [14]. The largest group of pollen allergens containing calcium-binding EF hand domains is the polcalcin family [15].

Animal food allergens could be classified into three main families, tropomyosins, EF-hand proteins and caseins, along with 14 minor families [3]. Tropomyosins play one of the key roles in muscle contraction. Vertebrate tropomyosins from mammals, birds and fish have never been reported to be allergenic. Allergenic tropomyosins are confined to the invertebrate groups, insects, crustaceans and mollusks [3]. The EF-hand family is composed almost exclusively of parvalbumins. Parvalbumins can be subdivided into two distinct evolutionary lineages, alpha- and beta-parvalbumins. The alpha-parvalbumins, abundant in the muscle of fish and amphibians and to a minor extent in the fast twitch muscle of birds and mammals, are not generally allergenic. Many beta-parvalbumin allergens are found in a variety of fish species. Caseins are exclusively mammalian proteins found in milk. The immunoglobulin E reactivity of the various types of caseins seems to be linked to their sequence similarities with human homologues. Caseins being the least like human are more reactive [3].

## 3. The evolutionary biology of allergens

The classification of proteins into protein families made the identification of a large number of proteins possible that are related to known allergens. These allergen homologues are either not allergenic or have never come into contact with the human immune system resulting in a lack of knowledge about their allergenicity. It is now possible to view allergenic proteins in their evolutionary context, trace back their origins to ancestral molecules and compare them to their non-allergenic counterparts.

The prolamin superfamily contains the highest number of allergens of all superfamilies [4]. Four of its eight member families contain allergens. The identification of the existence of the prolamin superfamily was based on the discovery of the presence of a conserved cysteine skeleton [16]. Besides that, the various member families have little sequence similarity to each other although they possess highly similar  $\alpha$ -helical architectures. Members of the alpha-globulin family of storage proteins that were originally identified in rice and corn are present in amounts comparable to allergenic members of the prolamin superfamily. Nevertheless, cereal alpha-globulins have never been identified as allergens. Members of the prolamin superfamily have only been reported in seed plants with the exception of proteins related to 2S albumins that are present in spores of the ostrich fern [17]. In contrast, the origin of the cupin scaffold can be traced back to archeal and bacterial predecessors [8]. The majority of the cupin allergens belongs to only two of the at least 18 cupin families, namely the vicilins and the legumins [4].

The major birch pollen allergen Bet v 1 is a member of the PR-10 family of the Bet v 1 superfamily [4]. At present the Bet v 1 superfamily comprises five member families. As structural homologues of Bet v 1 have been recently discovered in archaea and bacteria, the Bet v 1 superfamily is currently being expanded. Allergenicity, however seems to be confined to certain members of the PR-10 family, including the Bet v 1

homologues of pollen of the plant order Fagales, fruits of the Rosaceae and vegetables of the Apiaceae. Allergenic profilins are restricted to pollen, plant foods and *Hevea* latex although profilins are found in all eukaryotic cells and even pox viruses [1].

In contrast to plant food or pollen allergens, almost all animal food allergens have homologues in the human proteome which affects the way in which they are recognized by the human immune system. Especially allergenic mammalian proteins, as found for example in cow's milk, lie at the limits of the capability of the human immune system to discriminate between foreign and self proteins [18]. For the three major animal food allergen families, tropomyosins, parvalbumins and caseins, allergenicity is inversely related to their sequence identity to a homologous human protein [3]. In the sequence set that was analyzed by Jenkins et al. [3], potentially allergenic animal food proteins with a sequence identity of below 54% to human homologues were all allergenic, whereas those with a sequence identity greater than 65% were rarely allergenic. This observation relates to the requirement for proteins to be recognized as non-self to mount an immune response.

#### 4. The evolution of allergenicity

The inquiry into which features make a protein allergenic is of high interest to allergologists worldwide. It has been argued by some that any immunogenic protein presented in the right context and in sufficient quantity to the immune system of an atopic individual can become an allergen [19]. Other authors attribute allergenicity to certain architectural and structural features of the proteins [9,20]. For this discussion, the comparison of allergenic and non-allergenic members in the various protein superfamilies could provide further insights. Preliminary studies on the evolution of allergenicity are so far only available for certain plant proteins [4].

The situation of the prolamin superfamily is unique. There is very little sequence similarity between the proteins of the various member families besides the conserved cysteine skeleton. Hence, the presence of allergens in many of the member families of this superfamily indicates a tendency of a proto-prolamin to act as an allergen. In contrast, the other superfamilies discussed in this article show a very limited distribution of allergens among their member families. In these cases, allergenicity might have emerged much later only along selected phylogenetic lineages. For example, profilins from invertebrates, fungi and plants show very low sequence similarities to human profilins, however only plant profilins have been described as allergens.

