FARRP Common allergen database update

Goodman RE, Hefle SL, Taylor SL, Wise J. Food Allergy Research and Resource Program, University of Nebraska, Lincoln. USA

Abstract

Genetic engineering and modern processing methods make it possible to introduce novel proteins into foods and consumer products. Bioinformatic methods such as FASTA or BLAST may be efficiently used prior to marketing to reduce the probability of introducing a protein that is an allergen in another source, or one that could elicit an allergic cross-reaction in a sensitized individual. While the exact identity needed to induce cross-reactivity is not known, empirical evidence indicates that proteins that are less than 50% identical throughout their sequences are unlikely to serve as cross-reactive allergens. Bioinformatics approaches require the development of an allergen database for FASTA (or BLASTP) and other comparisons. The inclusiveness of the allergen database is crucial for reducing the risk of missing a potentially hazardous protein. However, many "allergens" in the literature are unproven and inclusion of many non-allergenic sequences could severely limit the potential benefits of new foods by mis-identifying safe proteins as allergens. Due to limited access to appropriate, fully-characterized allergic subject sera, and the uncertainties associated with *in vitro* IgE tests, false-positive bioinformatics results are not easily identified by serum screening. AllergenOnline with 1191 sequences in version 5.0 (updated October, 2004), was constructed by searching public sequence databases (NCBI-GenBank, SwissProt and others) as well as PubMed sources, primarily using the keyword *allergen*. The dataset was curated to remove duplicates and those representing obviously irrelevant proteins (e.g. IL-4, TLR, IgE). In order to improve the accuracy of this database for safety assessments, we have assembled a group of independent allergy experts including clinicians and immunologists/food chemists to develop scientifically defensible criteria for the selection of protein sequences that enter the database (including existing sequences). This panel will also review the list of entries. We are currently establishing criteria to help in determining those proteins that are not allergenic, or those that lack sufficient proof of identity or clinical reactivity based on published data. We are also developing criteria to grade the reliability and relevance of data that might be useful to differentiate matches to proven allergens (e.g. Ara h 2, Ovomucoid, peach LTP, Der p 2) from those of less certainty or potentially lower apparent potency. This database will be freely available to the public along with the criteria. The effort is being sponsored by seven companies (BASF, Bayer Cropscience, Dow AgroSciences, DuPont/Pioneer, Monsanto, Procter & Gamble, Syngenta).