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

P Somvanshi, V Singh, P Seth

Citation

P Somvanshi, V Singh, P Seth. *Prediction and mapping of IgE motif epitopes in proteins of Genetically Modified Foods for Immunotherapy strategy*. The Internet Journal of Infectious Diseases. 2007 Volume 6 Number 2.

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 acidc omposition along local order of amino acid. The fraction of each dipeptide was calculated using following equation:

Figure 2

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

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).

Figure 3

Table 1: Prediction of allergenic IgE motifs in transgene promoter, enhancer and structural gene introduced into genetically modified plants

Accession No	Gene product/Plant	Allergen	Non-allergen	Potential Allergen
CAA25390.1	ribulose bisphosphate carboxylase [*Pisum sativum*]		yes	
P69249	Ribulose bisphosphate carboxylase chloroplast precursor		Yes	
AAA82069.1	ribulose 1,5-bisphosphate carboxylase		yes	
AAA63413.1	cab precursor		Yes	
S04270	chlorophyll a/b-binding protein type II - wheat	Yes		
P04784	WHEAT Chlorophyll a-b binding protein, chloroplast precursor		Yes	
CAA28450.1	ST-LS1 protein [*Solanum tuberosum*]		yes	
P06515	Chalcone synthase		Yes	
CAE53111.1|	protease inhibitor [*Oryza sativa (indica)*	yes		
AAC97524.1	protease inhibitor [*Glycine max*]		Yes	
CAA25592.1|	patatin [*Solanum tuberosum*]		Yes	
1633290	Leghemoglobin Complexed With Nicotinate			Yes
GPSYC2	leghemoglobin c2 - soybean			Yes
CAA28471.1	nodulin [*Glycine max*]	Yes		
AAC28907.1	phaseolin G-box binding protein PG2 [*Phaseolus vulgaris*]		Yes	
21465633|pdb|1IPK|C	Recombinant And Native Soybean Beta- Conglycinin Beta Homotrimers			Yes
CAB96392.1|	lectin [*Phaseolus lunatus*]		Yes	
AAS45646.1|	lectin [*Glycine max*]		Yes	
1421113|pdb|1BCS|B	Wheat Serine Carboxypeptidase, Cpdw-Ii, With The Microbial Peptide Aldehyde Inhibitor, Chymostatin,		Yes	
1421112|pdb|1BCS| Degrees Kelvin	A Chain A, Complex Of The Wheat Serine Carboxypeptidase, Cpdw-Ii, With The Microbial Peptide Aldehyde Inhibitor, Chymostatin, And Arginine At 100	Yes		
BAA11642.1|D	hordein [*Hordeum vulgare subsp. vulgare*	Yes		
157879614|pdb|1P8B|A	A 37-Amino Acid Insecticidal Protein Extracted From Pea Seeds (*Pisum Sativum*)		Yes	
114176|sp|P11043|	AROA_PETHY 3-phosphoshikimate 1-carboxyvinyltransferase, chloroplast precursor (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase)		Yes	
BAA24017.1|	ribonuclease [*Nicotiana alata*]	Yes		
AAZ67565.1|52008_19	*Brassica rapa subsp. pekinensis*		Yes	

Figure 4

Table 2: Prediction of allergenic IgE motifs in transgene promoter, enhancer and structural gene introduced into genetically modified animals

Transgenic animals	Protein designation	Accession No	Gene product/Animal	Allergen	Non-allergen	Potential Allergen
Chicken	TVCHLV	gi|1070476|pir	epidermal growth factor receptor precursor	Yes		
	V-REL	NP_990181.1	reticuloendotheliosis viral oncogene homolog B, nuclear factor of kappa light polypeptide gene enhancer in B-cells 3		Yes	
	GRF/GHRH	NP_001032923.1	growth hormone releasing hormone receptor		Yes	
Cow	BLV-R	AAB32770.2	bovine leukemia virus cell receptor		Yes	
	Lactoferrin	AAA30610.1	lactoferrin	Yes		
	GRF	NP_847895.1	growth hormone releasing hormone		Yes	
Fish	MT	AAF22355.2	metallothionein [Dicentrarchus labrax]	Yes		
	MT	CAA07557.1	metallothionein [Pagothenia borchgrevinki]	Yes		
	cd-crystallin	BAA03718.1	PP-CAT [Photobacterium damselae subsp. piscicida]	Yes		
	SV/hygro	sp|Q61597|CRGC	MOUSE Gamma-crystallin C (Gamma-C crystallin)		Yes	
	Hygro	ABM60744.1	hygromycin phosphotransferase [Binary vector pCAMBIA1300-FAD2A]		Yes	
	SV40	NP_043128.1	hypothetical protein SV40gp7 [Simian virus 40]		Yes	
	AFP	ABA41373.1	type III antifreeze protein [Lycodichthys dearborni]	Yes		
	AFP	sp|Q01758|	ISP2_OSMMO Type-2 ice-structuring protein precursor (Type II antifreeze protein) (AFP)		Yes	
Pig	GH	gi|134715|sp|P01248|	SOMA_PIG Somatotropin precursor (Growth hormone)		Yes	
	MT	NP_001001266.1	metallothionein	Yes		
	Gal	AAZ03639.1	beta-glucuronidase		Yes	
	GH	NP_999034.1	growth hormone		Yes	
	MLV	gi|7546226|pdb|1D0E|B	Chain B, Crystal Structures Of The N-Terminal Fragment From Moloney Murine Leukemia Virus Reverse Transcriptase Complexed With Nucleic Acid: Functional Implications For Template-Primer Binding To The Fingers Domain		Yes	
	Prolactin	NP_999091.1	prolactin		Yes	
	GRF/GHRH	NP_999200.1	growth hormone releasing hormone receptor		Yes	
Sheep	Prolactin	NP_001009306.1	prolactin		Yes	
	MT	gi|54037843|sp|P67982	Metallothionein-1A (MT-1A) (Metallothionein-IA) (MT-IA) (MTC)	Yes		
	TK	NP_659638.1	Thymidine kinase [Sheeppox virus]			Yes
	GRF/GHRH	NP_001009454.1	growth hormone releasing hormone receptor		Yes	
	FIX	AAA31520.1	factor IX			Yes
Goat	GRF/GHRH	gi|54042072|sp|P63293|SLIB_CAPHI	Somatoliberin (Growth hormone-releasing factor) (Growth hormone-releasing hormone)		Yes	
	FKV	gi|13399947|pdb|1FKV|A Chain A	Recombinant Goat Alpha-Lactalbumin T29i			Yes

