

**In Silico Methods for Evaluating Human Allergenicity to Novel Proteins:
International Bioinformatics Workshop Meeting Report February 23–24, 2005**

Karluss Thomas,^{*,1} Gary Bannon,[†] Susan Hefle,[‡] Corinne Herouet,[§] Michael Holsapple,^{*} Gregory Ladics,[¶] Sue MacIntosh,^{//} and Laura Privalle,^{///}

* International Life Sciences Institute Health and Environmental Sciences Institute, One Thomas Circle NW, Ninth Floor, Washington, DC 20005.

† Monsanto Co. 800 North Lindbergh Boulevard, St. Louis, MO 63167

‡ Food Allergy Research and Resource Program, University of Nebraska, 255 Food Industry Building, Lincoln, NE 68583

§ Bayer CropScience, Regulatory Toxicology, 355 rue Doïstoïevski, Sophia Antipolis, France

¶ E.I. Dupont De Nemours and Company, Haskell Laboratory for Toxicology and Industrial Medicine, P.O. Box 50 1090 Elkton Road, Newark, DE 19714

// Bayer CropScience, P.O. Box 12014, 2 T.W. Alexander Drive, Research Triangle Park, NC 27709

/// BASF Plant Science, 26 Davis Drive, Research Triangle Park, NC 27709

¹ To whom correspondence should be addressed at One Thomas Circle NW, Ninth Floor, Washington, DC 20005. E-mail: kthomas@ilsi.org.

Abstract

The ILSI Health and Environmental Sciences Institute (HESI) hosted an expert workshop February 22–24, 2005 in Mallorca, Spain, to review the state-of-the-science for conducting a sequence homology/bioinformatics evaluation in the context of a comprehensive allergenicity assessment for novel proteins, to obtain consensus on the value and role of bioinformatics in evaluating novel proteins, and to discuss the utility and methods of allergen specific IgE testing in the diagnosis of food allergy. The workshop participants included over forty international experts from academia, industry, and government. The workshop was hosted by the HESI Protein Allergenicity Technical committee which has established a long-term program whose mission is to advance the scientific understanding of the relevant parameters for characterizing the allergenic potential of novel proteins.

Key words: allergy, antibody, antigen, IgE, food allergy, immunoglobulin

Introduction

A food allergy is an immunologic reaction to an otherwise harmless food or food component. Most food allergies are mediated by IgE antibody and are characteristic of type I hypersensitivity reactions. Food allergic reactions occur in individuals who are genetically predisposed to allergy and who have been previously sensitized to the allergen (Sicherer, 2000). Antigen specific IgE is produced and binds to the surface of mast cells and basophils distributed throughout the body. Subsequent exposure of sensitized subjects to the inducing antigen causes an allergic response (i.e., the release of chemical mediators from basophils and mast cells). The potential consequences of food allergy can be serious with severe reactions (i.e., anaphylaxis) occurring in approximately 3 individuals per 100,000 per year (Burks and Sampson, 1997). Food allergy occurs in only a small percentage of the population after ingestion of relatively small amounts of the offending food. The prevalence of food allergy appears to be about 3% of the adult population and 6 – 8% of young children (Sicherer *et al.* 2004). Overall, approximately 90 percent of all food allergies are associated with a small number of specific proteins represented by eight major allergens: peanuts, tree nuts, cows' milk, hens' eggs, fish, crustacea (e.g., shrimp), wheat, and soybeans (Metcalf *et al.*, 1996). The remaining 10 percent of food allergies are caused by less commonly allergenic proteins or minor allergens and affect a relatively small number of people (Hefle *et al.*, 1996). Individuals that produce IgE to specific proteins do not always exhibit allergic symptoms. Therefore, IgE binding is important but is not a surrogate for clinical allergy. Virtually all allergens are proteins but only a few of the many thousands of naturally occurring proteins found in foods are allergenic under typical circumstances of exposure. One of the characteristics of an allergenic protein is that it not only induces specific IgE production, but also contains multiple IgE binding epitopes, so that IgE on mast cells and basophils can be cross-linked. This cross-linking allows the progression of the well-known inflammatory response. In addition, to the need for a minimum of two IgE binding sites, the required distance between the two IgE binding sites for effective cross-linking is

estimated to be between 5 and 24 nm (DeLisi and Siragagnian, 1979; Metzger, 1983; Kane *et al.*, 1986); the allergen concentration should be sufficient to induce > 100 cross-links per cell and the affinity for the allergen should be such that binding to IgE should occur within 100 seconds (DeLisi and Siragagnian, 1979; Ortega *et al.*, 1988; Scheitzer-Stenner and Pecht, 1999).

