



The analysis of specific allergenicity of food allergens families

Jing Wang, Cui Zhou, Jing Tian, Na Sun, Jing Lu, Cuiyan Wang & Huilian Che

To cite this article: Jing Wang, Cui Zhou, Jing Tian, Na Sun, Jing Lu, Cuiyan Wang & Huilian Che (2014) The analysis of specific allergenicity of food allergens families, Food and Agricultural Immunology, 25:3, 375-385, DOI: [10.1080/09540105.2013.805733](https://doi.org/10.1080/09540105.2013.805733)

To link to this article: <http://dx.doi.org/10.1080/09540105.2013.805733>



Published online: 01 Aug 2013.



Submit your article to this journal [↗](#)



Article views: 91



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

The analysis of specific allergenicity of food allergens families

Jing Wang^a, Cui Zhou^a, Jing Tian^b, Na Sun^a, Jing Lu^a, Cuiyan Wang^a and Huilian Che^{a,c,*}

^aCollege of Food Science and Nutritional Engineering, China Agricultural University, Beijing, China; ^bDepartment of Dermatology, The First Affiliated Hospital of Harbin Medical University, Harbin, China; ^cThe Supervision, Inspection and Testing Center of Genetically Modified Organisms, Ministry of Agriculture, Beijing, China

(Received 24 March 2013; final version received 13 May 2013)

Classification of food allergens based on protein's structure and function contributes to the study of the relationship between bioinformatics and potential allergenicity of allergens, as well as the evaluation of novel proteins' allergenicity. Some researches were focused on classification within plant or animal food allergens respectively, but there is not any classification of the wholefood allergens. In this article, we classified all the food allergens included in the SDAP into different food allergen families and analysed their specific allergenicity. According to allergen families in AllFam Database, food allergens taken from SDAP (Structural Database of Allergenic Proteins) are classified into different allergen families. Moreover, Protean of DNASTar was applied to analyzing the allergenicity of food allergens. 60% of food allergens are included in the five allergen families: Prolamin superfamily, EF hand domain, Cupin superfamily Profilin and Bet v 1-related protein. Besides, three other cross-food allergen families are found: EF hand domain, Serpin serine protease inhibitor and Triose phosphate isomerase, which include food allergens from both plants and animals. Common characters of food allergens from the same family were easy to find. For instance, they share the same specific peptides and potential T lymphocyte antigen epitope. This classification and characterization of allergen allergenicity in different food allergen families provide foundation for studying the structure and function of novel food allergens as well as the data for the assessment of potential allergenicity.

Keywords: food allergen; classification; food allergen families; allergenicity

1. Introduction

Proteins can be classified into different families based on their similarity in amino acid sequence and three-dimensional structure. Mills has reported that most of food allergens fall into only a few of protein families (Mills, Jenkins, Alcocer, & Shewry, 2004). This makes it possible to evaluate potential allergenicity of novel proteins via bioinformatics.

A total of 65% of plant food allergens belong to four allergen families: Cupin superfamily, Prolamin superfamily, Bet v 1-related protein and Profilin (Jenkins, Griffiths-Jones, Shewry, Breiteneder, & Mills, 2005). It has been reported that most of plant food allergens are homologous to pathogenesis-related proteins (PRs;

*Corresponding author. Email: chehuilian@cau.edu.cn

Breiteneder & Ebner, 2000). Accordingly, proteins that are homologous to PRs are classified into plant defence system protein family.

A majority of animal food allergens could be included in three allergen families: Tropomyosins, EF hand proteins and Caseins (Jenkins, Breiteneder, & Mills, 2007). Unlike plant allergens, most animal food allergens share homologs with human proteins and this may affect the way they are recognised by human immune system.

SDAP is a Web server that provides various computational tools for the study of allergenic proteins. It can also provide the amino acid sequence of some common allergens, and search relative biological information from major protein servers (PDB, SWISS-PROT, PIR). SDAP has listed 528 food allergens so far.

Although food allergens could be classified by their protein families from SDAP, not all protein families contain allergenic protein (Radauer, Bublin, Wagner, Mari, & Breiteneder, 2008), and consequently the classification based on protein families is over-inclusive. However, AllFam Database only records the protein families that contain allergenic proteins forming the allergen families. That's conducive to the study on the special structure and function of allergens.

AllFam is based on Allergome database which is the currently most comprehensive collection of allergen data. Allergome provides the information of identified allergens causing an IgE-mediated (allergic, atopic) disease (anaphylaxis, asthma, atopic dermatitis, conjunctivitis, rhinitis, urticaria). It contains all the identified and characterised allergens including those in the official IUIS Allergen Nomenclature list. The classification of allergen families of AllFam is obtained from Pfam, a protein family database. Every allergen family in AllFam is homologous to one or several Pfam families. AllFam has included 186 allergen families and 1091 allergens so far. Allergens in AllFam database are sourced from various organisms, such as fungi, insect, tree pollen, etc. Only food allergens are tested and classified in our experiments since we concentrate on studying the characters of food allergens.

