

Prediction of the allergenicity of a molecule

D. Martin, M. Chapman*

Asthma & Allergic Diseases Center, Box 801355, University of Virginia, Charlottesville, VA 22908-1355, États-Unis

Summary

This presentation addresses the question as to whether it is possible to predict whether proteins will become allergens. The molecular properties of inhaled allergens and food allergens are now well known and detailed analysis of their structures have not revealed common structural features or specific motifs that are associated with induction of IgE responses. Rather, allergen dose, route of exposure and aerodynamic properties appear to be key properties that promote IgE responses in genetically predisposed individuals. These include repeated exposure to low doses of allergen (1-10 µg/year) without adjuvant on airborne particles of 10-40 µ diameter. Food allergens tend to be heat stable and pH resistant proteins that account for a high percentage of the total protein in foods such as milk, egg, peanut, fish and soybeans. Predictions of allergenicity have particular relevance to the food biotechnology industry, where proteins that would not normally be part of the food supply are being introduced in genetically modified foods. Decision tree based approaches have been developed to assess whether genetically modified foods have the potential to become allergenic. Using structural information and exposure assessments, it is possible to make rational predictions about the potential allergenicity of protein molecules. © 2001 Éditions scientifiques et médicales Elsevier SAS

This curious title, given to me by Professor Bernard David, is especially topical as risk assessment for allergenicity has come to the fore in the biotechnology industry. Is it possible in the 21st century to 'predict' whether proteins will have the capacity to induce IgE responses in humans, and what are the critical factors that influence the development of these responses? Allergens are now extremely well defined proteins – there are over 1,000 sequences of allergens and isoallergens in sequence databases. Most major allergens have been cloned and expressed, and the three dimensional structures of many are known, either from sophisticated molecular modeling techniques or a priori from X-ray crys-

tallography. Sequence homology searches and functional assays have defined the biologic function of many allergens and this has facilitated studies on whether biologic function plays a role in allergenicity. But as the late Nobel Prize winning immunologist, Sir Peter Medawar, pointed out, prediction is not science, and unpredictable results or effects are often the most challenging and stimulate the greatest scientific interest. This short essay will whether the art of prediction can be applied to scientific determination of allergenicity.

CAN WE PREDICT ALLERGENICITY FROM PROTEIN STRUCTURE?

Allergens belong to diverse protein families – they may be enzymes, enzyme inhibitors, structural proteins, or binding proteins, with a variety of biologic

* Correspondence and reprints.

E-mail address : mdc8m@virginia.edu (M. Chapman).

functions [1-3]. Detailed analysis of allergen structures has not revealed any common sequences, elements or features that are associated with the development of IgE responses. The short answer to the question as posed is 'No'! However, within protein families some scope for prediction, or at least room for manoeuvre, can be found. Most animal allergens are ligand binding proteins (or lipocalins) or, to a lesser extent, albumins [4, 5]. The lipocalin allergens have been cloned from several species including mouse, rat, dog, cow and horse. Although the lipocalins show only ~20% amino acid sequence homology, they have a conserved three dimensional structure of an eight stranded anti-parallel β barrel with an N-terminal α helix. The lipocalin allergens are secreted proteins which differ in the ligands that they bind: the rodent allergens are pheromone binding proteins, whereas the cow allergen is a palmitate and retinol binding protein. However, one might expect (or predict) that lipocalin allergens would be found in other mammalian species, such as cat, guinea pig and rabbit, which can also be allergenic. Indeed, given the potent allergenicity of cats, it is surprising that a feline lipocalin allergen has not yet been identified, making this a realistic target for future studies.

By extension, similar circumstantial predictions could be made for other allergens. Thus most (but not all) mite allergens are digestive enzymes excreted with the feces and enzymes would be targeted for investigation of allergens produced by any new mite species. Grass, tree and weed pollen allergens are usually soluble proteins that elute from pollen grains within 20-30 minutes. Among food allergens, profilins are a common cause of IgE mediated sensitization in fresh fruits and vegetables (and in tree pollens). Sensitization to profilins is widespread among food allergic patients, especially those with oral allergy syndrome, and the evidence suggests that introducing a new food containing profilins into the diet would lead to an increased prevalence of food allergy.

Similarly, invertebrate tropomyosins are common food allergens (present in shellfish and snails) and inhaled tropomyosin allergens have also been recently described in mites and cockroach [1]. Mammalian tropomyosins are not usually allergens and show only ~40% homology with their inverte-

brate counterparts. Again, the prediction would be that foods containing invertebrate tropomyosins would have the potential to become allergenic.

KEY DETERMINANTS ALLERGEN DOSE, ROUTE OF EXPOSURE AND AERODYNAMIC PROPERTIES

Inhalation of environmental allergens such as pollens, dust mites and animal dander is the most common cause of IgE responses in humans. This form of sensitization is characterized by repeated exposure to low dose of allergen (1-10 $\mu\text{g}/\text{year}$). The allergen is carried on airborne particles (pollen grains or mite feces) that are 10-40 μ in diameter and is rapidly eluted upon contact with the respiratory mucosa. The important allergens are readily soluble proteins or glycoproteins with molecular size 10-50 kDa, which facilitates passage through mucosal surfaces and into the epithelium. The ability to recognize and respond to inhaled allergen distinguishes atopic individuals from non-atopics. The immune response is characterized by induction of Th2 like cells, which produce IL-2, IL-4 and IL-5, and modulate the production of IgG1, IgG4 and IgE antibody responses. One could predict that any protein which fulfills these properties and exposure conditions would have a high probability of being allergenic in individuals who are genetically predisposed to make IgE antibody responses. Evidence from occupational allergy supports this concept. Crustacea are important allergens in seafood processing workers, where inhalation of allergens, e.g. from shrimp or snow crabs, from the vats where the foods are prepared, can result in occupational asthma [6]. Thus these allergens normally associated with food allergy become inhaled allergens under different exposure conditions. Many individuals who would previously have been considered non-atopic, develop occupational respiratory symptoms when persistently exposed to inhaled allergen. Examples include laboratory animal handlers or workers involved in industrial enzyme production. The evidence suggests that the mode of environmental exposure to allergens is a more relevant predictor of allergenicity than any intrinsic properties of the protein.

