



Published in final edited form as:

Regul Toxicol Pharmacol. 2009 August ; 54(3 Suppl): S11–S19. doi:10.1016/j.yrtph.2008.11.007.

Structural analysis of linear and conformational epitopes of allergens

Ovidiu Ivanciuc, Catherine H. Schein, Tzintzuni Garcia, Numan Oezguen, Surendra S. Negi, and Werner Braun*

Sealy Center for Structural Biology and Molecular Biophysics, Departments of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-0857, USA

Abstract

In many countries regulatory agencies have adopted safety guidelines, based on bioinformatics rules from the WHO/FAO and EFSA recommendations, to prevent potentially allergenic novel foods or agricultural products from reaching consumers. We created the Structural Database of Allergenic Proteins (SDAP, <http://fermi.utmb.edu/SDAP/>) to combine data that had previously been available only as flat files on Web pages or in the literature. SDAP was designed to be user friendly, to be of maximum use to regulatory agencies, clinicians, as well as to scientists interested in assessing the potential allergenic risk of a protein. We developed methods, unique to SDAP, to compare the physicochemical properties of discrete areas of allergenic proteins to known IgE epitopes. We developed a new similarity measure, the property distance (*PD*) value that can be used to detect related segments in allergens with clinical observed crossreactivity. We have now expanded this work to obtain experimental validation of the *PD* index as a quantitative predictor of IgE cross-reactivity, by designing peptide variants with predetermined *PD* scores relative to known IgE epitopes. In complementary work we show how sequence motifs characteristic of allergenic proteins in protein families can be used as fingerprints for allergenicity.

Keywords

Structural database of allergenic proteins (SDAP); Sequence and structural motifs; Linear and conformational epitopes

1. Introduction

It is well known that clinically important cross-reactivities among environmental triggers of allergy and asthma can be accounted for by proteins in those sources that have common molecular properties (Breiteneder and Ebner, 2000; Mari, 2001; Ferreira et al., 2004; Breiteneder and Mills, 2005; Jenkins et al., 2005). For example, the major allergenic proteins isolated from peanut (Burks et al., 1997; Shin et al., 1998; Rabjohn et al., 1999) have homologues in other foods that are known to elicit clinically significant responses in atopic individuals (Schein et al., 2005a), such as tree nuts (de Leon et al., 2003), soy (Eigenmann et al., 1996), and legumes (Lopez-Torrejon et al., 2003; Wensing et al., 2003). Pathogenesis response (PR) proteins of plants are another major group of protein families that are commonly found among allergens (Midoro-Horiuti et al., 2001). The cedar pollen allergen Jun a 3, classified as pathogenesis response protein group 5 (PR5) (Soman et al., 2000), was

*Corresponding author. Fax: +1 409 747 6000. E-mail address: E-mail: webraun@utmb.edu (W. Braun).

subsequently shown to be similar to allergenic proteins isolated from many different plants (Midoro-Horiuti et al., 2001; Elbez et al., 2002; Hoffmann-Sommergruber, 2002; Asensio et al., 2004). These included many food sources, including cherries, bell pepper, apple and tomato (Ivanciuc et al., 2003a). Other cedar pollen allergens (Midoro-Horiuti et al., 1999a,b, 2003, 2006; Czerwinski et al., 2005; Varshney et al., 2007) were shown to be similar to proteins from other pollen allergens including birch (Fedorov et al., 1997; Ferreira et al., 1998; Spangfort et al., 1999), and grass (Lalla et al., 1996; Schramm et al., 1997; Petersen et al., 1998; Flicker et al., 2000).

There are now several databases that contain sequences and information about allergenic proteins, as reviewed in several publications (Hileman et al., 2002; Brusica et al., 2003; Gendel, 2004; Gendel and Jenkins, 2006; Goodman, 2006; Schein et al., 2007). Most of these databases are simply lists of allergenic proteins or sources, with limited cross-indexing, including the IUIS (International Union of Immunological Societies) Website, (<http://www.allergen.org>), AllAllergy (<http://allallergy.net/>), or the Biotechnology Information for Food Safety Database (National Center for Food Safety and Technology, <http://www.iit.edu/~sgendel/fa.htm>). Cross-indexed databases, more useful for identifying potentially cross-reacting allergens and their natural sources, include Allergome (<http://www.allergome.org>), CSL (Central Science Laboratory, UK, <http://www.csl.gov.uk/allergen/index.htm>), Protall (<http://www.ifrn.bbsrc.ac.uk/protall/>), and the FARRP database (<http://allergenonline.com/asp/public/login.asp>). Here, we will present details of our Structural Database of Allergenic Proteins (SDAP, <http://fermi.utmb.edu/SDAP/>) (Ivanciuc et al., 2002, 2003b), described as the “most ambitious of the molecular databases” in a recent review (Gendel and Jenkins, 2006), which is unique in that it contains several bioinformatics search tools beyond standard FASTA to identify cross-reactive allergens. We will discuss and highlight those specific features that have been recently added to SDAP (Ivanciuc et al., 2008a; Oezguen et al., 2008) and that are of particular relevance for regulatory purposes (Schein et al., 2007).

Current guidelines recommend the use of standard sequence comparison methods, such as BLAST or FASTA, to determine whether a protein could cause reactions in allergic individuals (Goodman, 2006). We have implemented these tools in SDAP, and one can now conveniently test their ability to: first, discriminate allergenic proteins based on their identity to officially recognized allergens, and secondly, distinguish proteins that should not be allergenic. As we show below, we have found the first task to be easier than the second one. Using the sequence information of known allergens as archived in SDAP, we used a large-scale statistical analysis to test the bioinformatics guidelines proposed by WHO and EFSA committees (WHO, 2000, 2001, 2003; EFSA, 2004). We found that, in seeking to identify all proteins that could be allergens, strict adherence to these guidelines would suggest eliminating about a third of all known proteins from our environment! Conundrums abounded in our results: proteins known to cause anaphylaxis in sensitive individuals, such as the tropomyosins of shrimp and other crustaceans, have high sequence identity to mammalian homologues that are not allergens (Schein et al., 2007). Our results emphasized that the nature of allergenicity was local, and that to identify the true allergenic potential of a protein one had to catalogue discrete areas of allergenic proteins that would bind IgE. Thus we included in SDAP a cross-referenced list of sequences known to bind IgE from patient sera, coupled with tools designed to compare their sequences to those of other allergens.

As the current bioinformatics guidelines for allergenicity, based on simple sequence comparisons, are far from optimal (van Ree et al., 2006; Schein et al., 2007; Goodman, 2008), we present alternative classification methods that use analysis of local sequence and structure to identify common features of allergenic proteins that distinguish them from related, non-allergenic proteins. Recent efforts to include structural information on the allergens in

predicting cross-reactivity (Aalberse, 2007; Chapman et al., 2007; Bonds et al., 2008; Oezguen et al., 2008) require classification into discrete PFAM classes (Ivanciuc et al., 2008a; Radauer et al., 2008). In addition, we developed and validated a *PD* (“physicochemical property distance”) scale (Ivanciuc et al., 2002, 2003b) expressly to identify, with statistical significance, areas of allergens catalogued in SDAP that are similar to known IgE binding sequences (Schein et al., 2007; Ivanciuc et al., 2008b). The compiled information in SDAP, in addition to the sequences and epitopes for all allergens listed in the IUIS Website, from published literature and from other databases, also includes substantial 3D-structural information. Classification of all the allergens in SDAP according to their protein family (Pfam) also allowed us to characterize sequence motifs, which can be used as fingerprints for allergenicity (Ivanciuc et al., 2008a). Those sequence motifs are publicly available on the MotifMate web server (<http://born.utmb.edu/motifmate/>). We also explored the 3D-structural characteristics of conformational epitopes that can be of importance for more refined bioinformatics rules in the future (Oezguen et al., 2008).

