Prediction and mapping of IgE motif epitopes in proteins of Genetically Modified Foods for Immunotherapy strategy

P Somvanshi, V Singh, P Seth

Citation

Abstract
There is a constant need for assessment of potential allergenic and their patterns of cross-reactivity of genetically modified food (GMF) and pharmaceutical; in which novel proteins are introduced into the human food chain. The aspect for its identification is still in its nascent stage, which includes detection of motifs commonly, occurred in many allergens but rare in ordinary proteins. Transgenic proteins were used by GMF are evaluated on the basis of potential allergenic properties, identification of short identical amino acid sequences were detected from transgenic proteins In this study, bioinformatics approach had been used to detect the potential allergens motifs in amino acid sequences. The purpose of the analysis is to suggest a guideline for food safety, which will aid in human welfare and thereby, controlling the extra use of allergen based transgene introduced into the transgenic foods. Therefore, computationally predicted IgE epitopes from GMF allergen might be an ideal peptide-based novel candidate for immunotherapy.

INTRODUCTION
In most nations assessment before authorities is required for the safety of genetically engineered foods prior to its launch in the market needs approval. An important issue in current food safety assessment is the evaluation of the potential allergenicity of food derived from biotechnology products. Since many food allergens are proteins, introduction of a new protein in food by genetic engineering can in assumption cause allergic reactions. Therefore, the allergenicity of novel proteins needs to be evaluated. Potential allergenicity of a protein is a complex issue and various tests can be made for the prediction, includes in silico as well as in vitro digestibility and binding antisera of allergic patients.

Allergy is caused by adverse immune responses to otherwise innocuous proteins, the allergens. It involves a series of reactions both intrinsic and extrinsic factors contributes in the development of disease and triggering the symptoms. Type I hyper - sensitive reaction is induced by certain types of antigens referred to as allergens that elicit specific IgE antibodies or from cross reactivity between common homologous allergens from different sources (Santos et al 1999). Atopic allergy and other forms of hypersensitivity affect up to 15–20% of the population in industrial nations. The estimated prevalence of food allergenicity among general population in European Union ranges from about 2.5 to 3.2% (Jansen et al., 1994; Kanny et al., 2001; SCP, 1998). Typical allergy (Type I hyper sensitivity reaction) symptoms are rhinitis, asthma and atopic eczema, but more severe reactions such as acute and possibly fat an anaphylactic shock can occur also.

In atopic individuals, sensitizing T-cell epitopes can trigger a cascade of events that leads to synthesis of allergen-specific immunoglobulin E (IgE) antibodies as well as other immunological reactions. The IgE antibodies bind the intruding allergen, or a structurally similar cross-reacting protein, leading to release of mediators, which causes allergic reactions. Mediators released by activated cells cause the symptoms of allergy, such as sneezing and swelling of the mucosa, characteristic for allergic rhinitis, allergic conjunctivitis, and asthma. Hence, T-and B- (IgE) epitopes are both relevant targets for models and the detection of protein allergens. The former type is generally confined to continuous motifs of about 8–24 amino acid residues, whereas the latter may occur as scattered regions, which are brought together on the three-dimensional surface of the protein (Bredehorst and David, 2001).

The prediction of an allergenic protein is important presently due to the specific use of modified proteins in edible foods, therapeutics and biopharmaceuticals (Goodman et al 2005). World Health Organization (WHO) and the Food and
Agriculture Organization (FAO) proposed a guideline to assess the potential allergenicity of proteins (FAO/WHO, 2003). The aim of the present study was to predict the IgE specific allergenic motif and novel strategies for immunotherapy against the genetically modified foods for safe food and health management.

MATERIALS AND METHODS

COLLECTION OF ALLERGENS

The data sets used in this study were taken from NCBI and allergen database. Different types of transgenic proteins food allergen proteins sequences were scanned for IgE epitopes. We have scanned total 52 proteins for detection of allergenic IgE epitopes. These IgE epitopes were compared with dataset of allergic and non-allergic proteins.

PROTEIN FEATURES AND AMINO ACID COMPOSITION

Amino acid composition is the reaction of each amino acid in a protein. The fraction of all 20 natural amino acids was calculated using the following equations:

**Figure 1**

\[
\text{Fraction of amino acid } i = \frac{\text{Total number of amino acids } i}{\text{Total number of amino acids in protein}}
\]

where \(i\) can be any amino acid.

DIPEPTIDE COMPOSITION

Dipeptide composition was used to encapsulate the global information about each protein sequence, which gives a fixed pattern length of 400 (20 x 20). This representation encompassed the information about amino acid composition along local order of amino acid. The fraction of each dipeptide was calculated using following equation:

**Figure 2**

\[
\text{Fraction of dipep } i = \frac{\text{Fraction of dipep } i}{\text{Total number all possible dipeptides}}
\]

Where dipep \(i\) is 1 out of 400 dipeptides.

FIVE FOLD CROSS-VALIDATION

The performance of all methods developed in this study is evaluated using 5-fold cross validation. In 5-fold cross validation the data set has been divided into five sets, where each set has nearly equal number of allergens and non-allergens. The training and testing of every method has been carried out five times, each time using one distinct set for testing and remaining four sets for training. The over all performance of a method is the average performance over five sets. Performance measures a standard set of parameters has been used to evaluate the performance of various methods developed in this study. Following is a brief description of the parameters (Baldi et al 2000): (i) sensitivity, also referred to as recall, is the percent of correctly predicted allergen epitopes; (ii) specificity is the percent of correctly predicted non- allergen epitopes; (iii) accuracy is the proportion of correctly predicted epitopes; (iv) PPV (positive prediction value, also referred to as precision) is the probability of correct positive prediction (Li et al 2004); (v) NPV (negative prediction value) is the probability of correct negative prediction; and (vi) Matthew’s correlation coefficient (MCC).