DNASar, as a software, is usually adopted to analyse sequence in a comprehensive way. Protean is one module of DNASar, which contains a variety of methods for the analysis and prediction of protein structure and allergenicity. To analyse the features of allergenicity of the main food allergen families, we adopted two methods – Rothbard-Taylor and Jameson-Wolf. Rothbard-Taylor is an analysis method to forecast the specific motif which contains potential T lymphocyte antigen epitope, and Jameson-Wolf is used to analyse the antigenic index.

The food allergens tested in our laboratory are obtained from SDAP, and are classified into different food allergen families based on total allergen families from AllFam. After the classification of food allergen families, typical allergens of main food allergen families are selected and analysed using Protean module in DNASar.

2. Methods

2.1. Classification of food allergens

In this study, food allergens are classified into different allergen families listed in AllFam. This classification is based on the total allergen families in AllFam database.

The list of 528 food allergens (including the isoallergens) is gained from SDAP (<http://fermi.utmb.edu/SDAP/index.html>). Pfam ID of the allergen is also obtained from the database in SDAP. With the shared Pfam ID as a standard, allergens in the

same family could be identified and retrieved from AllFam (<http://www.meduniwien.ac.at/allergens/allfam/>). Then, all the allergen families are ranked on the basis of the number of allergens that each family contains.

Take Ara h 1 as an example. Ara h 1 is a primary peanut allergen and a kind of 7S vicilins from Cupin superfamily. Its Pfam ID, which is obtained from SDAP, is adopted as a parameter to identify all the other allergens in AllFam which are classified into the same family as Ara h 1.

2.2. Analysis of allergenicity of food allergens

The sequences of food allergens are obtained from NCBI. We downloaded the FASTA files of typical allergens of main food allergen families, and used Protean of DNASTar to analyse the allergenicity of those allergens.

3. Results

About 466 of the 528 food allergens listed in SDAP have corresponding allergen families in AllFam. And 60% of these 466 food allergens belong to five allergen families: Prolamin superfamily, EF hand domain, Cupin superfamily, Profilin and Bet v 1-related protein. Typical food allergens of those allergen families are shown in (Table 1). 28 Allergen families contain fewer than three food allergens, and make up 60% of all the food allergen families. This suggests most of food allergens could be categorised into the main food allergen families. This result is consistent with Mills' study to some extent (Mills et al., 2004).

3.1. Prolamin superfamily

As one of the major plant food allergen families, Prolamin superfamily contains 57 food allergens and 51 isoallergens. Most of the allergens from this family have large amounts of α -helix structures as well as similar three-dimensional structure and a conserved skeleton consisted of eight cysteine residues within the proteins' sequences (Heimo & Christian, 2004; Kreis, Forde, Rahman, Mifflin, & Shewry, 1985). Members of this family also share a domain, which forms a four-helical bundle in a right-handed superhelix with a folded leaf topology (Ivanciuc, Garcia, Torres, Schein, & Braun, 2009). The domain, which is stabilised by disulfide bonds, has an internal cavity (Ivanciuc et al., 2009). Protean's analysis of Prolamin superfamily's allergens is shown in (Figure 1). Most potential T lymphocyte antigen epitopes of allergens from this family characterise an α -helix structure. Several frequently seen

Table 1. Typical food allergens of five allergen families that contain most of food allergens.

Allergen family	Common allergen
Prolamin superfamily	Ara h 2 (peanut), Ara h 6 (peanut), Ber e 1 (Brazil nut)
Bet v 1-related protein	Api g 2 (celery), Dau c 1 (carrot), Pyr c 1 (pear)
EF hand domain	Gad c 1 (cod), Cyp c 1 (carp)
Cupin superfamily	Gly m 5 (soybean), Ara h 3 (peanut), Ses i 6 (sesame)
Profilin	Ara h 5 (peanut), Lyc e 1 (tomato), Mus a 1 (banana)