1.1. Extracting information from the catalogue of allergens in the SDAP database

SDAP has integrated search tools to allow a user to rapidly compare the molecular properties of allergenic proteins and their epitopes (Ivanciuc et al., 2002). SDAP was developed for basic research to determine common molecular characteristics of group of allergens, and to provide regulatory agencies, food scientists and biomedical researchers software support to determine if a novel protein has allergenic potential (Ivanciuc et al., 2003a). No special training is needed to access the data, and the tools are implemented in a user friendly fashion. Software tools integrated in SDAP include the FAO/WHO bioinformatics rules, standard BLAST (Schaffer et al., 2001) and FASTA (Pearson, 1994) search methods, ExPaSy (Schneider et al., 2004), PIR (Barker et al., 1999) and PRO-SITE (Hulo et al., 2006). The special tools of SDAP, such as the *PD* scale (Ivanciuc et al., 2002, 2003b), were developed to compare short sequences to one another in a mathematically rigorous, unbiased fashion (as opposed to using simple sequence comparisons, “by eye”, or applying limited rules with respect to identity or homology).

SDAP is also integrated with other bioinformatics servers, allowing the user to investigate structural similarity and neighbors using SCOP (Structural Classification Of Proteins) (Conte et al., 2000), TOPS (TOPOlogical representation of Protein Structure) (Gilbert et al., 1999), CATH (Class, Architecture, Topology and Homologous superfamily) (Pearl et al., 2001), CE (Combinatorial Extension of the optimal path) (Shindyalov and Bourne, 1998), FSSP (Fold Classification based on Structure–Structure alignment of Proteins) (Holm and Sander, 1996), and VAST (Vector Alignment Search Tool) (Gibrat et al., 1996).

The information content in SDAP for a given allergen is illustrated for the Ole e 8 protein from olive trees (Fig. 1). This descriptive page is shown after a user selects the allergen of interest. The page contains a summary of all the data archived in SDAP for the selected allergen, including the official name (according to the IUIS Website listing, <http://allergen.org/>), scientific and common name for the species, general source of the allergens, allergen type; species; systematic name; brief description; sequence accession numbers from SwissProt, PIR, NCBI and, where available, the PDB file name for a structure. All of this information is also cross-referenced to other data sources, which can be directly accessed by clicking on the appropriate links.

1.2. Search methods in SDAP

Several methods implemented in the SDAP web server are designed for regulatory purposes. The most widely used method to determine potential allergenicity of a novel protein is to do a global sequence search method to other allergens by FASTA (Pearson, 1990). FASTA can be

run automatically from any sequence file in SDAP by a mouse-click, and outputs a table that lists all similar allergens in SDAP according to their “*E*-value”, or expectation value, to the target, which indicates the statistical significance of the hit. The *E*-value is a measure of how many matches with the same sequence similarity one would expect to occur randomly in a database of a given size. Thus a low *E*-value (e.g. less than 10^{-6}) indicates a high significance of the sequence match.

The FAO/WHO reports (Bindsley-Jensen et al., 2003; WHO, 2003) proposed that cross-reactivity between a query protein and a known allergen has to be considered when there is (a) more than 35% identity in the amino acid sequence of the query protein, using a window of 80 amino acids and a suitable gap penalty, or: (b) identity of six contiguous amino acids of the query protein in a known allergen. To carry-out a search based on these criteria, the SDAP user only needs to cut and paste a query protein sequence in the appropriate window at the SDAP Website (Fig. 2). The output lists all similar proteins in SDAP (i.e. those that satisfy the FAO/WHO cross-reactivity conditions). Several variations of the search can be performed by altering the parameters; e.g. a full length FASTA search in SDAP or searching for larger segments of contiguous identical residues. The output is a summary table listing allergens that have an *E*-score alignment with the query protein lower than the user-set maximum. The output also contains the individual pairwise alignments and full sequence identities. The user can examine each pairwise alignment and use as a guide to estimate the allergenic potential of the query sequence.

1.3. Statistical validation of the WHO/FAO rules

The first questions about these rules are: how many proteins will be incorrectly determined to be allergenic, and more importantly, how many allergens will be missed? To validate the bioinformatics guidelines of the FAO/WHO committee, we used all SDAP entries as positive controls, and we filtered the SwissProt database to generate a set of non-allergenic proteins (negative control). For the negative control set, we removed all SDAP entries in the SwissProt database and then used keyword filters to remove all SwissProt records that (a) contain an allergen-related keyword (allergen, allergy, lipid transfer protein, profilin, lipocalin, pectate lyase, tropomyosin, melittin, thaumatin, seed storage protein), (b) have a sequence shorter than 80 amino acids, or (c) belong to Inter-Pro, Pfam, or Prosite allergen-related classes. For every SDAP protein we recorded the best match among all windows of 80 amino acids. A protein is classified as an allergen if the sequence identity to an allergen is higher than a given threshold in a window of 80 amino acids (Fig. 3). A comparison between the fraction of positive controls (SDAP allergens, blue line) and the fraction of negative controls (set of non-allergenic proteins, red line) suggests that a good threshold for the sequence identity should be between 35% and 45%. We found that the threshold of 35% for sequence identity is a good estimate for separating allergens from non-allergens, but with 6.6% of non-allergenic proteins classified as allergenic there is still a relatively high number of false positives.

The sensitivity of criterion 1 was evaluated by comparing each SDAP allergen with the remaining SDAP sequences (blue line). For a threshold sequence identity of 35%, the test correctly identifies 92.29% allergens. Increasing this threshold to 45% decreases the fraction of SDAP allergens identified to 90.45%. Decreasing the threshold sequence identity to 15% will identify 99.10% of known allergens, but at this level 78.25% from SwissProt (95725 sequences) would also be considered allergenic! Thus, sequence identify alone cannot be used to absolutely identify allergens. The results from Fig. 3 indicate that while this bioinformatics test is able to filter non-allergenic proteins when the sequence identity is between 35% and 45%, the overall sequence identity is not the only determinant for allergenicity. Additional quantitative descriptors need to be developed for computational predictions.

1.4. Sequence similarity ranking in SDAP: the “property-distance” (PD) scale

The FASTA search in SDAP is a rapid way to determine the overall similarity of large proteins. However, FASTA was not designed to compare short sequences, such as the linear IgE epitopes that have been identified by peptide mapping for many allergens (Jarvinen et al., 2001; Elsayed et al., 2004; Shreffler et al., 2004; Schein et al., 2005a). Two different tools were incorporated in SDAP to look for short sequences in other known allergens, an “exact search”, that finds short sequences identical to that of a known epitope, and a second tool, to determine sequences that are close to the IgE epitope in the *PD* “property-distance space” (Ivanciuc et al., 2002, 2003b). The *PD* tool determines similar sequences in other allergen entries in SDAP that have similar overall physicochemical properties. Peptides with identical sequences have a *PD* value of 0, and peptides with conservative substitutions of a few amino acids have a small *PD* value, typically in the range of 0–3. Peptides with a recognizable similarity in their physicochemical properties generally have *PD* values lower than 10, while unrelated peptides have *PD* values that are much higher.

