

ALLERGEN DOMAIN ONLINE: A DATABASE FOR DOMAINS OCCURRING IN THE ALLERGENS

Monendra Grover, Gurinder Jit Randhawa and Vijay Kumar Sharma

National Bureau of Plant Genetic Resources,

New Delhi-110012

ABSTRACT

For the commercialization of GM food crops the biosafety issues pertaining to their allergenicity need to be thoroughly investigated. Sequence homology with the known allergens can serve as possible indicator of allergenicity since sequence homology between the proteins has high probability of causing cross-reactions. However, structures with low sequence similarity may also cause cross-reactivity due to the fact that the structural similarity is possible even in absence of sequence similarity. The database developed in this study provides the information about the domains in the allergens listed in the allergen online database (<http://www.allergenonline.org>). The database presented in this paper can be searched for the domains deduced to be present in transgenic proteins or other query proteins and can serve as an indicator for detection of possible allergenicity.

Keywords : GM , Biosafety, Allergenicity, Codex Alimentarius

INTRODUCTION

The safety of the GM crops needs to be thoroughly assessed before their introduction into the food chain. The Codex Alimentarius commission has recommended that following factors be taken into account for assessing the safety of the GM crops. I) the source of the protein ii) sequence similarity with known allergenic proteins iii)in vitro digestibility and degradability, iv) sera binding tests v) animal models and clinical tests. A weight of evidence approach giving different importance to different factors listed above has been recommended by the Codex Alimentarius commission.

(http://library.wur.nl/frontis/allergy_matters/10_kleter.pdf).

As mentioned above sequence similarity with the known allergenic proteins has been one of the important in silico tests advocated by Codex Alimentarius commission. It has been suggested that the sequence similarity may

lead to structural similarities. Accordingly a decision tree has been advocated by FAO/WHO [1] for allergenicity assessment. While this is indeed true, the structural similarities are possible even without sequence similarities [2]. Accordingly we have proposed modifications to the decision tree [3], where we have suggested that domain similarity should be included in the in silico tests for allergenicity risk assessment.

Thus, besides conducting sequence similarity searches between the GM protein and the allergens, domain similarity between the two also needs to be addressed. To achieve this, domains present in the novel GM proteins can be assessed by using bioinformatic tools. Following this one can check whether these domains are present in the known allergens. To facilitate this we have created an independent database which lists all the domains present in the allergen online

database. Though one can find the domains present in the allergens using allergen online database, there is no option for searching the domains present in allergens for a particular domain. Our database provides the option for searching the domains present in allergens for a particular domain. The database is available online at <http://nabg.iasri.res.in/allergen/>.

The results of the search are finally used to identify recombinant proteins expressed in transgenic plants, which are tested by specific serum screens, and to identify the specific allergic populations that might be at risk and should be used as serum donors for specific testing [4]

MATERIALS AND METHODS:

Allergen Domain online was constructed using SQL server 2005 at back end and Microsoft visual studio 2005 at front end. The entries in the allergen online database (<http://www.allergenonline.org>) were searched for the domains present in allergens. An excel table was prepared which was transferred to a SQL database. A screenshot of the Allergen Domain Online is given in Fig. 1

RESULTS AND DISCUSSION

Commercialization of GM crops is a highly sensitive issue and needs to be thoroughly regulated. A two tier weight of evidence approach strategy has been advocated by International Food Biotechnology Committee (ILSI) to determine the safety of novel GM proteins introduced in the food chain. The Tier I includes an assessment of the biological function or mode of action and intended application of the protein, history of safe use, comparison of amino acid sequence of the protein to other proteins as well as the biochemical and physico-chemical properties of the proteins. The second tier consists of studies (hazard characterization) are conducted when the results from tier I are not sufficient to permit conclusions on a case by case basis.

An important factor to be considered before the commercialization of GM crops is that the newly introduced gene in the transgenic plants may code for an allergen. Sequence and structural similarity with the known allergens may be important determinants for in silico analysis of the allergenicity of the novel proteins. Our database facilitates the search of similarity at the domain level.

Many computational approaches have been used for prediction of allergenicity. Zorzet et al. (2002) [5] used a supervised learning algorithm to predict allergenicity. The algorithm was trained on the data derived from sequence alignment matches using defined sets of amino acids. Approximately 81% predictions of the known allergens were found to be correct and 98% accuracy was achieved for prediction of sequences devoid of any allergenic potential. Goodman and Wise (2006) [6] reported the computational prediction of potential allergenicity for the proteins expressed in Golden rice 2. The bioinformatics approaches and a database of allergens and gliadins for prediction of potential allergenicity in six *Bacillus thuringiensis* insecticidal proteins, 50 randomly selected corn proteins and 3 common non allergenic food proteins, [7].

The epitopes present in the known allergens can also be searched for the presence of epitopes detected in the potential allergens. However, a comprehensive database of the authenticated epitopes is not available. In this context the algorithms such as ALLERGENSEARCH can be used for identification of short allergenic epitopes.

In conclusion it can be said that in silico approaches can complement the 'wet' approaches for determination of potential allergenicity. In this study we report the creation of Allergen domain online database which is useful for matching the domains present in the potential and known allergens.

REFERENCES

1. FAO/WHO, 2001. Evaluation of allergenicity of genetically modified foods derived from biotechnology. Rome, Italy.
2. Opiyo, S.O. Moriyama, E.N., 2007. Protein family classification with partial least squares. *J. Proteome Res.* 6, 846-853.
3. Randhawa, G.J., Singh, M. and Grover, M. (2011) Bioinformatic analysis for allergenicity assessment of *Bacillus thuringiensis* Cry proteins expressed in insect-resistant food crops. *Food and Chem. Toxic.* 49: 356-362
4. Codex Alimentarius Commission, 2003. Alinorm 03/34: Appendix III. Guideline for the conduct of food safety assessment of foods derived from recombinant DNA plants. Annex IV. Annex on the assessment of possible allergenicity, Rome, Italy.
5. Zorzet, A., Gustafsson, M., Hammerling, U., 2002. Prediction of food protein allergenicity: a bioinformatic learning systems approach. *In Silico Biol.* 2, 0048.
6. Goodman, R.E., Wise, J., 2006. Bioinformatics analysis of proteins in Golden rice2 to assess potential allergenic cross-reactivity. In: *Food Allergy Research & Resource Program, Univ. of Nebraska (Study no. BIO-02-2006).* pp. 1-24.
7. Hileman, R.E., Silvanovich, A., Goodman, R.E., Rice, E.A., Holleschak, G., Astwood, J.D., Hefle, S.L., 2002. Bioinformatic methods of allergenicity assessment using a comprehensive allergen database. *Int. Arch. Allergy Immunol.* 128, 280-291.

Fig,1 : Screenshot of allergen online domain database

The screenshot displays the 'Allergen Domain Online Database' website. The page is titled 'Allergen Domain Online Database' and includes a search bar labeled 'Search Domain'. The main text describes the database's role in assessing the allergenicity of GM crops using bioinformatics and in silico methods. It mentions the Codex Alimentarius guidelines and the FAO/WHO's 'weight of evidence' approach. A 3D ribbon diagram of a protein structure is shown on the right side of the page. The footer credits the developer as Mr. Vijay Kumar Sharma, Research Associate at NBPGR, New Delhi. The browser window shows the URL <http://localhost:49239/allergene/Home.aspx>.