Mol. Nutr. Food Res. 2006, *50*, 625 – 627 DOI 10.1002/mnfr.200500270

Review

In silico predictability of allergenicity: From amino acid sequence *via* 3-D structure to allergenicity

Rob C. Aalberse[1] and Beda M. Stadler[2]

[1] Sanquin Research at CLB, Landsteiner Laboratory, Academic Medical Centre, Department of Immunopathology, Amsterdam, The Netherlands
[2] Institute of Immunology, Inselspital, Bern, Switzerland

In relation to the prediction of allergenicity three aspects have to be discussed: IgE immunogenicity, IgE cross-reactivity, and T-cell cross-reactivity. IgE immunogenicity depends largely on factors other than the protein itself: the context and dose and "history" of the protein by the time it reaches the immune system. It is, therefore, not fully predictable from structural information. In contrast, IgE cross-reactivity can be much more reliably assessed by *in-silico* homology searches in combination with *in vitro* IgE antibody assays. The *in-silico* homology search is unlikely to miss potential cross-reactivity with sequenced allergens. So far, no biologically relevant cross-reactivity at the antibody level has been demonstrated between proteins without easily demonstrable homology. T-cell cross-reactivity is much more difficult to predict than B-cell cross-reactivity. Moreover, its effects are more diverse. Yet, pre-existing cross-reactive T-cell activity is likely to influence the outcome not only of the immune response, but also of the effect phase of the allergic reaction. The question of whether any antigen can be allergenic is still a matter of debate.

Keywords: Allergen / Allergenicity / Cross-reactivity / IgE

Received: June 8, 2005; revised: February 10, 2006; accepted: April 8, 2006

1 Introduction

The "allergenic potential" of a novel protein is a reflection of at least two distinct propensities: (i) the propensity to induce IgE antibodies (*i. e*. *de novo* IgE immunogenicity, or "true" allergenicity). This involves interactions of the potential allergen with the key cells of the immune system: antigen-presenting cells (APCs), various types of T- and B-cells all interact with the allergen. In most cases, other cells, such as cells forming the nonspecific immune barriers, are also involved; (ii) the propensity to react with IgE antibodies induced by other proteins (*i. e.* cross-reactivity). The potential allergen can have consequences at two levels. It may influence the immune response in a quantitative or qualitative way or, more commonly, may interact with cell-bound IgE antibodies and trigger the effector phase of the allergic reaction. These two propensities are to a large degree independent. *De novo* IgE immunogenicity is more difficult to predict or evaluate. Cross-reactivity is more amenable to bioinformatics analysis, particularly if it is restricted to cross-reactivity at the antibody (B-cell) level. Cross-reactivity at the T-cell level is more complex, but will briefly be discussed as well.

Correspondence: Dr. Rob C. Aalberse, Sanquin Research at CLB, Landsteiner Laboratory, Academic Medical Centre, Department of Immunopathology, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands
E-mail: r.aalberse@sanquin.nl
Fax: +31-20-5123170

2 *De novo* IgE immunogenicity

The question of "what makes an allergen an allergen" has elicited much debate, which demonstrates the need for more research. In a review on the structural features of the allergens for which such information was available at the time, it was concluded that "allergens have no characteristic structural features other than that they need to be able to reach (and stimulate) immune cells and mast cells. Within this constraint, any antigen may be allergenic, particularly if it avoids activation of TH2-suppressive mechanisms (CD8 cells and TH1 cells)" [1]. At the current knowledge level, one would have added "or other regulatory T-cells, or the production of regulatory cytokines such as IL-10 or transforming growth factor beta." Structural information on additional allergens has become available, but this does not really alter the situation. A number of papers discuss algorithms to distinguish allergens from "nonallergens" based on searches

in databases of protein primary sequences [2 – 12]. One limitation of these studies is that there are no well-established criteria for designating a protein as an intrinsic nonallergen. Moreover, homology among allergens in the database and the potential effects of the human homologs of allergens is not always taken into account.

Additional (sometimes opposing) hypotheses have been proposed for *de novo* IgE immunogenicity. It has been argued that only a few antigens are allergenic [8] and that allergens share a minimal set of structures (motifs). It has been hypothesized that these few motifs (approximately 100) have a structural characteristic that plays a role in the *de novo* induction of an IgE antibody response. Therefore, the chance of a protein *not* belonging to these few families and/or motifs being an allergenic risk is thought to be small if not negligible. Acid- and protease-resistant food allergens may represent an example of such a set. Through affinity maturation during the B-cell response, cross-reactions arise against distantly related antigens from the same motif group causing cross-reactions with proteins incapable of inducing an IgE response by themselves. Furthermore, it has been proposed that strong and chronic exposure to certain antigens that are only weakly related to an allergen motif may generate low affinity antibodies that stimulate a primary response to an allergen from the same motif group [8]. Thus in order to claim that any antigen can be allergenic, it is necessary to analyze such cases with an eye for proteins with weak sequence similarity to a known allergen.