The *PD* score is based on the amino acid descriptors E_1 – E_5 that were determined by the multidimensional scaling of 237 physico-chemical properties of amino acids (Venkatarajan and Braun, 2001). Using the amino physicochemical descriptors E_1 – E_5 , the properties of the 20 naturally occurring amino acids can be numerically summarized as five values. These five dimensions define a physicochemical property space for all amino acids, with each axis representing a distinct feature. For example, the first three *E* descriptors correlate with the amino acid’s hydrophobicity, size, and polarity, respectively. Each amino acid is represented as a point in the five-dimensional space E_1 – E_5 , and the similarity between two amino acids is inversely correlated to the distance between the two points representing the two amino acids. The *PD* sequence similarity score for two sequences *A* and *B* each containing *N* amino acids is (Ivanciuc et al., 2002, 2003b):

$$PD(A, B) = \frac{1}{N} \sum_{i=1}^N \left[\sum_{j=1}^5 \lambda_j (E_j(A_i) - E_j(B_i))^2 \right]^{1/2} \quad (1)$$

where λ_j is the eigenvalue of the *j*-th *E* component, $E_j(A_i)$ is the E_j value for the amino acid in the *i*-th position from sequence *A*, and $E_j(B_i)$ is the E_j value for the amino acid in the *i*-th position from sequence *B*.

Table 1 illustrates the usefulness of using the *PD* value to identify related potential epitopes and potentially cross-reactive allergenic proteins for the IgE epitope VQGKEKEP of Par j 1. The *PD* search identifies a fragment from the related allergen Par j 2 (VKGEEKEP; Table 1) as the most similar region to the IgE epitope of Par j 1. More distant similarities are identified in a number of SDAP allergens. We should at this point emphasize that the *PD* search is a computational way to define the sequence relationship between known IgE epitopes and other sequences in allergenic proteins. Our initial tests indicate that *PD* is a reliable index to quantify local similarities in known allergens.

To obtain experimental validation of the *PD* index as a quantitative predictor of IgE cross-reactivity we designed peptide variants with predetermined *PD* scores relative to three linear IgE epitopes of Jun a 1 (Midoro-Horiuti et al., 2003, 2006). The peptides synthesized on a derivatized cellulose membrane were probed with sera from patients allergic to Jun a 1, and the experimental data were interpreted with a *PD* classification method, giving a percentage of correct predictions up to 80% (Ivanciuc et al., 2008b). Peptides similar to a Jun a 1 epitope ($PD < 6$) were more likely to bind IgE from the sera than were those with *PD* values larger than 6. Control sequences, with *PD* values between 18 and 20 to all the three epitopes, did not

bind patient IgE, thus validating our procedure for identifying negative control peptides. These results demonstrate that the *PD* index may identify peptides that have a high probability of cross-reacting with IgE from allergic patients.

1.5. Grouping allergenic proteins according to major Pfam families

Classification of allergens into functional groups of proteins can indicate important relationships and has the additional advantage that structural and sequence groupings allow one to identify significant similarities in proteins with diverse origins. We annotated all allergens in SDAP according to their Pfam classification (Ivanciuc et al., 2008a). Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) is a list of multiple sequence alignments of related protein domains, classified in two ways. The Pfam-A database lists protein families that are grouped by their common function as well as sequence, using expert knowledge and experimental data. Pfam-B is computer-generated and contains alignments of proteins sequences selected based on a minimum level of sequence identity, regardless of their protein function. Most SDAP entries have now been classified to families from the Pfam-A database. Easy access to this Pfam classification for any allergen can be accessed from the “List SDAP” menu item.

Allergens from the same Pfam class exhibit a high structural similarity, as it is shown in Fig. 4 for three pairs of allergens: Act c 1 (kiwi, PDB 2ACT) and Car p 1 (papaya, PDB 1KHQ) from the family PF00112, Papain family cysteine protease; Phl p 5 (timothy, PDB 1L3P) and Phl p 6 (PDB 1NLX) from the family PF01620, Ribonuclease (pollen allergen); Der f 2 (American house dust mite, PDB 1XWV) and Der p 2 (European house dust mite, PDB 1KTJ), from the family PF02221, ML domain. We found that allergens populate only a small subset of all known Pfam families, as all allergenic proteins in SDAP could be grouped to only 130 (of 9318 total) Pfams, and only 31 families contain more than 4 allergens, which is consistent with results obtained by others (Radauer et al., 2008). The limited number of Pfam families suggests new criteria to estimate the potential risk of allergenic recombinant protein products. For example, if a novel protein product belongs to a Pfam class different from all listed Pfam classes as found in SDAP, it should be considered to have little allergenic potential.

1.6. Motif-based methods for allergenicity prediction

Alternatively, one can define discrete areas of residue conservation, “motifs”, in related allergenic proteins of known clinical cross-reactivity, as possible areas for IgE binding. Several groups have defined conserved sequences in groups of allergens (Mills et al., 2002; Brusic and Petrovsky, 2003; Stadler and Stadler, 2003; Li et al., 2004; Marti et al., 2007). Unlike motifs defined by others, which can be quite long (to the point that they be more properly called protein domains), we define areas more likely to be discrete IgE epitopes, with a normal length is between 6 and 15 amino acids. In our work, we look for areas where the side chains show conserved physicochemical properties (PCPs), such as hydrophobicity, size or alpha-helical propensity, rather than strict identity. The underlying assumption is that for a group of cross-reactive allergenic proteins, the IgE epitopes areas have similar binding affinities for the same antibodies, and have thus common physico chemical properties in the antibody binding sites.

Our method begins by aligning the sequences of known allergens that are related to one another, such as those in the tropomyosin or vicilin family. The PCPMer suite (available at <http://landau.utmb.edu:8080/WebPCPMer/HomePage/index.html>) finds sequence motifs in protein families by identifying regions with highly conserved physicochemical properties. These “PCP-motifs” are determined by conservation of the five quantitative property vectors E_1 – E_5 which summarize many different physicochemical properties of the side chains of the amino acids, including size, hydrophobicity, and tendency to form helical or strand secondary structures (Venkatarajan and Braun, 2001; Venkatarajan et al., 2003). Sequence motifs are

contiguous segments of high relative entropy values for at least one of the five descriptors. Alternatively, the program allows the user to set thresholds of relative entropy, gap cutoff and minimum motif length to balance the specificity and sensitivity of motifs. Each motif identified by PCPmer is quantitatively expressed as a profile, in this case (for a motif of length N), a series of $N \times 5$ matrices consisting of the average values, standard deviations and the relative entropies of the descriptors E_1 – E_5 at each position (column in the multiple sequence alignment) in the motif. This profile can be used to search for similar sequences in protein databases. Details of the algorithm for motif generation are described in our previous publications (Venkatarajan et al., 2003; Schein et al., 2005b).

As an illustration for the generation of sequence motifs we show the (truncated) multiple sequence alignment for the walnut allergen Jug r 1 with other allergens in the same Pfam classification (Fig. 5A) and the corresponding output of PCPmer in Fig. 5B. The sequence of the first protein in the alignment is given as reference for those columns where the relative entropy values exceed the value given in the first column of Fig. 5B. Two motifs, CQYYLR and CCQQLS, are identified as local maxims of the relative entropy values, and are regions of high conservation of physicochemical properties. Our PCP motifs do not require that residues within a motif are identical among all sequences, just that the overall pattern of property presentation be similar.