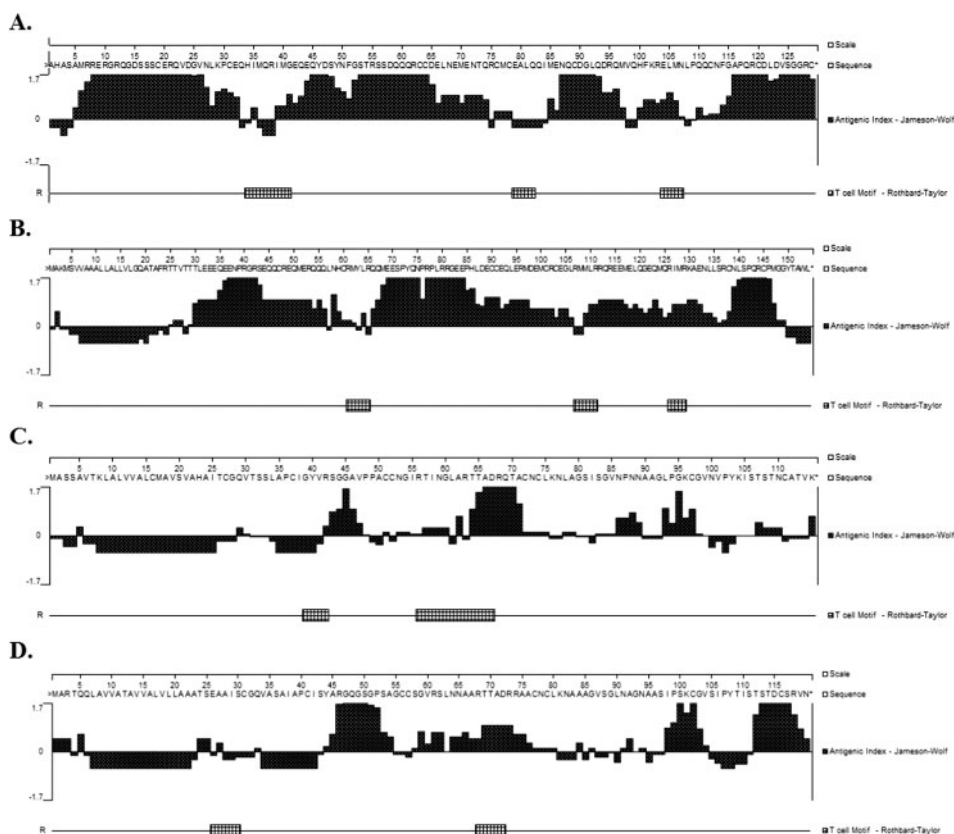


Figure 1. The protean analysis of typical allergens from Prolamin superfamily. (A) represents Ara h 6; (B) represents Ber e 1; (C) represents Mal d 3; (D) represents Zea m 14.

allergens from soy, cereal, vegetable and fruit belong to this family: for instance, 2S albumins, non-specific lipid transfer proteins, cereal α -amylase and protease inhibitors. The well-defined peanut allergens, such as Ara h 2, Ara h 6 and Ara h 7, are also included in Prolamin superfamily.

3.2. EF hand domain

EF hand domain includes 26 food allergens and 26 isoallergens, most of which are from animals. Most of their allergenic proteins are calcium-binding proteins, and share a motif which has a 12 residue loop flanked on both sides by a 12 residue α -helical domain (Ivanciuc et al., 2009). Analysis results from Protean on EF hand domain's typical allergens are shown in (Figure 2). Potential T lymphocyte antigen epitopes of many allergens have an α -helix structure. The biological functions of these proteins can be divided into two categories: signalling and calcium buffering (or transport). Some reports suggest that the IgE-binding capacity of some EF hand-containing allergens was related to calcium-binding (Gao, Gillen, & Wheatly, 2006).

