

# The Necessity for Improvement of the Algorithms Used for *In Silico* Allergenicity Assessment of Novel Proteins

Najaf Allahyari Fard<sup>1</sup>, Zarrin Minuchehr<sup>1</sup>

<sup>1</sup>National Institute of Genetic Engineering and Biotechnology (NIGEB),  
Bioinformatics Group, 14965/161, Tehran, Iran  
allahyar@nigeb.ac.ir

**Abstract.** Prevalence of allergic diseases is increasing worldwide, particularly in low and middle income countries. Also upsurge in the prevalence of allergies is observed as countries become more urbanized. Allergy is a global public health problem and greater efforts should be made towards its prevention and optimal treatment. Prediction and *In Silico* allergenicity assessment is very important for safety evaluation of foods, Protein-based therapeutics, and other uses of recombinant proteins. Several bioinformatics approaches have been developed for evaluation the potential allergenicity of recombinant proteins. In this investigation, we study approaches and algorithms used for *In Silico* allergenicity evaluation of novel proteins. Based on our results there are two types of *In Silico*-based allergenicity prediction. Some allergen sequence databases such as SDAP, Allergenonline, Allermatch, ADFS, PSD, and AllerTool follow FAO/WHO guidelines and searches for primary structure of protein and sequence similarity. In this approach we achieve to high sensitivity (true positives/ (true positives + false negatives)) and many false positives and have low precision or positive predictive value (true positives/ (true positives + false positives)). Another approach includes identifying conserved allergenicity related motifs. Some allergen sequence databases such as AlgPred is a motif based server which combines four methods for motif search: Support Vector Machines (SVM), MEME/MAST, IgE epitopes and Allergen Representative Peptides (ARP). Deficiencies of the algorithms used for *In Silico* allergenicity evaluation of novel proteins are; 1) In both approaches allergenicity is assumed linearly and conformational epitopes are not considered. 2) Discovery of new allergen will be restricted by their lack of identity to known allergens. 3) 3D structure of proteins, the positions of IgE epitopes on surface of protein, and stability of IgE-Allergen complex that are important for molecular allergenicity mechanism are not considered. Thus current algorithms used for *In Silico* allergenicity assessment of novel proteins and FAO/WHO guidelines should be improved.

**Keywords:** Improvement, Algorithms, *In Silico*, Allergenicity