1.7. Combining sequence and structural information to improve prediction

Motifs can also be mapped onto the 3D-structure of a protein to identify epitopes and conserved functional areas (Schein et al., 2005b). Combining sequence analysis with structural representations can answer many questions about the nature of the IgE epitopes of allergens. For example, why do some individuals show cross-reactivity to homologous proteins in peanuts and tree nuts, while others react to one or another of the homologous proteins (Teuber and Beyer, 2004)? While single amino acid differences may be quite important in individual reactivity, a 3D view of the identified IgE binding sites can provide missing information about the possible relationships between structure and sequence. If IgE binding sequences of related proteins have similar properties, the proposed methods that combine PD values with structural details will have higher predictive ability, if properly calibrated. Thus we are building up a library of models, based on homology of allergens to proteins of known structure, to determine clusters of residues that are conserved on the surface of allergens.

Once similar sequences have been identified by PD values, the structural information in SDAP can be used to understand which parts of an allergen sequence are likely to be surface exposed, and thus able to form an IgE binding surface. In order to investigate the structural features of allergens we computed reliable models for more than 80% of allergens in SDAP for which the experimental structure is unknown (Oezguen et al., 2008). We initially attempted to generate 3D homology models for 645 allergens in SDAP for which no experimental structure or close homolog is deposited in the Protein Databank. Each model of our automatic procedure was evaluated critically by three quality criteria, namely: (1) negative overall conformational energy after FANTOM minimization, which indicates favorable local packing of the side chains; (2) an RMSD to the template for the aligned regions less than 1.8 Å; and (3) not more than 5% of the ϕ/ψ dihedral angles situated in the disallowed region of a Ramachandran plot. Overall, 433 allergen sequences passed these criteria and gave reliable 3D homology models that are currently deposited in SDAP and are available for viewing or for download. These allergen models can be used to determine areas of local structure that correlate with allergenicity. For example, using our models, linear IgE epitopes taken from SDAP are mapped onto the surface of the pollen allergen Par j 1 (Fig. 6): epitope 1, VQGKEKEP, red; epitope 2, SKGCCSGAKRLD, green; epitope 3, KTGPRQRV, gold; epitope 4, PKHCGIVD, blue. The surface mapping of linear epitopes on the 3D models of allergens may be used to identify buried

residues and surface accessible residues, thus highlighting the amino acids that may bind to the IgE.

Phage display technology is an alternative approach to characterize conformational epitopes of proteins. It identifies a discontinuous group of amino acids on the protein surface by binding to a monoclonal antibody (Smith and Petrenko, 1997). Therefore, to locate the interaction site on the protein surface mimicked by the epitope is not possible using sequence analysis alone. We developed a fully automated method EpiSearch (available at <http://curie.utmb.edu/episearch.html>), that locates the antibody binding site on the antigen surface using the peptide sequences obtained from phage display. The method is a further development of our approach to predict interface residues in a monomeric protein (Negi et al., 2006, 2007; Negi and Braun, 2007).

2. Conclusions

Bioinformatics analysis of the properties of allergens has progressed greatly in the last few years. As we have shown, SDAP has reliable tools that go beyond the initial guidelines for determining the potential allergenicity of new food products for regulatory purposes. SDAP contains now a broad array of bioinformatics and computational tools that: (1) can evaluate the overall sequence similarity to a known allergen based on FASTA alignments, (2) evaluate the WHO/FAO rules, (3) find regions identical with known IgE epitopes, (4) identify regions similar with known IgE epitopes and rank them with the *PD* score, (5) use 3D homology models to identify the amino acids that are important in IgE binding.

We regard our studies important in providing a solid scientific foundation in the general discussion on the potential risk of genetically modified (GM) foods. The statistical results and the novel bioinformatics tools can help regulatory agencies in the US and other countries that grow GM plants to find more specific bioinformatics guidelines for these crops. Since food allergies can result in fatal reactions, the allergenic potential of genetically-engineered food products needs to be carefully assessed prior to their entry into the market. There is a vital need for faster and reliable methods to evaluate the potential allergenicity of proteins that have not previously been part of the food supply. Our novel approaches can reduce some uncertainty for those crops that may be potentially allergenic for some sensitive sub-population.

Acknowledgments

This work was supported by a contract from the US Food and Drug Administration (HHSF223200710011) and grants from the National Institute of Health (R01 AI 064913), and the US Environmental Protection Agency under a STAR Research Assistance Agreement (No. RD 833137).

References

- Aalberse RC. Assessment of allergen cross-reactivity. *Clinical and Molecular Allergy* 2007;5:2. [PubMed: 17291349]
- Asensio T, Crespo JF, Sanchez-Monge R, Lopez-Torrejon G, Somoza ML, Rodriguez J, Salcedo G. Novel plant pathogenesis-related protein family involved in food allergy. *Journal of Allergy and Clinical Immunology* 2004;114:896–899. [PubMed: 15480331]
- Barker WC, Garavelli JS, McGarvey PB, Marzec CR, Orcutt BC, Srinivasarao GY, Yeh LS, Ledley RS, Mewes HW, Pfeiffer F, Tsugita A, Wu C. The PIR-International Protein Sequence Database. *Nucleic Acids Research* 1999;27:39–43. [PubMed: 9847137]
- Bindsley-Jensen C, Sten E, Earl LK, Crevel RWR, Bindslev-Jensen U, Hansen TK, Skov PS, Poulsen LK. Assessment of the potential allergenicity of ice structuring protein type III HPLC 12 using the FAO/WHO 2001 decision tree for novel foods. *Food and Chemical Toxicology* 2003;41:81–87. [PubMed: 12453731]