Note: ALV: avain leukosis virus, X/AT: x1 antitrypsis, BPV: bovine papilloma virus, Eu: Immunoglobulin heavy chain, FIX: Factor IX, GH: growth hormone, GRF: growth releasing factor, Gal: beta-galactosidase, Hygro: hygromycin, BLG: beta-lactoglobulin, MT: metallothionein, MLV: moleny murine leukemia virus, c-myc: murine leukemia virus, REV: reticuloendotheliosis, PRL: prolactin, SV: SV40, TK: thymidine kinase, AFP: antifreeze protein, TPA: human tissue plasmonogen activator., BLV-R: bovine leukemia virus cell receptor.

Six protein sequences from plant and animal showed the potential of allergenic motif where as 14 protein sequences confirmed that they were allergen from insilico 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

1. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A. and Nielsen, H. (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics, 16,412ñ424.
2. Bredehorst, R., and David, K. (2001) What establishes a protein as an allergen? J. Chromatogr. B Biomed. Sci. Appl. 756, 33-40.
3. de Lalla C, Sturniolo T, Abbruzzese L, Hammer J, Sidoli A, Sinigaglia F, Panina-Bordignon P. (1999). Cutting edge: identification of novel T cell epitopes in Lol p5a by computational prediction. J Immunol, 163(4):1725-9.
4. FAO / WHO(2003) Report of the fourth session of the codex ad hoc intergovernmental task force on foods derived from biotechnology. (http://www.codexalimentarius.net/download/report/46/Al0334ae.pdf).
5. FAO/WHO (2001) Evaluation of allergenicity of genetically modified foods. Report of a joint FAO/WHO expert consultation on allergenicity of foods derived from biotechnology (http://www.fao.org/es/ESN/food/ pdf / allergygm. pdf).
6. Goodman RE, Hefle SL, Taylor SL, van Ree R. 2005. Assessing genetically modified crops to minimize the risk of increased food allergy: a review. Int Arch Allergy Immunol. 137(2):153-66.
7. Hileman, R. E., Silvanovich, A., Goodman, R. E., Rice, E. A., Holleschak, G., Astwood, J. D. and Hefle, S. L. (2002). Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. Int. Arch. Allergy Immunol. 128, 280- 291.
8. IFT (2000). IFT Expert Report in Biotechnology and Foods: Introduction. Food Technol. 54, 124-136.
9. Jansen, J.J., Kardinaal, A.F., Huijbers,G., Vlieg-Boerstra, B.J., Martens, B.P. and Ockhuizen,T. (1994) Prevalence of food allergy and intolerance in the adult Dutch population. J. Allergy Clin. Immunol., 93,446-456.
10. Kanny,G., Moneret-Vautrin,D.A., Flabbee,J., Beaudouin,E., Morisset,M. and Thevenin, F.(2001) Population study of food allergy in France. J. Allergy Clin. Immunol., 108,133-140.
11. Li K-B, Issac P and Krishnan A 2004. Predicting allergenic proteins using wavelet transform. Bioinformatics. 20: 2575-2578.
12. Santos,A.B.,Chapman,M.D.,Aalberse,R.C.,Vailes,L.D.,Ferriani,V.P., Oliver, C., Rizzo,M.C., Naspitz,C.K. and Arruda, L.K. (1999). Cockroach allergens and asthma in Brazil: identification of Tropomyosin as a major allergen with potential cross-reactivity with mite and shrimp allergens. J. Allergy Clin. Immunol., 104, 329-337.
13. SCP (1998). The occurrence of severe food allergies in the EU. Report of experts participating in task 7.2, Scientific Cooperation Programme in the EU.
14. Zorzet A, Gustafsson M, Hammerling U. (2002) Prediction of food protein allergenicity: a bioinformatic learning systems approach. In Silico Biology. 2(4): 525-34.

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