Novel Protein Evaluation

Novel proteins introduced into genetically engineered plants are assessed for their potential to induce human allergic responses. A comparison of the amino acid sequence of the novel (introduced) protein to the amino acid sequences of known allergens is one of several criteria used to evaluate product safety (Metcalf *et al.*, 1996; Astwood *et al.*, 2003; Codex Alimentarius, 2003). The purpose of this screen is to determine if the introduced protein shares any sequence similarities with known allergens or gliadins (i.e., proteins derived from wheat, rye, barley, and perhaps oats that are involved in the etiology of celiac disease) that might indicate the protein could elicit a clinical reaction in a sensitized individual. It is important to note however that bioinformatics alone cannot be used to predict whether a protein will be allergenic. If a novel protein demonstrates some similarity to a known allergen, further testing such as IgE binding studies using sera from appropriately allergic subjects should be considered to confirm the match. The criteria for identifying a protein as a potential allergen have changed as more information about allergens has accumulated and search algorithms have become more sophisticated.

An important consideration for homology comparisons is the availability of representative repositories of protein data that can be used as the basis for the bioinformatics evaluation. Large scale DNA sequencing efforts have resulted in the characterization and sequencing of a large number of protein allergens that have subsequently been included in a number of allergen databases (a list of current databases is in Table 1). For proteins to be included in allergen databases, participants agreed that the protein must have clear evidence of IgE-binding and minimal criteria should be established for inclusion of a protein in a particular

database. The workshop group also agreed that standards for database maintenance, management, archival, retrieval, and updating procedures should be designed prior to the development of any database. Additionally, it was agreed that a diversity of database approaches (such as exists now) is desirable and that allergen databases should be publicly accessible.

The current practice for a bioinformatics assessment of a novel protein consists of a global sequence similarity search against an allergen database and a sliding window search designed to determine if very small fragments of the novel protein had any similarity to known allergens. The global sequence similarity search is performed using publicly available search engines such as FASTA or BLAST. Cutoff values for determining whether a novel protein has the potential to cross-react with IgE directed against known allergens of 35% identity over at least 80 amino acids were established by a FAO/WHO scientific advisory panel (2001). Their recommendation was based largely on the observation that homologous Bet v 1 proteins either failed to bind IgE from birch pollen allergic patients or bound very poorly when the percent identity of the homologous proteins dropped below 40% (Scheurer *et al.*, 1999). The Codex report (2003) also recommended using matches of 35% identity over 80 or more amino acids as a specific criterion. Conference participants were in general agreement that this type of search with these cutoffs was useful for identifying proteins that could potentially cross-react with IgE to known allergens. However, new computer assisted algorithms for identifying specific protein structures or motifs specific to allergens was presented. Three independent research groups (Stadler and Stadler, 2003; Ivanciuc *et al.*, 2003 Jenkins *et al.*, 2005) found that certain protein structures and motifs were associated with protein allergens or less than 20 protein families that contain a number of common allergens, while there are thousands of protein families that are not known to contain any allergens (Jenkins *et al.*, 2005). The general consensus of meeting participants was that this type of search showed promise and encouraged further research to establish the connection between structure and protein allergenicity.