- Bonds RS, Midoro-Horiuti T, Goldblum R. A structural basis for food allergy: the role of cross-reactivity. *Current Opinion in Allergy and Clinical Immunology* 2008;8:82–86. [PubMed: 18188023]
- Breiteneder H, Mills ENC. Molecular properties of food allergens. *The Journal of Allergy and Clinical Immunology* 2005;115:14–23. [PubMed: 15637541]
- Breiteneder H, Ebner C. Molecular and biochemical classification of plant-derived food allergens. *Journal of Allergy and Clinical Immunology* 2000;106:27–36. [PubMed: 10887301]
- Brusic V, Millot M, Petrovsky N, Gendel SM, Gigonzac O, Stelman SJ. Allergen databases. *Allergy* 2003;58:1093–1100. [PubMed: 14616118]
- Brusic V, Petrovsky N. Bioinformatics for characterisation of allergens, allergenicity and allergic crossreactivity. *Trends in Immunology* 2003;24:225–228. [PubMed: 12738409]
- Burks AW, Shin D, Cockrell G, Stanley JS, Helm RM, Bannon GA. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *European Journal of Biochemistry* 1997;245:334–339. [PubMed: 9151961]
- Chapman MD, Pomés A, Breiteneder H, Ferreira F. Nomenclature and structural biology of allergens. *Journal of Allergy and Clinical Immunology* 2007;119:414–420. [PubMed: 17166572]
- Conte LL, Ailey B, Hubbard TJP, Brenner SE, Murzin AG, Chothia C. SCOP: A Structural Classification of Proteins Database. *Nucleic Acids Research* 2000;28:257–259. [PubMed: 10592240]
- Czerwinski EW, Midoro-Horiuti T, White MA, Brooks EG, Goldblum RM. Crystal structure of Jun a 1, the major cedar pollen allergen from *Juniperus ashei*, reveals a parallel beta-helical core. *The Journal of Biological Chemistry* 2005;280:3740–3746. [PubMed: 15539389]
- de Leon MP, Glaspole IN, Drew AC, Rolland JM, O’Hehir RE, Suphioglu C. Immunological analysis of allergenic cross-reactivity between peanut and tree nuts. *Clinical and Experimental Allergy* 2003;33:1273–1280. [PubMed: 12956750]
- EFSA. Guidance Document of the GMO Panel for the Risk Assessment of Genetically Modified Plants and Derived Food and Feed. European Food Safety Authority; 2004. Available from: <http://www.efsa.eu.int/science/gmo/gmo_guidance/660_en.html>
- Eigenmann PA, Burks AW, Bannon GA, Sampson HA. Identification of unique peanut and soy allergens in sera adsorbed with cross-reacting antibodies. *The Journal of Allergy and Clinical Immunology* 1996;98:969–978. [PubMed: 8939161]
- Elbez M, Kevers C, Hamdi S, Rideau M, Petit-Paly G. The plant pathogenesis-related PR-10 proteins. *Acta Botanica Gallica* 2002;149:415–444.
- Elsayed S, Hill DJ, Do TV. Evaluation of the allergenicity and antigenicity of bovine-milk alpha s1-casein using extensively purified synthetic peptides. *Scandinavian Journal of Immunology* 2004;60:486–493. [PubMed: 15541041]
- Fedorov AA, Ball T, Mahoney NM, Valenta R, Almo SC. The molecular basis for allergen cross-reactivity: crystal structure and IgE epitope mapping of birch pollen profilin. *Structure* 1997;5:33–45. [PubMed: 9016715]
- Ferreira F, Ebner C, Kramer B, Casari G, Briza P, Grimm R, Jahn-Schmid B, Breiteneder H, Kraft D, Breitenbach M, Rheinberger HJ, Scheiner O. Modulation of IgE reactivity of allergens by site-directed mutagenesis: potential use of hypoallergenic variants for immunotherapy. *FASEB J* 1998;12:231–242. [PubMed: 9472988]
- Ferreira F, Hawranek T, Gruber P, Wopfner N, Mari A. Allergic cross-reactivity: from gene to the clinic. *Allergy* 2004;59:243–267. [PubMed: 14982506]
- Flicker S, Vrtala S, Steinberger P, Vangelista L, Bufe A, Petersen A, Ghannadan M, Sperr WR, Valent P, Norderhaug L, Bohle B, Stockinger H, Suphioglu C, Ong EK, Kraft D, Valenta R. A human monoclonal IgE antibody defines a highly allergenic fragment of the major timothy grass pollen allergen, Phl p 5: molecular, immunological, and structural characterization of the epitope-containing domain. *Journal of Immunology* 2000;165:3849–3859.
- Gendel SM. Bioinformatics and food allergens. *Journal of AOAC International* 2004;87:1417–1422. [PubMed: 15675454]
- Gendel SM, Jenkins JA. Allergen sequence databases. *Molecular Nutrition & Food Research* 2006;50:633–637. [PubMed: 16764016]
- Gibrat JF, Madej T, Bryant SH. Surprising similarities in structure comparison. *Current Opinion in Structural Biology* 1996;6:377–385. [PubMed: 8804824]

- Gilbert D, Westhead D, Nagano N, Thornton J. Motif-based searching in TOPS protein topology databases. *Bioinformatics* 1999;15:317–326. [PubMed: 10320400]
- Goodman RE. Practical and predictive bioinformatics methods for the identification of potentially cross-reactive protein matches. *Molecular Nutrition & Food Research* 2006;50:655–660. [PubMed: 16810734]
- Goodman RE. Performing IgE serum testing due to bioinformatics matches in the allergenicity assessment of GM crops. *Food Chem Toxicol* 2008;46(Suppl 10):S24–S34. [PubMed: 18715545]
- Hileman R, Silvanovich A, Goodman R, Rice E, Holleschak G, Astwood J, Hefle S. Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *International Archives of Allergy and Immunology* 2002;128:280–291. [PubMed: 12218366]
- Hoffmann-Sommergruber K. Pathogenesis-related (PR)-proteins identified as allergens. *Biochemical Society Transactions* 2002;30:930–935. [PubMed: 12440949]
- Holm L, Sander C. Mapping the protein universe. *Science* 1996;273:595–602. [PubMed: 8662544]
- Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ. The PROSITE database. *Nucleic Acids Research* 2006;34:D227–D230. [PubMed: 16381852]
- Ivanciuc O, Garcia T, Torres M, Schein CH, Braun W. Characteristic motifs for families of allergenic proteins. *Molecular Immunology*. 2008a;10.1016/j.molimm.2008.07.034
- Ivanciuc O, Mathura V, Midoro-Horiuti T, Braun W, Goldblum RM, Schein CH. Detecting potential IgE-reactive sites on food proteins using a sequence and structure database, SDAP-food. *Journal of Agricultural and Food Chemistry* 2003a;51:4830–4837. [PubMed: 14705920]
- Ivanciuc O, Midoro-Horiuti T, Schein CH, Xie L, Hillman GR, Goldblum RM, Braun W. The property distance index PD predicts peptides that cross-react with IgE antibodies. *Molecular Immunology*. 2008b;10.1016/j.molimm.2008.09.004
- Ivanciuc O, Schein CH, Braun W. Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics* 2002;18:1358–1364. [PubMed: 12376380]
- Ivanciuc O, Schein CH, Braun W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Research* 2003b;31:359–362. [PubMed: 12520022]
- Jarvinen KM, Chatchatee P, Bardina L, Beyer K, Sampson HA. IgE and IgG binding epitopes on alpha-lactalbumin and beta-lactoglobulin in cow's milk allergy. *International Archives of Allergy and Immunology* 2001;126:111–118. [PubMed: 11729348]
- Jenkins J, Griffiths-Jones S, Shewry P, Breiteneder H, Mills ENC. Structural Relatedness of plant food allergens with specific reference to cross-reactive allergens: an *in silico* analysis. *Journal of Allergy and Clinical Immunology* 2005;115:163–170. [PubMed: 15637564]
- Lalla C, Tamborini E, Longhi R, Tresoldi E, Manoni M, Siccardi A, Arosio P, Sidoli A. Human recombinant antibody fragments specific for a rye-grass pollen allergen: characterization and potential applications. *Molecular Immunology* 1996;33:1049–1058. [PubMed: 9010244]
- Li KB, Issac P, Krishnan A. Predicting allergenic proteins using wavelet transform. *Bioinformatics* 2004;20:2572–2578. [PubMed: 15117757]
- Lopez-Torrejon G, Salcedo G, Martin-Esteban M, Diaz-Perales A, Pascual CY, Sanchez-Monge R. Len c 1, a major allergen and vicilin from lentil seeds: protein isolation and cDNA cloning. *The Journal of Allergy and Clinical Immunology* 2003;112:1208–1215. [PubMed: 14657885]
- Mari A. Multiple pollen sensitization: a molecular approach to the diagnosis. *International Archives of Allergy and Immunology* 2001;125:57–65. [PubMed: 11385289]
- Marti P, Truffer R, Stadler MB, Keller-Gautschi E, Cramer R, Mari A, Schmid-Grendelmeier P, Miescher SM, Stadler BM, Vogel M. Allergen motifs and the prediction of allergenicity. *Immunology Letters* 2007;109:47–55. [PubMed: 17303251]
- Midoro-Horiuti T, Brooks EG, Goldblum RM. Pathogenesis-related proteins of plants as allergens. *Annals of Allergy, Asthma and Immunology* 2001;87:261–271.
- Midoro-Horiuti T, Goldblum RM, Kurosky A, Goetz DW, Brooks EG. Isolation and characterization of the mountain cedar (*Juniperus ashei*) pollen major allergen, Jun a 1. *The Journal of Allergy and Clinical Immunology* 1999a;104:608–612. [PubMed: 10482835]
- Midoro-Horiuti T, Goldblum RM, Kurosky A, Wood TG, Schein CH, Brooks EG. Molecular cloning of the mountain cedar (*Juniperus ashei*) pollen major allergen, Jun a 1. *The Journal of Allergy and Clinical Immunology* 1999b;104:613–617. [PubMed: 10482836]