RESULTS AND DISCUSSION

In the current study, prediction of IgE allergenic motif (IgE allergenic epitope) from 52 known transgenic proteins previously introduced into the animals (27) and plants (25) for their growth rate, disease resistance and cold resistance (Table 1 & 2).
Six protein sequences from plant and animal showed the potential of allergenic motif where as 14 protein sequences confirmed that they were allergen from in silico study. It indicates future transgenesis of any animal or plant specifically, researcher should attempt to remove the allergenic motif or mutate the region that was introduced into the protein. The first genetically modified (GM) crops approved for eating (tomato and soybean) were evaluated for safety by USFDA prior to its commercial production. Among other factors, all additional GM crops that have been grown commercially were evaluated for potential increases
in allergenic properties using methods that are consistent with the current understanding of food allergens and knowledge regarding the prediction of allergenic activity. Although there have been refinements, the key aspects of the evaluation have not been changed. The allergenic properties of the gene donor and the host (recipient) organisms are considered in determining the appropriate testing strategy. The amino acid sequence of the encoded protein is compared with all known allergens to determine whether the protein is a known allergen or is sufficiently similar to any known allergen to indicate an increased probability of allergic cross-reactivity (Goodman et al 2005).

Introduction of foods derived from genetically modified crops in the marketplace, scientific community, regulatory bodies and international associations have intensified discussions on risk assessment procedures to identify potential food allergenicity of the newly introduced proteins. In this work, a novel biocomputational methodology was presented for the classification of amino acid sequences with regard to food allergenicity and non-allergenicity. In this study, food allergens from several specialized public repositories of food allergy and the SWALL database were identified pre-processed and stored yielding one of the most extensively characterised repositories of allergenic sequences known today. The biocomputational approach presented here should consider as a significant extension and refinement of earlier attempts suggested for in silico food safety assessment (Zorzet 2002). The group 5 grass allergens are characterized by repeated structural motifs. Using a new algorithm, TEPITOPE we predicted promiscuous HLA-DR ligands within the repeated motifs of the Lol p5a allergen from rye grass. In vitro binding studies confirmed the promiscuous binding characteristics of these peptides. Moreover, most of the predicted ligands were novel T cell epitopes that were able to stimulate T cells from atopic patients. We generated a panel of Lol p5a-specific T cell clones, the majority of which recognized the peptides in a cross-reactive fashion. The computational prediction of DR ligands might thus allow the design of T cell epitopes with potential useful application in novel immunotherapy strategies (de Lalla et al 1999).

Laboratory assessment of protein allergenic potential, including immunogenicity, cross-reactivity and clinical symptoms studies are both time-consuming and cumbersome. While animal studies of immune responses by means of measuring immunoglobulin E (IgE) antibody or T-cell response, provide the ultimate validation of an allergen, they are too expensive to be applied to every protein. Therefore, a reliable prediction of potential allergens is imperative for allocating the money to study the real potential allergens, rather than all proteins.

Biotechnology allows for relatively precise target base modification of the genome in plants, which can serve the purpose of inhibiting or even disrupting gene activity activating existing genes or more commonly accomplish the introduction of new genes. The most common genetically modified crops currently used in agriculture are soybean, corn and rapeseed expressing new proteins that impart either herbicide tolerance or insecticide resistance to the crop (IFT, 2000). Because of the relatively high incidence of food allergenicity in Western societies, entailed to serious outcomes in many cases, special attention is given to risk assessment of the potential allergenicity of transgenic proteins expressed by GM crops. Concerning sequence similarity, the FAO/WHO report delineates a bifurcated query procedure for determination of sequence similarity between a test protein and a set of allergenic molecules. The first criterion is based on identical amino acids (6 or more). There is an appreciable risk of 6 contiguous amino acids occurring by chance. Therefore, verification of cross-reactivity is warranted when only this criterion is met in the alignment procedure between a test sequence and the allergenic sequence data (FAO, 2001). A recent report demonstrates large numbers of spurious hits using an alignment setting of 6 amino acids joined in order as a limit for alarm, and 8 amino acids is shown to produce more relevant sequence identifications (Hileman, et al 2002).

These transgenic foods might be safer than the normal transgenic foods if someone will be suffering from asthma or any allergic disease. It shall be a great challenge for researcher in implying the safety foods. Commercial production of meat from the genetically modified animals is important for increasing the country's economy.

In conclusion, the study was carried out for transgenic proteins which had been identical stretches of six or seven amino acids in common with allergenic proteins. As evident in this study, identical stretch can be further screened for relevant comparison with linear IgE-binding epitopes. In the absence of data on epitopes, antigenicity prediction by computer aide to select potential antibody binding sites that will need verification of IgE binding by sera binding tests. Foods produced through agricultural biotechnology are
directly reaching the consumer marketplace. These novel foods should assess for their safety, including their potential allergenicity. Agricultural biotechnology involves the introduction of novel proteins into the modified foods, and proteins, which can be allergenic. Additional factors, such as level of expression of novel protein in the modified food and expression in the edible portion of food may also yield valuable insights. The result showed that the transgenic novel proteins are likely to become an allergen. Finally, the positive outcomes of this approach warrant further clinical testing for potential allergenicity and strategies for novel immunotherapy against the allergy and asthma. However, further investigations were needed to evaluate the predicted allergenic motifs in genetically modified foods.

References
Author Information

Pallavi Somvanshi
Sector-G Jankipuram

Vijay Singh
Department of Biotechnology, National Bureau of Fish Genetic Resources

P. K. Seth
CEO, Biotech Park, Sector-G Jankipuram