A sliding window search was first recommended by Metcalfe *et al.* (1996) to determine if small regions of proteins that represented IgE-binding epitopes or T-cell epitopes shared identities with known allergens that would be missed in the global sequence alignment. The recommendation of a sliding window search size of 8 amino acids was considered a conservative approach that took into account the known sizes of these types of epitopes and the fact that very limited experimental data existed regarding the size and structure of epitopes on food allergens. In 2001, an FAO/WHO scientific advisory panel recommended that the size of the sliding window be decreased due to experimental evidence that IgE binding epitopes could be as small as 6 amino acids. However, several publications have demonstrated that matches of 6 amino acids with any allergen occurs frequently by chance, which limits the utility of this criteria for predicting allergenicity (Hileman *et al.*, 2002; Stadler and Stadler, 2003). Information presented at the meeting indicated that more than 150 linear IgE-binding epitopes have now been determined and have been published on a variety of allergens. Sizes of these epitopes ranged from 4 amino acids in length to as much as 23 amino acids long with the vast majority of epitopes consisting of at least 8 amino acids. Even with this large increase in experimentally determined IgE-binding epitopes workshop participants were reluctant to recommend that novel proteins be searched using an IgE epitope library, either in conjunction with or as an alternative to the sliding search window approach. Participants cited the variable quality of the IgE epitope data, the variability in the methods used to identify the epitopes, the fact that all epitopes have not been identified, and the recognition that there is very little information available regarding conformational epitopes as reasons for their reluctance to utilize an IgE epitope library. Additionally, data were presented that confirm using a sliding window search of 6 amino acids resulted in an unacceptable number of false-positive matches. The workshop participants agreed that the 6-mer sliding amino acid window search does not provide any value for the bioinformatics evaluation of novel proteins in a safety assessment. This was in general agreement with the Codex Ad Hoc Inter-governmental Task Force on Foods Derived from

Biotechnology Report (2003) that stated any sliding window search should be performed with a scientifically justified search window size. Most participants agreed that further research was required to define a useful window size to use in this type of bioinformatics search.

IgE Testing

IgE binding studies and methods for assessing functional IgE binding for diagnostic purposes and their utility and weaknesses for use in the safety assessment were also discussed. The general consensus of the workshop participants was that sera screening done for the purpose of public-health decision-making should be performed using individual sera instead of pooled sera. Participants recognized, however, that there might be availability constraints from the standpoint of individual patient sera to conduct such studies. There was also general agreement that in most cases IgE binding to carbohydrate domains is irrelevant to clinical allergy, unless there are multiple carbohydrates on the potentially allergenic protein. The need for careful clinical characterization of serum donors was also noted in order to reduce the potential for false negative or false positive testing. In particular, participants felt it important to establish a link between the sera sample and a double blind placebo controlled food challenge (DBPCFC) study. The need to demonstrate that IgE binding is functional either by in vivo methods (e.g., skin prick testing), or by basophil histamine release assays in vitro was discussed. It was noted that all of these assays can be technically challenging and require method validation in laboratories that would use them to further evaluate whether a protein that has a meaningful bioinformatics match with a known allergen is likely to cause clinical cross-reactivity. The workshop group also discussed the value of targeted sera screening as recommended by FAO/WHO (2001). The consensus of the participants was that the utility of targeted sera screening for predicting protein allergenicity has not been demonstrated, nor have specific methods for this screen been validated.

Conclusion

Overall, there was general consensus from workshop participants that bioinformatics plays a vital role in the allergy assessment of biotech proteins. Together with IgE binding studies that utilize well-characterized patient sera, Bioinformatics plays a critical role in identifying proteins with sequence or structural similarities to allergens, thus identifying proteins with the potential to cross-react with IgE and cause clinical symptoms of allergy. Workshop participants recognized that bioinformatics methods associated with allergy assessment were evolving and new methods, once they are tested and validated, may lead to improved methods for identifying potentially allergenic characteristics of novel proteins.

References

Astwood, J. D., Bannon, G. A., Dobert, R. L., and Fuchs, R. L. (2003). Food biotechnology and genetic engineering. In *Food Allergy* (D. D. Metcalfe, H. A. Sampson, and R. A. Simon, Eds.), 3rd ed., pp. 51-70, Blackwell Science, Cambridge, MA.

Burks, A. W., and Sampson, H. A. (1997). Anaphylaxis and food allergy. In *Food Allergy: Adverse Reactions to Foods and Food Additives* (D. D. Metcalfe, H. A. Sampson, and R. A. Simon, Eds.), pp. 245-257, Blackwell Science, Cambridge, MA.