- Midoro-Horiuti T, Mathura V, Schein CH, Braun W, Yu S, Watanabe M, Lee JC, Brooks EG, Goldblum RM. Major linear IgE epitopes of mountain cedar pollen allergen Jun a 1 map to the pectate lyase catalytic site. *Molecular Immunology* 2003;40:555–562. [PubMed: 14563374]
- Midoro-Horiuti T, Schein CH, Mathura V, Braun W, Czerwinski EW, Togawa A, Kondo Y, Oka T, Watanabe M, Goldblum RM. Structural basis for epitope sharing between group 1 allergens of cedar pollen. *Molecular Immunology* 2006;43:509–518. [PubMed: 15975657]
- Mills EN, Jenkins J, Marigheto N, Belton PS, Gunning AP, Morris VJ. Allergens of the cupin superfamily. *Biochemical Society Transactions* 2002;30:925–929. [PubMed: 12440948]
- Negi SS, Braun W. Statistical analysis of physical–chemical properties and prediction of protein–protein interfaces. *Journal of Molecular Modeling* 2007;13:1157–1167. [PubMed: 17828612]
- Negi SS, Kolokoltsov AA, Schein CH, Davey RA, Braun W. Determining functionally important amino acid residues of the E1 protein of Venezuelan equine encephalitis virus. *Journal of Molecular Modeling* 2006;12:921–929. [PubMed: 16607494]
- Negi SS, Schein CH, Oezguen N, Power TD, Braun W. InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics* 2007;23:3397–3399. [PubMed: 17933856]
- Oezguen N, Zhou B, Negi SS, Ivanciuc O, Schein CH, Labesse G, Braun W. Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes. *Molecular Immunology* 2008;45:3740–3747. [PubMed: 18621419]
- Pearl FMG, Martin N, Bray JE, Buchan DWA, Harrison AP, Lee D, Reeves GA, Shepherd AJ, Sillitoe I, Todd AE, Thornton JM, Orengo CA. A rapid classification protocol for the CATH Domain Database to support structural genomics. *Nucleic Acids Research* 2001;29:223–227. [PubMed: 11125098]
- Pearson W. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology* 1990;183:63–98. [PubMed: 2156132]
- Pearson WR. Using the FASTA program to search protein and DNA sequence databases. *Methods in Molecular Biology* 1994;25:365–389. [PubMed: 8004177]
- Petersen A, Schramm G, Schlaak M, Becker WM. Post-translational modifications influence IgE reactivity. *Clinical and Experimental Allergy* 1998;28:315–321. [PubMed: 9543081]
- Rabjohn P, Burks AW, Sampson HA, Bannon GA. Mutational analysis of the IgE-binding epitopes of the peanut allergen, Ara h 3: a member of the glycinin family of seed-storage proteins. *Journal of Allergy and Clinical Immunology* 1999;103:S101.
- Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *The Journal of Allergy and Clinical Immunology* 2008;121:847–852. e847. [PubMed: 18395549]
- Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Research* 2001;29:2994–3005. [PubMed: 11452024]
- Schein CH, Ivanciuc O, Braun W. Common physical–chemical properties correlate with similar structure of the IgE epitopes of peanut allergens. *Journal of Agricultural and Food Chemistry* 2005a;53:8752–8759. [PubMed: 16248581]
- Schein CH, Ivanciuc O, Braun W. Bioinformatics approaches to classifying allergens and predicting cross-reactivity. *Immunology and Allergy Clinics of North America* 2007;27:1–27. [PubMed: 17276876]
- Schein CH, Zhou B, Braun W. Stereophysicochemical variability plots highlight conserved antigenic areas in Flaviviruses. *Virology Journal* 2005b;2:40. [PubMed: 15845145]
- Schneider M, Tognolli M, Bairoch A. The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools. *Plant Physiology and Biochemistry* 2004;42:1013–1021. [PubMed: 15707838]
- Schramm G, Bufe A, Petersen A, Haas H, Schlaak M, Becker WM. Mapping of IgE-binding epitopes on the recombinant major group I allergen of velvet grass pollen, rHol 1. *Journal of Allergy and Clinical Immunology* 1997;99:781–787. [PubMed: 9215246]
- Shin DS, Compadre CM, Maleki SJ, Kopper RA, Sampson H, Huang SK, Burks AW, Bannon GA. Biochemical and structural analysis of the IgE binding sites on Ara h1, an abundant and highly

- allergenic peanut protein. *Journal of Biological Chemistry* 1998;273:13753–13759. [PubMed: 9593717]
- Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering* 1998;11:739–747. [PubMed: 9796821]
- Shreffler WG, Beyer K, Chu TH, Burks AW, Sampson HA. Microarray immunoassay: association of clinical history, in vitro IgE function, and heterogeneity of allergenic peanut epitopes. *Journal of Allergy and Clinical Immunology* 2004;113:776–782. [PubMed: 15100687]
- Smith PG, Petrenko AV. Phage display. *Chemical Reviews* 1997;97:391–410. [PubMed: 11848876]
- Soman KV, Midoro-Horiuti T, Ferreon JC, Goldblum RM, Brooks EG, Kurosky A, Braun W, Schein CH. Homology modeling and characterization of IgE binding epitopes of mountain cedar allergen Jun a 3. *Biophysical Journal* 2000;79:1601–1609. [PubMed: 10969020]
- Spangfort MD, Mirza O, Holm J, Larsen JN, Ipsen H, Lowenstein H. The structure of major birch pollen allergens–epitopes, reactivity and cross-reactivity. *Allergy* 1999;50:23–26. [PubMed: 10466032]
- Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *The FASEB Journal* 2003;17:1141–1143. [PubMed: 12709401]
- Teuber SS, Beyer K. Peanut, tree nut and seed allergies. *Current Opinion in Allergy and Clinical Immunology* 2004;4:201–203. [PubMed: 15126942]
- van Ree R, Vieths S, Poulsen LK. Allergen-specific IgE testing in the diagnosis of food allergy and the event of a positive match in the bioinformatics search. *Molecular Nutrition & Food Research* 2006;50:645–654. [PubMed: 16764014]
- Varshney S, Goldblum RM, Kearney C, Watanabe M, Midoro-Horiuti T. Major mountain cedar allergen, Jun a 1, contains conformational as well as linear IgE epitopes. *Molecular Immunology* 2007;44:2781–2785. [PubMed: 16423400]
- Venkatarajan MS, Braun W. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical–chemical properties. *Journal of Molecular Modeling* 2001;7:445–453.
- Venkatarajan SM, Schein CH, Braun W. Identifying property based sequence motifs in protein families and superfamilies: application to APE. *Bioinformatics* 2003;19:1381–1390. [PubMed: 12874050]
- Wensing M, Knulst AC, Piersma S, O’Kane F, Knol EF, Koppelman SJ. Patients with anaphylaxis to pea can have peanut allergy caused by cross-reactive IgE to vicilin (Ara h 1). *Journal of Allergy and Clinical Immunology* 2003;111:420–424. [PubMed: 12589366]
- WHO. Safety Aspects of Genetically Modified Foods of Plant Origin. Report of a Joint FAO/WHO Expert Consultation. World Health Organization; Geneva: 2000.
- WHO. Evaluation of Allergenicity of Genetically Modified Foods. Report of a Joint FAO/WHO Expert Consultation. World Health Organization; Geneva: 2001.
- WHO. Joint FAO/WHO Food Standards Programme. Codex Ad Hoc Intergovernmental Task Force on Foods Derived from Biotechnology. World Health Organization; Yokohama: 2003. Available from: <<http://www.codexalimentarius.net/>>