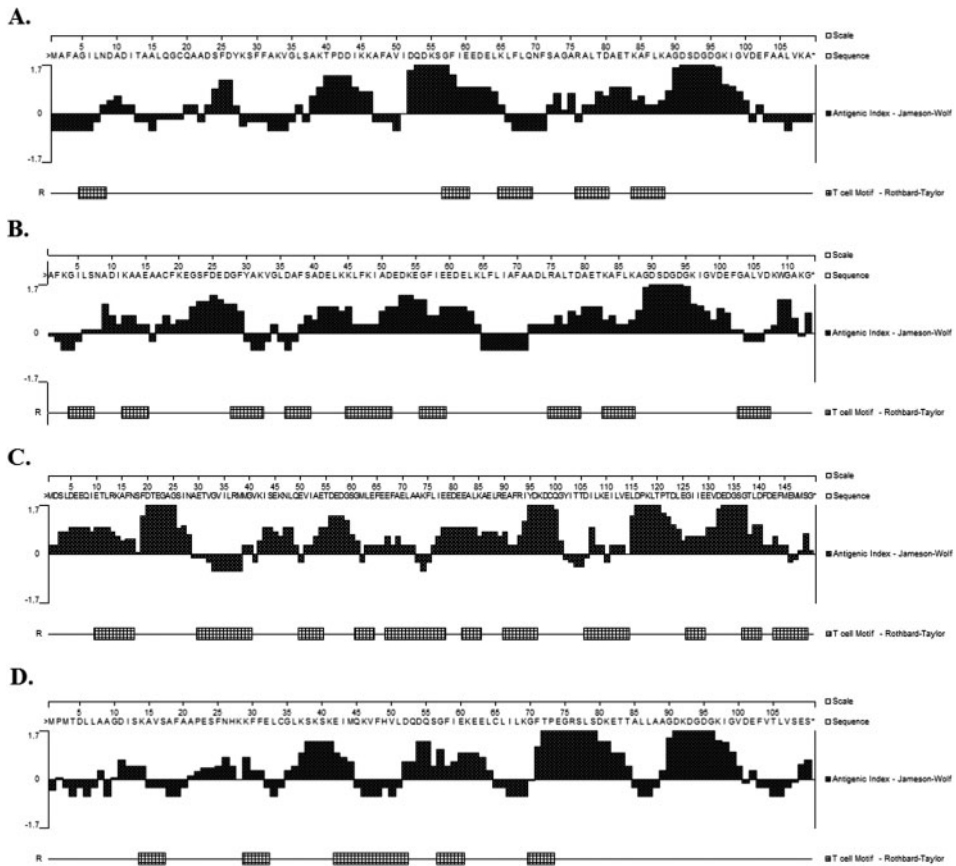


Figure 2. The protean analysis of typical allergens from EF hand domain. (A) represents Cyp c 1.01; (B) represents Gad c 1; (C) represents Pen m 6.0101; (D) represents Ran e 1.

3.3. Cupin superfamily

Cupin superfamily includes 24 food allergens and 26 isoallergens. Most allergens are from legumes, nuts and seeds (Mills et al., 2004). These proteins share two short conserved consensus sequence motifs and a β -barrel structure (Jenkins et al., 2007). Protean's analyses on some of this family's typical allergens are shown (Figure 3). The result of Jameson-Wolf indicates that peptides with high antigenic index carry large percentage of glutamine. And some allergens share the same potential T lymphocyte antigen epitopes that characterise an α -helix structure. The members of this family mainly belong to seed storage globulins that can be divided into two subfamilies: 7/8S vicilins and mature 11S legumins.

3.4. Profilin

Profilin includes 24 food allergens and 20 isoallergens, most of which are 12–15 kDa proteins mainly found in eukaryotic cells. At least 75% of the family members show sequence identities. Protean's analysis of Profilin's typical allergens is shown (Figure 4). Most allergens in this family share 3 or 4 peptides that show high

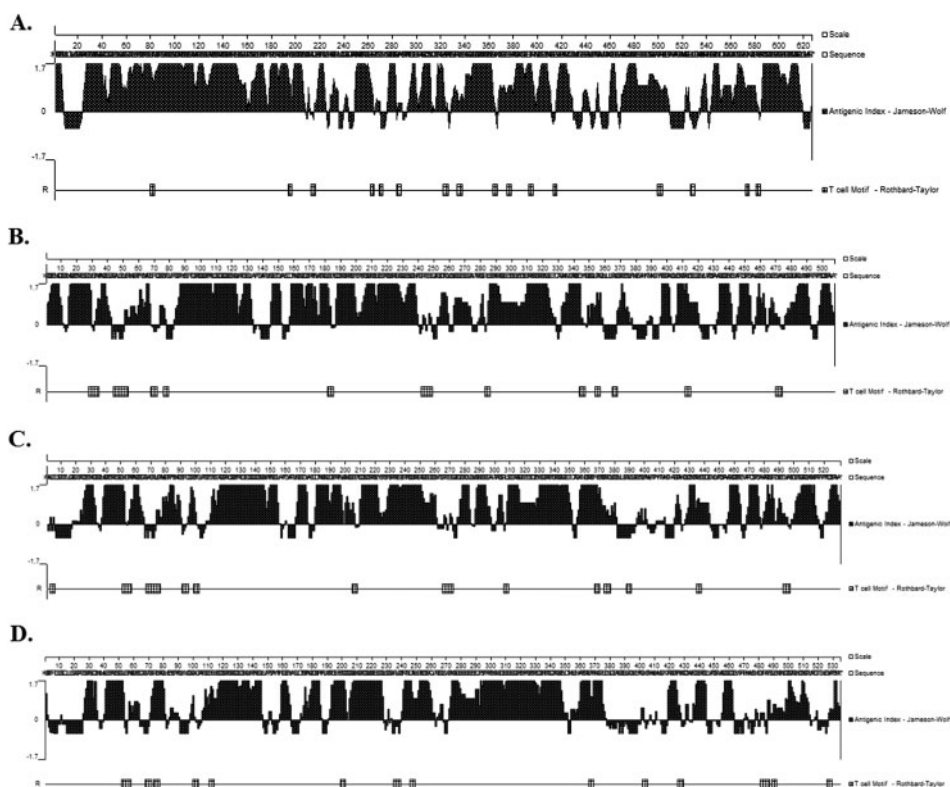


Figure 3. The proteomic analysis of typical allergens from Cupin superfamily. (A) represents Ara h 1; (B) represents Ara h 3; (C) represents Ara h 4; (D) represents Gly m 6.