#### 5. Concluding remarks

In the last two decades, hundreds of allergens have been identified, cloned and sequenced. This wealth of data now enables us to classify allergens into protein families that are defined by sequence and structural similarity. Employing different methods, several groups have shown that most

allergens belong to a limited set of protein families [1–3,21] thus disproving the assumption that any protein can become an allergen. Querying the AllFam Database, it is shown that the distribution of allergens into protein families is highly biased. They are found in only 2% of the 9318 families defined by the Pfam Protein Family Database. Hence, the classification of allergenic proteins based on structural similarities including the comparison of allergenic and non-allergenic members of a protein family will lead to new insights into factors that contribute to allergenicity.

#### Acknowledgement

The support of the Austrian Science Fund grant SFB F01802 is gratefully acknowledged.

#### References

- [1] Radauer C, Breiteneder H. Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J Allergy Clin Immunol* 2006;117:141–7.
- [2] Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteneder H, Mills EN. Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. *J Allergy Clin Immunol* 2005;115:163–70.
- [3] Jenkins JA, Breiteneder H, Mills EN. Evolutionary distance from human homologs reflects allergenicity of animal food proteins. *J Allergy Clin Immunol* 2007;120:1399–405.
- [4] Radauer C, Breiteneder H. Evolutionary biology of plant food allergens. *J Allergy Clin Immunol* 2007;120:518–25.
- [5] Breiteneder H, Mills C. Nonspecific lipid-transfer proteins in plant foods and pollens: an important allergen class. *Curr Opin Allergy Clin Immunol* 2005;5:275–9.
- [6] Pastorello EA, Pompei C, Pravettoni V, Brenna O, Farioli L, Trambaioli C, et al. Lipid transfer proteins and 2S albumins as allergens. *Allergy* 2001;56(Suppl. 67):45–7.
- [7] Pastorello EA, Farioli L, Conti A, Pravettoni V, Bonomi S, Iametti S, et al. Wheat IgE-mediated food allergy in European patients: alpha-amylase inhibitors, lipid transfer proteins and low-molecular-weight glutenins. Allergenic molecules recognized by Double-Blind, Placebo-Controlled Food Challenge. *Int Arch Allergy Immunol* 2007;144:10–22.
- [8] Dunwell JM, Purvis A, Khuri S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 2004;65:7–17.
- [9] Breiteneder H, Radauer C. A classification of plant food allergens. *J Allergy Clin Immunol* 2004;113:821–30.
- [10] Arasada R, Gloss A, Tunggal B, Joseph JM, Rieger D, Mondal S, et al. Profilin isoforms in *Dictyostelium discoideum*. *Biochim Biophys Acta* 2007;1773:631–41.
- [11] Liscombe DK, MacLeod BP, Loukanina N, Nandi OI, Facchini PJ. Evidence for the monophyletic evolution of benzyloquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry* 2005;66:2501–20.
- [12] Sampedro J, Cosgrove DJ. The expansin superfamily. *Genome Biol* 2005;6:242.
- [13] Valenta R, Duchene M, Pettenburger K, Sillaber C, Valent P, Bettelheim P, et al. Identification of profilin as a novel pollen allergen; IgE autoreactivity in sensitized individuals. *Science* 1991;253:557–60.
- [14] Radauer C, Hoffmann-Sommergruber K. Profilins. In: Mills ENC, Shewry PR, editors. *Plant food allergens*. Blackwell Publishing; 2004. p. 105–24.
- [15] Wopfner N, Dissertori O, Ferreira F, Lackner P. Calcium-binding proteins and their role in allergic diseases. *Immunol Allergy Clin North Am* 2007;27:29–44.
- [16] Kreis M, Forde BG, Rahman S, Mifflin BJ, Shewry PR. Molecular evolution of the seed storage proteins of barley, rye and wheat. *J Mol Biol* 1985;183:499–502.

- [17] Rodin J, Rask L. Characterization of matteuccin, the 2.2S storage protein of the ostrich fern. Evolutionary relationship to angiosperm seed storage proteins. *Eur J Biochem* 1990;192:101–7.
- [18] Spitzauer S. Allergy to mammalian proteins: at the borderline between foreign and self? *Int Arch Allergy Immunol* 1999;120: 259–69.
- [19] Aalberse RC. Structural features of allergenic molecules. *Chem Immunol Allergy* 2006;91:134–46.
- [20] Breiteneder H, Clare Mills EN. Plant food allergens`structural and functional aspects of allergenicity. *Biotechnol Adv* 2005;23:395–9.
- [21] Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *FASEB J* 2003;17:1141–3.