The screenshot shows the SDAP website interface. At the top, it says "UTMB The University of Texas Medical Branch" and "SDAP - All Allergens". There are navigation links for "SDAP Home Page", "SDAP All", and "SDAP Food". A search bar is present with the URL "http://fermi.utmb.edu/cgi-bin/SDAP/sdap_02?db_Type=0&allid=63".

The main content area is titled "Allergen Ole e 8". Below this title is a table with the following data:

Allergen	Ole e 8
Species - Systematic Name	Olea europea
Species - Trivial Name	olive
Source	tree Lamiales: Oleaceae
Description	Ca2+-binding protein

Below the table is a section "Ole e 8 - Protein Sequences" with a table of links to various databases:

Source	Link to Source	View Sequence	FASTA@SDAP	BLAST@ExPASy	BLAST@PIR	FASTA@PIR	PROSITE@PIR
SwissProt	Q9M7R0	Go!	Go!	Go!	Go!	Go!	Go!
GenBank	6901654	Go!	Go!	Go!	Go!	Go!	Go!

Below this table are instructions for using the search tools: FASTA@SDAP, BLAST@ExPASy, BLAST@PIR, FASTA@PIR, and PROSITE@PIR.

Next is a section "Ole e 8 - Pfam domains" with a table of Pfam domain information:

Sequence	Pfam sequence	Pfam family	Allergens from this Pfam family	Pfam database	Pfam region	First amino acid	Last amino acid
6901654	ALL8_OLEEU	PF00036	Go!	Pfam A	1	20	48
	ALL8_OLEEU	PF00036	Go!	Pfam A	2	56	84
	ALL8_OLEEU	PF00036	Go!	Pfam A	3	96	124

Below the Pfam table is a note: "Pfam is a database of multiple sequence alignments and hidden Markov models covering protein domains and families".

At the bottom, there is a section "Ole e 8 - Sequences from the Entrez Nucleotide database".

Fig. 1. Example of the information content for allergens in SDAP, illustrated here for the allergen Ole e 8. Links to software tools, other websites and information in other databases for this protein, can be direct accessed from this page.

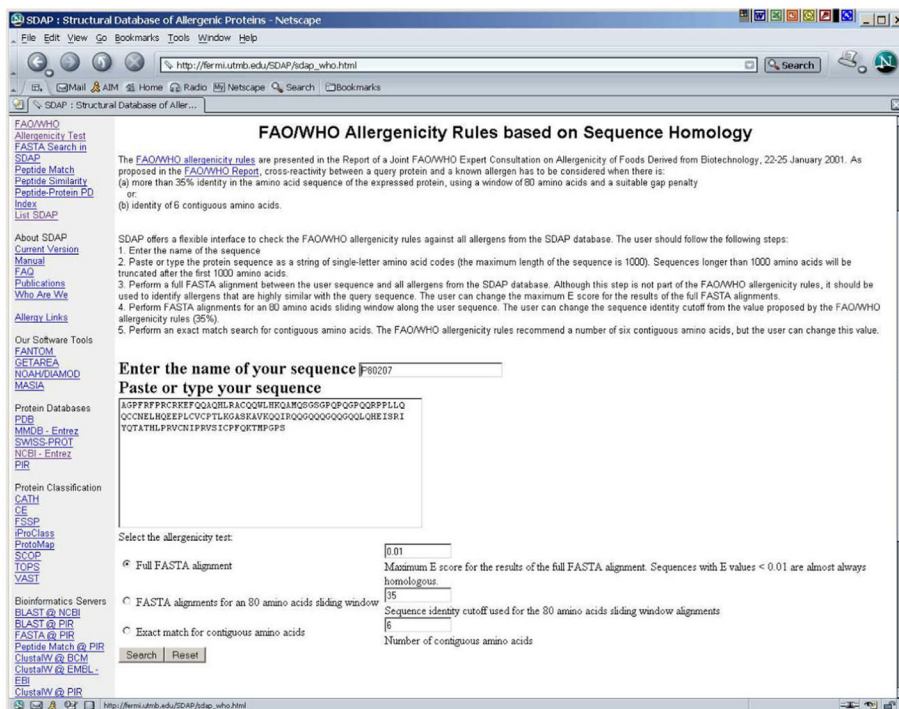


Fig. 2. Implementation of the FAO/WHO allergenicity guidelines in the SDAP Website. A user supplied sequence can be pasted into the window, and the FASTA search run according to the user selected criteria.

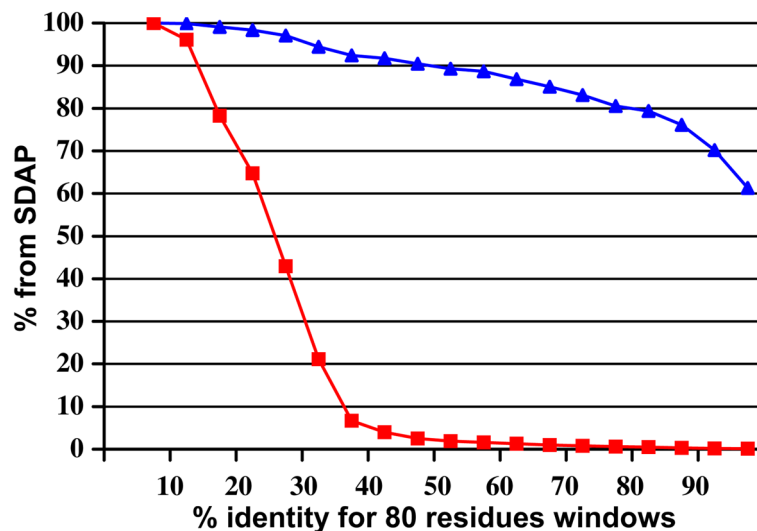


Fig. 3.

Evaluation of criterion 1 of the WHO/FAO rules: A protein is classified as an allergen if the sequence identity to an allergen is higher than a given threshold in a window of 80 aa residues. The fraction of positive controls (SDAP allergens) in blue and the fraction of negative controls (set of non-allergenic proteins) in red satisfying criterion 1 are given as a function of the threshold value. A cutoff value of 35–40% identities most allergens, but classifies about 6% of the control group as allergenic.