antigenic index. In addition, Rothbard-Taylor analysis shows that most allergens share three or four potential T lymphocyte antigen epitopes, most of which are in β -strand or turn structure. 10% to 20% of patients allergic to allergens of Profilin also suffer from pollen allergy (Valenta et al., 1992). That indicates these allergens have been considered as a risk factor for allergic reactions to multiple pollen sources (Mari, 2001).

3.5. Bet v 1-related protein

Bet v 1-related protein includes 18 food allergens and 44 isoallergens, most of which are widely distributed among vascular plants, such as Rosaceae (apple, pear, stone fruits), Apiaceae (celery, carrot) and Fabaceae (soybean, peanut). Allergens from this family are formed by seven anti-parallel β -strands and three α -helices. Besides, they also share a C-terminal amphipathic helical motif, which consists of a helix (Ivanciuc et al., 2009). Proteomic analyses of some of this family's typical allergens are shown in (Figure 5). A short peptide is found between 120th to 140th amino acid, and all of these 20 amino acids have high antigenic index. And the short peptide often forms α -helix structure. The result of Rothbard-Taylor shows that many allergens of this family share two peptides, which may contain potential T lymphocyte antigen

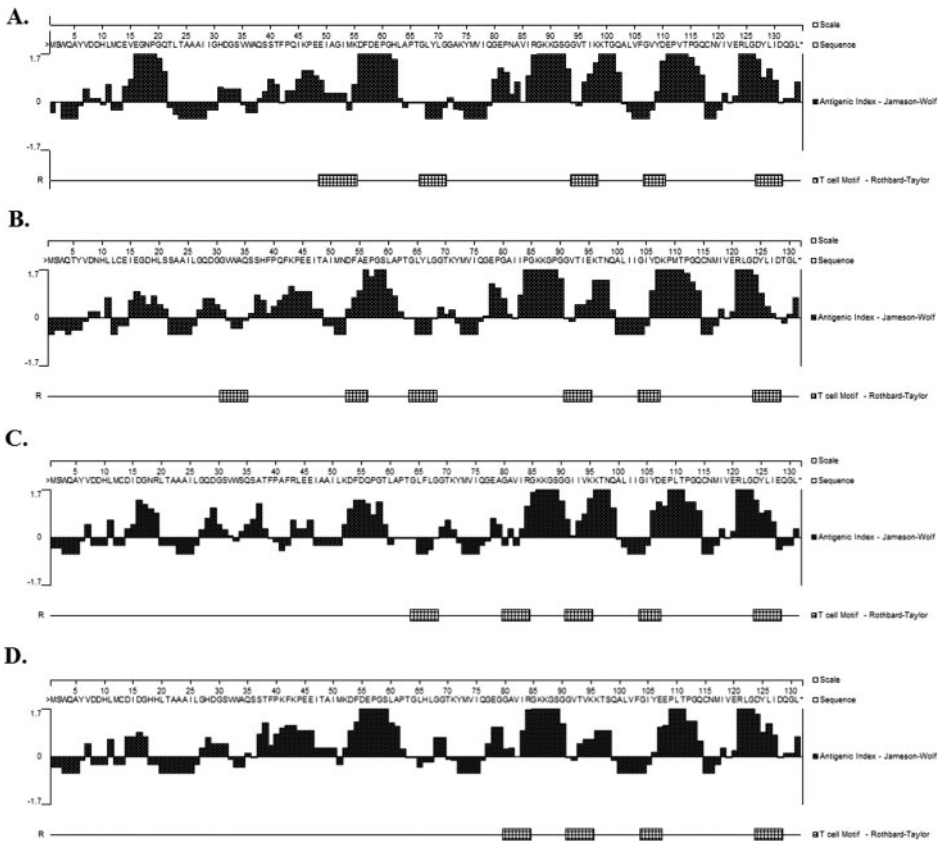


Figure 4. The proteomic analysis of typical allergens from Profilin. (A) represents Api g 4; (B) represents Ara h 5; (C) represents Cit s 2; (D) represents Pyr c 4.