Codex Alimentarius Commission (2003). Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants and Appendix IV, Annex on the assessment of possible allergenicity, 25th Session, Rome, Italy 30 June – 5 July, 2003. Pp. 47 – 60.

DeLisi, C., and Siragagnian, R. P. (1979). Receptor cross-linking and histamine release. I. The quantitative dependence of basophil degranulation on the number of receptor doublets. *J. Immunol.* **122**, 2286-2292.

Food and Agriculture Organization/World Health Organization. (2001). Evaluation of the Allergenicity of Genetically Modified foods: Report of a Joint FAO/WHO Expert Consultation, Rome, Italy, 22-25 January 2001.

Hefle, S. L., Nordlee, J. A., and Taylor, S. L. (1996). Allergenic foods. *Crit. Rev. Food Sci. Nutr.* **36**, 69-89.

Hileman, R. E., Silvanovich, A., Goodman, R. E., Rice, E. A., Holleschak, G., Astwood, J. D., Hefle, S. L. (2002). Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int. Arch. Allergy Immunol.* **128**, 280-291.

Ivanciuc, O., Schein, C. H., Braun, W. (2003). SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res.* **31**, 359-362.

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 in silico analysis. *J. Allergy Clin. Immunol.* **115**, 163-170.

Kane, P., Erickson, J., Fewtrell, C., Baird, B., and Holowka, D. (1986). Cross-linking of IgE receptor complexes at the cell surface: synthesis and characterization of a long bivalent hapten that is capable of triggering mast cells and rat basophilic leukemia cells. *Mol. Immunol.* **23**, 783-790.

Metcalfe, D. D., Astwood, J. D., Townsend, R., Sampson, H. A., Taylor, S. L., and Fuchs, R. L. (1996). Assessment of the allergenic potential of foods from genetically engineered crop plants. *Crit. Rev. Food Sci. Nutr.* **36**(Suppl), 165-186.

Metzger, H. (1983). The receptor on mast cells and related cells with high affinity for IgE. *Contemp. Top. Mol. Immunol.* **9**, 115-145.

Ortega, E., Schweitzer-Stenner, R., and Pecht, I. (1988). Possible orientational constraints determine secretory signals induced by aggregation of IgE receptors on mast cells. *EMBO J.* **7**, 4101-4109.

Scheurer, S., Son, D. Y., Boehm, M., Karamloo, F., Franke, S., Hoffman, A., Haustein, D., Vieths S. (1999). Cross-reactivity and epitope analysis of Pru a 1, the major cherry allergen. *Mol. Immunol.* **36**, 155-167.

Sicherer, S. H. (2000). Determinants of systemic manifestations of food allergy. *J. Allergy Clin. Immunol.* **106**, 251-257.

Sicherer, S. H., Munoz-Furlong, A., Sampson, H. A. (2004). Prevalence of seafood allergy in the United States determined by random telephone survey. *J. Allergy Clin. Immunol.* **114**, 159-165.

Schweitzer-Stenner, R., and Pecht, I. (1999). Parameters determining the stimulatory capacity of the type I Fc epsilon receptor. *Immunol. Lett.* **68**, 59-69.

Stadler M. B., Stadler B. M. (2003). Allergenicity prediction by protein sequence. *FASEB J.* **17**, 1141-1143.

Table 1. Allergen Databases	
Database	URL
International Union of Immunological Societies (IUIS)	http://www.allergen.org
Allergome	http://www.allergome.org
Structural Database of Allergenic Proteins (SDAP)	http://fermi.utmb.edu/SDAP
Food Allergy Research and Resource Program (FARRP)	http://allergenonline.com
PROTALL	http://www.ifr.bbsrc.ac.uk/Protall
Bioinformatics for Food Safety	http://www.iit.edu/~sgendel/fa.htm
Swiss-Prot's Allergen Index	http://www.ca.expasy.org/sprot
AllAllergy	http://allallergy.net/
Informall	http://www.foodallergens.info
Central Science Laboratory	http://www.csl.gov.uk/allergen/
AllerPRedict	http://sdmc.i2r.a-star.edu.sg/Templar/DB/Allergen/
Allermatch	http://www.allermatch.org/
International Immunogenetics (IMGT) Information System	http://imgt.cines.fr/