The statistics of the databases (how many allergens belong to how many families/motifs, how often is a new allergen with a "new" family/motif identified) are likely to be affected by "guided" searches: as soon as a certain type of allergen is described in one food, this type of allergen will be searched for and identified in other foods. This might lead to bias in the current databases.

It should be stressed that IgE immunogenicity is, to a very large degree, determined by factors other than the protein itself. One very important factor is the way we, humans, are exposed to the protein (as a food, as an airborne protein, or as an injected protein). The discussion is largely initiated by concerns over the potential allergenicity of novel proteins introduced into the food chain. Therefore, oral exposure is usually (even if implicitly) assumed. This introduces important aspects of thermal stability upon food preparation and digestibility in relation to gastric or intestinal fluids [13]. Yet, the databases used for *in silico* allergenicity prediction usually include not only food allergens, but also allergens from airborne particles (pollen grains, mites, and animal dandruff). Another factor is the environment in which the protein is introduced to the immune system, be it the food matrix or the intestinal milieu of the consumer [14]. This may prevent or promote loss of structure and/or steer the immune system away from the TH2 pathway.

Prediction of *de novo* IgE immunogenicity in a well-defined context (*i. e.* the same food matrix, the same immunization route, and the same host) depends largely on structural stability and only to a minor degree on intrinsic immunogenicity. This lowers the predictive value of various assessment protocols (such as the combination of *in-silico* and *in vitro* procedures) for intrinsic *de novo* IgE immunogenicity of a "novel" protein.

3 IgE cross-reactivity

Compared to predicting *de novo* IgE immunogenicity, IgE cross-reactivity can be much more reliably assessed by a combination of *in silico* homology searches and *in vitro* IgE antibody assays (http://www.fao.org/documents/show_cdr.asp?url_file=/docrep/007/y0820e/y0820e00.htm) [15, 16]. Two criteria have been suggested by the Food and Agriculture Organization of the United Nations (FAO)/World Health Organization (WHO) expert consultation: short ($N = 6$, 7, or 8) contiguous sequences of identical amino acids, or 35% identity over a sliding window of 80 amino acids. The first approach does not result in meaningful hits that fail by the second approach, whereas false-positive hits were substantially more common with the first approach. The second approach is unlikely to miss potential cross-reactivity with sequenced allergens. False positives (*i. e.* protein pairs identified as homologous by structural analysis, but not obviously cross-reactive) will be found. So far, no biologically relevant cross-reactivity at the antibody level has been demonstrated between proteins without easily demonstrable homology. One possible exception might be a type of synthetic peptides called mimotopes, which show cross-reactivity without insignificant sequence homology and that influence immune reactivity in an animal model [17]. The relevance of mimotopes to food allergy needs further study.

The major limitation of bioinformatics at the moment is the number of allergens missing from the database. In particular, minor allergens from airborne allergen sources (not only pollen, but also insects, moulds, *etc.*) are not fully represented at present. Thanks to proteomics, these deficiencies will undoubtedly be filled rapidly.

By focusing on the relationship between primary sequence and cross-reactivity, one important source of cross-reactivity is easily overlooked: post-translational modification. These processes are not fully defined by the DNA sequence, but are determined largely by the host in which the protein is expressed. Such cross-reactivity would be mostly due to an anticarbohydrate response that often leads to an extraordinarily broad cross-reactivity pattern.

4 Cross-reactivity of T-cells

T-cell cross-reactivity is much more difficult to predict than B-cell cross-reactivity, and its effects are more diverse. Yet, pre-existing cross-reactive T-cell activity is likely to influence the outcome not only of the immune response, but also of the effector phase of the allergic reaction. Most work on T-cell cross-reactivity stems from research on autoimmune disorders. In this context, molecular mimicry is the term most commonly used. It specifically relates to T-cell cross-reactivity between microbial antigens and autoantigens, in which the micro-organism avoids detection by the immune system by mimicking host structures. There is an impressive body of data showing the striking lack of specificity (*i. e.* degeneracy) of many T-cell-peptide-MHC interactions [18 – 25]. Despite the well-established fact that longer peptides are involved in the T_H cell-MHC class II interaction than in the T_C-MHC class I interaction, the specificity of the former (CD4+ T-cell-MHC class II) is even lower than the latter. Using larger combinatorial peptide libraries, cloned CD4+ T-cells can in some case be stimulated by peptides (in MHC class II context) that lack recognizable homology.