epitope, at around 50th to 54th and 140th to 148th. Those two peptides are also in α -helix structure. This family was classified by sequence similarity into three subfamilies: pathogenesis-related (PR)-10, Bet v 1 major allergen and ripening-related (RRP) proteins (Liscombe, MacLeod, Loukanina, Nandi, & Facchini, 2005). PR-10 includes a majority of allergens in those three subfamilies (Wen, Vanek-Krebitz, Hoffmann-Sommergruber, Scheiner, & Breiteneder, 1997). Bet v 1 is the major birch pollen allergen. Many patients with pollen allergy are also allergic to vegetable and fruit, which suggests plant food allergens have IgE cross-reactivity with Bet v 1 (Vieths, Scheurer, & Ballmer-Weber, 2002).

3.6. The cross-food allergen families including allergens from both plants and animals

Most allergen families only include plant or animal food allergens, which explains why most studies classify plant or animal food allergens respectively. In our study, we found three cross-food allergen families: Serpin serine protease inhibitor, Thioredoxin and Triose phosphate isomerase, which cover both plant and animal food allergens. EF hand domain includes only one plant food allergen, Bra n 2 (from rapeseed); Serpin serine protease inhibitor includes Gal d 2 (from chicken) and Tri a

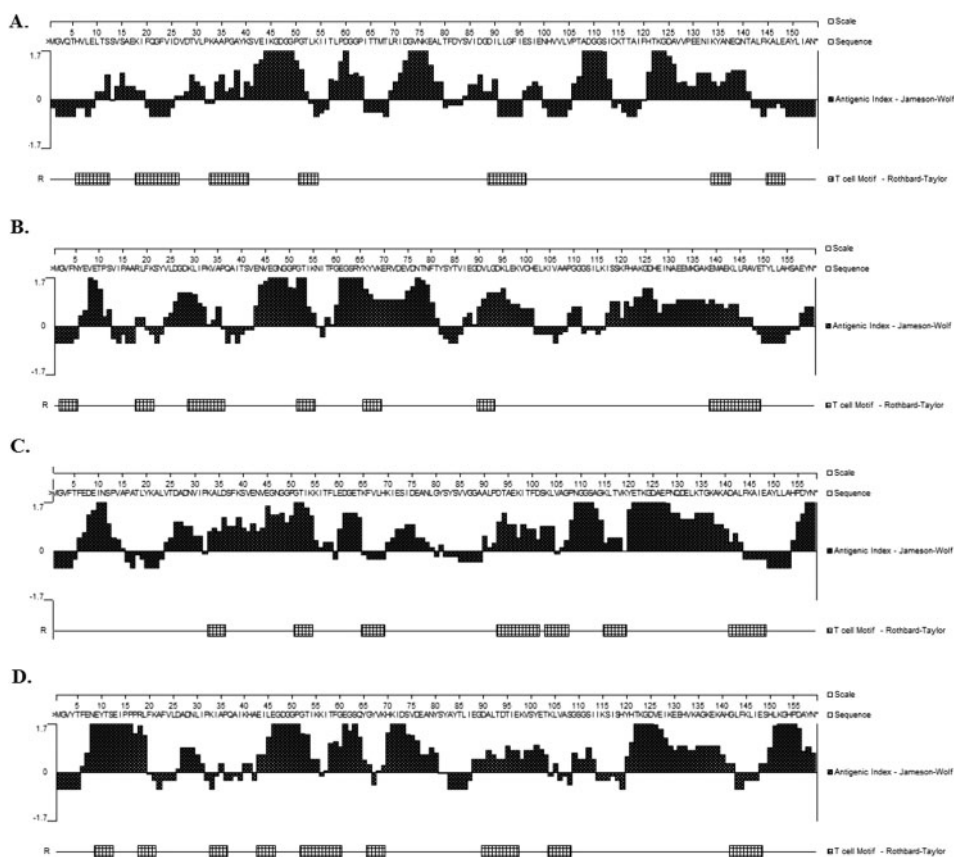


Figure 5. The proteomic analysis of typical allergens from Bet v 1-related protein. (A) represents Ape g 1; (B) represents Cor a 1; (C) represents Gly m 4; (D) represents Mal d 1.

33 (from wheat); Triose phosphate isomerase includes Arc s 8 (from Crustacean species), Cra c 8 (from North Sea shrimp), Tri a 31 (from wheat) and Tri a TPIS (from wheat). The similarity in these three families is that two of them contain allergens from wheat, which suggests that wheat allergens may share similar function or structure with animal food allergens.

Protein's analyses of some of this family's typical allergens are shown (Figure 6). The result shows that most allergens in these three cross-food families show α -helix structures. Allergens in the same family often share lots of specific peptides such as potential T lymphocyte antigen epitopes. Moreover, most of those peptides are in α -helix structure.