RELEVANCE TO BIOTECHNOLOGY

The ability to make rational predictions and risk assessments of allergenicity has been highlighted by the development of genetically modified (GM) foods for animal and human use. Food allergy is more common among infants and young children, but can persist into adulthood. The prevalence of food allergy ranges from 5–8% among children to 1–3% among adults. The GM foods pose a particular problem because in many cases they involve the introduction into the food supply of proteins to which humans would not normally be exposed in the diet. For example, these proteins may extend the shelf life of the food, act as cryo-protectants, or confer pesticide resistance and are produced in hosts to which humans are not naturally exposed. Food allergy presents more difficulties for diagnosis than inhaled allergy, in part because the most affected patients are often infants and also because the presence of IgE ab to foods does not show as strong a correlation with allergic symptoms. The gold standard for demonstrating that a food causes symptoms is the double blind placebo controlled food challenge (DBPCFC). However, many patients may have circulating IgE ab to food without positive DBPCFC. As with inhaled allergens, the most clinically relevant food allergens have been cloned, eg from milk, egg, peanuts, soybeans, fish and rice [7, 8]. They include albumins, globulins, caseins, storage proteins, structural proteins, and enzymes. Sensitization to foods involves repeated exposure to foods in which the allergen generally comprises a high proportion of the total food protein. This may range from 5–30% or more in some cases and, conversely, proteins that comprise less than 1% of the food are unlikely to be allergenic. In general, food allergens are heat stable and resistant to digestion. However, there are exceptions, including allergens derived from fresh fruits and vegetables.

When a Brazil nut allergen was introduced into soybeans to enrich their methionine content, the transgenic soybeans were shown to react in IgE antibody assays using sera from Brazil nut allergic patients [9]. These results showed that if a known allergen is introduced into a GM food, the allergenic properties will be transferred to that food. In that instance, the biotechnology company did not proceed

with marketing the transgenic soybeans. That study highlighted concerns about the potential allergenicity of transgenic foods, but in some respects represents a worse case scenario. Most GM foods will contain proteins that have no history of being allergens. In the US, the EPA and the FDA have adopted a decision tree approach to assessing the potential allergenicity of transgenic foods [10, 11]. If the food is not commonly allergenic, its properties are compared with those of known food allergens, in addition to standard toxicological tests, to determine whether the GM food represents a safety concern. The main properties assessed by the regulatory agencies are the sequence similarity between the GM food protein and known allergens, heat and pH stability, the abundance of the protein in the food, and whether any cross-allergenicity between the GM food and known food allergens can be demonstrated. Proteins that satisfy these requirements would not be considered a cause for concern and would generally be approved for human consumption. This approach emphasizes the value of structural and molecular studies of food allergens and how detailed analysis of allergen structure can be applied in the biotechnology industry. It is clear from the above that discussions concerning the potential allergenicity of a protein need to take into account a variety of factors, which include the protein structure, relatedness to known allergens, aerodynamic properties, dose and route of exposure and, in the case of food allergens, abundance in the food. The accumulation of evidence on each of these parameters suggests that reasonable predictions can be made about the potential allergenicity of new proteins that are introduced into the environment.

REFERENCES

- 1 Chapman MD, Smith AM, Vailes LD, Arruda LK, Dhanaraj V, Pomés A. Recombinant allergens for diagnosis and therapy of allergic disease. *J Allergy Clin Immunol* 2000 ; 106 : 409-18.
- 2 Valenta R, Lidholm J, Niederberger V, Hayek B, Kraft D, Gronlund H. The recombinant allergen-based concept of component resolved diagnostics and immunotherapy (CRD and CRIT). *Clin Exp Allergy* 1999 ; 29 : 896-904.
- 3 Stewart GA, Thompson PJ. The biochemistry of common aeroallergens. *Clin Exp Allergy* 1996 ; 26 : 1020-44.
- 4 Virtanen T., Zeiler T., Rautiainen J., Mantylarvi, R. Allergy to lipocalins: a consequence of misguided T-cell recognition of self and nonself? *Immunol Today* 1999 ; 20 : 398-400.

- 5 Spitzauer S. Allergy to mammalian proteins: at the borderline between foreign and self? *Int. Arch. Allergy Immunol* 1999 ; 120 : 259-69.
- 6 Lehrer SB, Ibanez MD, McCants ML, Daul CB, Morgan JE. Characterization of water-soluble shrimp allergens released during boiling. *J Allergy Clin Immunol* 1990 ; 85 : 1005-13.
- 7 Taylor SL, Lehrer SB. Principles and characteristics of food allergens. *Crit Rev Food Sci Nutr* 1996 ; 36 : S91-118.
- 8 Bush RR, Hefle SL. Food Allergens. *Crit Revs Food Sci Nutr*. 1996 ; 36 : S119-64.
- 9 Nordlee JA, Taylor SL, Townsend JA, Thomas LA, Bush RK. Identification of Brazil nut allergen in transgenic soybeans. *N Engl J Med* 1996 ; 334 : 688-92.
- 10 Nestle M. Allergies to transgenic foods - questions of policy. *N Engl J Med* 1996 ; 334 : 726-8.
- 11 Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. *Crit Rev Food Sci Nutr* 1996 ; 36 : S165-86.