These data support several conclusions concerning IgE-mediated allergy, two of which are: (i) we have to be extremely careful in interpreting allergen-induced T-cell activation (proliferation or cytokine production) as evidence of prior contact between the allergen and the immune system; (ii) the accidental history (type of infections, *etc.*) of the T-cell branch of the immune system in an individual may markedly influence (positively or negatively) the outcome of a subsequent contact with an allergen (*i. e.* a food allergen) that happens to generate a peptide that cross-reacts on the T-cell level. This may sound like a rephrasing of the hygiene hypothesis, but it is important to appreciate that in this case the infectious agent acts in an immunologically specific (if degenerate) way, rather than by stimulating a TH2–suppressive milieu in a nonspecific way.

5 Conclusions

(i) Consensus is still lacking on the predictability of *de novo* allergenicity. It has been argued that bioinformatics can predict *de novo* allergenicity of a protein only to a very limited degree because *de novo* allergenicity is strongly influenced by extrinsic factors. A different, more optimistic, view is that allergens belong to a small subset of protein families, which may be predicted by bioinformatics on the basis of inherent molecular structures (*i. e.* epitope motifs).

(ii) Cross-reactivity at the level of the B-cell or the IgE antibody is more predictable than allergenicity. (The discussion of the value, or lack of value, of 6, 7, or 8 contiguous peptides is by now obsolete.) However, *in-silico* homology search can only be performed for allergens that are in the database. For allergens whose sequence is unknown (largely minor allergens), serum screening protocols may be useful.

(iii) T-cell cross-reactivity in relation to allergic sensitization has not been studied in sufficient detail (because it is a very complicated subject), but it may have a significant impact on IgE-dependent allergenicity.

6 References

[1] Aalberse, R. C., *J. Allergy Clin. Immunol*. 2000, *106*, 228 – 238.

[2] Gendel, S. M., Adv. *Food Nutr. Res.* 1998, *42*, 45 – 62.

[3] Gendel, S. M., *Adv. Food Nutr. Res.* 1998, *42*, 63 – 92.

[4] Gendel, S. M., *Ann. N. Y. Acad. Sci.* 2002, *964*, 87 – 98.

[5] Ivanciuc, O., Schein, C. H., Braun, W., *Bioinformatics* 2002, *18*, 1358 – 1364.

[6] Ivanciuc, O., Schein, C. H., Braun, W., *Nucleic Acids Res.* 2003, *31*, 359 – 362.

[7] Ivanciuc, O., Mathura, V., Midoro-Horiuti, T., Braun, W., *et al.*, *J. Agric. Food Chem.* 2003, *51*, 4830 – 4837.

[8] Stadler, M. B., Stadler, B. M., *FASEB J.* 2003, *17*, 1141 – 1143.

[9] Brusic, V., Petrovsky, N., *Trends Immunol.* 2003, *24*, 225 – 228.

[10] Soeria-Atmadja, D., Zorzet, A., Gustafsson, M. G., Hammerling, U., *Int. Arch. Allergy Immunol.* 2004, *133*, 101 – 112.

[11] Li, K. B., Issac, P., Krishnan, A., *Bioinformatics* 2004, *20*, 2572 – 2578.

[12] Furmonaviciene, R., Shakib, F., *Mol. Pathol.* 2001, *54*, 155 – 159.

[13] Thomas, K., Aalbers, M., Bannon, G. A., Bartels, M., *et al*., *Regul. Toxicol. Pharmacol.* 2004, *39*, 87 – 98.

[14] Untersmayr, E., Scholl, I., Swoboda, I., Beil, W. J., *et al.*, *J. Allergy Clin. Immunol.* 2003, *112*, 616 – 623.

[15] Aalberse, R. C., Stapel, S. O., *Pediatr. Allergy Immunol.* 2001, *12(S14)*, 10 – 14.

[16] Aalberse, R. C., Akkerdaas, J., Van Ree, R., *Allergy* 2001, *56*, 478 – 490.

[17] Jensen-Jarolim, E., Wiedermann, U., Ganglberger, E., Zürcher, A., *et al.*, *FASEB J.* 1999, *13*, 1586 – 1592.

[18] Hunig, T., Bevan, M. J., *Nature* 1981, *294*, 460 – 462.

[19] Scherer, A., Noest, A., de Boer, R. J., *Proc. R. Soc. Lond. Biol. Sci.* 2004, *271*, 609 – 616.

[20] Parnes, O., *Mol. Immunol.* 2004, *40*, 985 – 991.

[21] Wucherpfennig, K. W., *Mol. Immunol.* 2004, *40*, 1009 – 1017.

[22] Ford, M. L., Evavold, B. D., *Mol. Immunol.* 2004, *40*, 1019 – 1025.

[23] Wilson, D. B., Wilson, D. H., Schroder, K., Pinilla, C., *et al*., *Mol. Immunol.* 2004, *40*, 1047 – 1055.

[24] Zhou, D., Hemmer, B., *Mol. Immunol.* 2004, *40*, 1057 – 1061.

[25] Nino-Vasquez, J. J., Allicotti, G., Borras, E., Wilson, D. B., *et al*., *Mol. Immunol.* 2004, *40*, 1063 – 1074.