4. Discussion

A restricted distribution of allergens was previously found in food allergens from plant (Heimo & Christian, 2004) and animal (Jenkins et al., 2007) respectively. However, we extend the range to cover all food allergens, including both plant food allergens and animal food allergens.

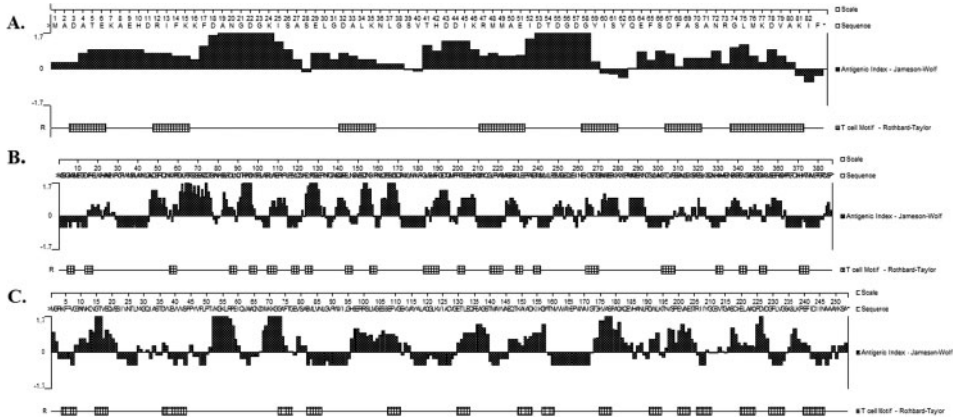


Figure 6. The protean analysis of typical allergens from three cross food allergen families. (A) represents Arc s 8.0101; (B) represents Gal d 2; (C) represents Tri a 25; (D) represents Zea m 25.

In this study, we take as reference the food allergen list from SDAP. According to SDAP classification, there are five major food allergen families: Prolamin superfamily, Bet v 1-related protein, EF hand domain, Cupin superfamily and Profilin. The five major families take up 60% of all the already-known food allergens. That suggests most of food allergens are included in a particular minority food allergen families. Breiteneder and Jenkins also indicate that the majority of allergens belong to several protein families and some protein families even don't contain any allergenic protein (Breiteneder & Ebner, 2000; Jenkins et al., 2007), which is in line with the results of our study. Therefore, using total allergen families from AllFam to classify food allergen is more significant than classifying allergenic protein directly. Studying the structure and function of allergens from food allergen families would make it possible to understand the biological characteristics of most of the food allergens.

In this classification, Prolamin superfamily, which includes only plant food allergens, is the biggest food allergen family. Some reports have shown that Cupin superfamily includes more plant food allergens than Prolamin superfamily (Heimo & Christian, 2004; Jenkins et al., 2005). This difference may suggest that more plant food allergens in the Prolamin superfamily have been found in the past 6 years.

One of the highlights of our classification is the three cross-food allergen families. The adoption of Protean makes possible the classification and analysis of the amino acid sequence in those three cross-food allergens. As a result, we could find the shared characters in those allergens. Two of three families contain allergens from wheat. The reason why wheat allergens are included in those two cross-food allergen families may be that wheat allergens share a similar function with animal food allergens, such as metal-binding, for Bra n 2 has the capacity of binding calcium. However, the structure of allergen is most important in the assessment of potential allergenicity. Most cross-food allergens are constitutive of α -helix, and allergens from the same cross family often share similar specific peptides which contain potential T lymphocyte antigen epitopes. Because the spatial structures of most allergens from

these three cross families are still unknown, further investigation about relationship of the structure of wheat allergens and that of animal food allergens is still necessary.

Ovidiu's study has reported that allergic proteins have specific motifs when compared with non-allergic proteins from the same protein family (Ivanciuc et al., 2009). In this study, we also summarised some specific motifs of allergens in the same allergen families. However, the research of motif is still limited. Although the specific motifs of allergens are useful for the evaluation of potential allergenicity of novel protein, more information of the structures and motifs of allergens is needed.

The analysis of Protean shows that allergens from the same family may share some peptides with relation to allergenicity. Some researches suggest that turn structure is protruding, which makes it easy to be exposed on the surface of antigen and become antigenic epitopes. β -strand and α -helix are stable structures, which makes it difficult for them to form antigenic epitopes. However, in this study, we found most potential T lymphocyte antigen epitopes are in α -helix or β -strand structure, and only a few epitopes are in turn structure. This indicates that it may be inaccurate to forecast antigenic epitopes according to amino acid sequence only.

The study of specific allergenicity in food allergen families leads to a better understanding of food allergen's general allergenicity. The scientific classification of food allergens contributes to an in-depth study of the biological characteristics of common food allergens and the assessment of potential allergenicity of novel proteins.

Acknowledgements

The authors acknowledge all colleagues who contributed to this research. This work was supported by National Natural Science Foundation of China (No.81072305).

References

- Breiteneder, H., & Ebner, C. (2000). Molecular and biochemical classification of plant-derived food allergens. *Journal of Allergy and Clinical Immunology*, 106(1), 27–36. doi:[10.1067/mai.2000.106929](https://doi.org/10.1067/mai.2000.106929)
- Gao, Y., Gillen, C. M., & Wheatly, M. G. (2006). Molecular characterization of the sarcoplasmic calcium-binding protein (SCP) from crayfish *Procambarus Clarkii*. *Comparative Biochemistry and Physiology – B Biochemistry and Molecular Biology*, 144, 478–487. doi:[10.1016/j.cbpb.2006.04.007](https://doi.org/10.1016/j.cbpb.2006.04.007)
- Heimo, B., & Christian, R. (2004). A classification of plant food allergens. *Journal of Allergy and Clinical Immunology*, 113(5), 821–830. doi:[10.1016/j.jaci.2004.01.779](https://doi.org/10.1016/j.jaci.2004.01.779)
- Ivanciuc, O., Garcia, T., Torres, M., Schein, C. H., & Braun, W. (2009). Characteristic motifs for families of allergenic proteins. *Molecular Immunology*, 46(4), 559–568. doi:[10.1016/j.molimm.2008.07.034](https://doi.org/10.1016/j.molimm.2008.07.034)
- Jenkins, J. A., Griffiths-Jones, S., Shewry, P. R., Breiteneder, H., & Mills, E. N. C. (2005). Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: An insilico analysis. *Journal of Allergy and Clinical Immunology*, 115(1), 163–170. doi:[10.1016/j.jaci.2004.10.026](https://doi.org/10.1016/j.jaci.2004.10.026)
- Jenkins, J. A., Breiteneder, H., & Mills, E. N. C. (2007). Evolutionary distance from human homologs reflects allergenicity of animal food proteins. *Journal of Allergy and Clinical Immunology*, 120(6), 1399–1405. doi:[10.1016/j.jaci.2007.08.019](https://doi.org/10.1016/j.jaci.2007.08.019)
- Kreis, M., Forde, B. G., Rahman, S., Mifflin, B. J., & Shewry, P. R. (1985). Molecular evolution of the seed storage proteins of barley, rye and wheat. *Journal of Molecular Biology*, 183(3), 499–502. doi:[10.1016/0022-2836\(85\)90017-8](https://doi.org/10.1016/0022-2836(85)90017-8)

- Liscombe, D. K., MacLeod, B. P., Loukanina, N., Nandi, O. I., & Facchini, P. J. (2005). Evidence for the monophyletic evolution of benzyloquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry*, 66, 2501–2520. Retrieved from <http://www.sciencedirect.com/science/article/pii/S003194220500186X>
- Mari, A. (2001). Multiple pollen sensitization: A molecular approach to the diagnosis. *International Archives of Allergy and Immunology*, 125(1), 57–65. doi:10.1159/000053797
- Mills, E. N. C., Jenkins, J. A., Alcocer, M. J. C., & Shewry, P. R. (2004). Structural, biological and evolutionary relationships of plant food allergens sensitising via the gastrointestinal tract. *Critical Reviews in Food Science and Nutrition*, 44(5), 379–407. doi:10.1080/10408690490489224
- Radauer, C., Bublin, M., Wagner, S., Mari, A., & Breiteneder, H. (2008). Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *Journal of Allergy and Clinical Immunology*, 121(4), 847–852.e7. doi:10.1016/j.jaci.2008.01.025
- Valenta, R., Duchene, M., Ebner, C., Valent, P., Sillaber, C., Deviller, P., ... Scheiner, O. (1992). Profilins constitute a novel family of functional plant pan-allergens. *Journal of Experimental Medicine*, 175(2), 377–385. doi:10.1084/jem.175.2.377
- Vieths, S., Scheurer, S., & Ballmer-Weber, B. (2002). Current understanding of cross-reactivity of food allergens and pollen. *Annals of the New York Academy of Sciences*, 964, 47–68. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12023194>
- Wen, J., Vanek-Krebitz, M., Hoffmann-Sommergruber, K., Scheiner, O., & Breiteneder, H. (1997). The potential of Bet v 1 homologues, a nuclear multigene family, as phylogenetic markers in flowering plants. *Molecular Phylogenetics and Evolution*, 8(3), 317–333. doi:10.1006/mpev.1997.0447