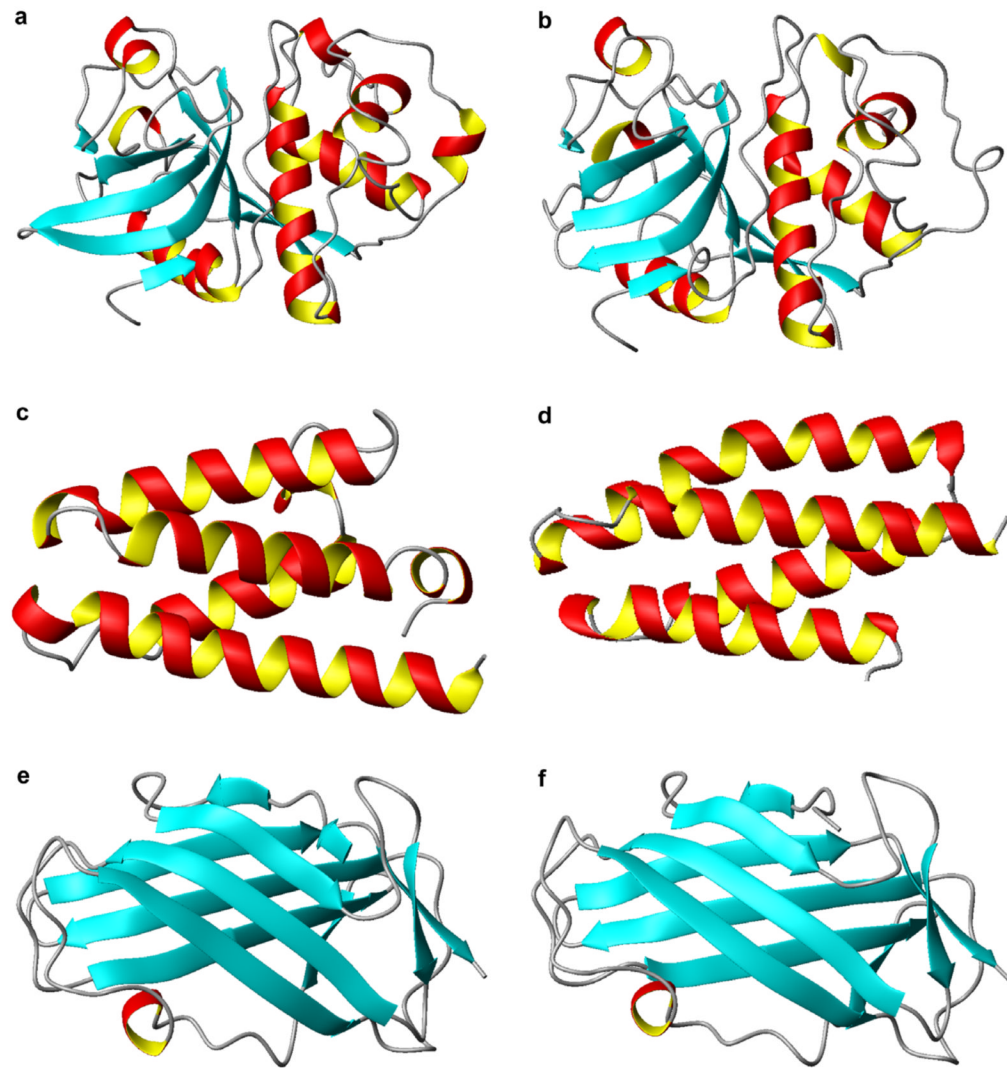


Fig. 4. PDB structures of allergens from three Pfam families: PF00112, Papain family cysteine protease: (a) Act c 1, 2ACT; (b) Car p 1, 1KHQ. PF01620, Ribonuclease (pollen allergen): (c) Phl p 5, 1L3P; (d) Phl p 6, 1NLX. PF02221, ML domain: (e) Der f 2, 1XWV; (f) Der p 2, 1KTJ.

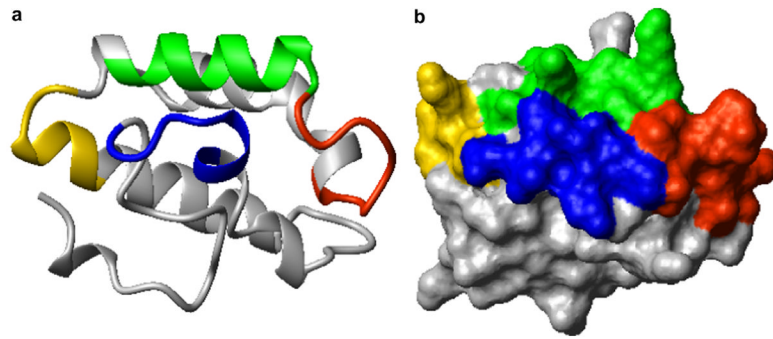


Fig. 6. IgE epitopes mapped on the MPACK model of Par j 1: (a) ribbon plot and (b) solvent accessible surface.

SDAP search with the sequence similarity index *PD* for epitope I (VQGKEKEP) from Par j 1 to identify similar regions in other allergens.

Table 1

No	Allergen	Sequence	<i>PD</i>	$z(PD, \text{all})$	Start residue	Matching region	End residue
1	Par j 1	P43217	0.00	6.8278	18	VQGKEKEP	25
2	Par j 2	P55958	2.52	5.8904	49	VKGEEKEP	56
3	Mala s 10	28564467	5.46	4.7988	599	VQSEEEKK	606
4	Bos d 8	162650	6.27	4.4965	21	IHAQQKEP	28
5	Tri a gliadin	170730	6.50	4.4123	202	VQPQQQP	209
6	Ole e 1	2465131	6.51	4.4098	42	LQCREKEN	49
7	Tri a glutenin	21773	6.72	4.3301	220	IQSQQQP	227
8	Ara h 2	9186485	6.73	4.3258	117	LQGRQQEQ	124
9	Der f 15	5815436	6.90	4.2641	401	INGDEKNS	408
10	Fra e 1	33327133	6.91	4.2618	41	LQCKDKEN	48
11	Cul n 1	12656498	6.96	4.2426	21	VDSKEAET	28
12	Hom s 3	929619	7.14	4.1742	52	TDSKEKEK	59
13	Dol m 2	P49371	7.14	4.1736	237	VQGRVKEA	244
14	Ves v 2	P49370	7.14	4.1736	237	VQGRVKEA	244
15	Pol a 2	Q9U6V9	7.36	4.0932	266	VQGRKEA	273
16	Gly m conglycinin	169929	7.51	4.0365	169	HQKKESEE	176
17	Jug n 2	31321944	7.62	3.9983	42	QQGRERGP	49
18	Jug r 2	6580762	7.62	3.9983	154	QQGRERGP	161
19	Chi k 10	7321108	7.67	3.9790	71	LEGKEKAL	78
20	Blot 12	Q17282	7.70	3.9667	26	TRGRHTEP	33
21	Ara h 1	P43237	7.70	3.9653	409	VKPDKKNP	416
22	Dic v a	763532	7.96	3.8696	1329	VPSDEKEK	1336
23	Solat 2	P16348	8.07	3.8303	11	TNGKELNP	18
24	Mal d 1	4590388	8.07	3.8289	134	MAGKEKAH	141
25	Hom s 5	1346344	8.10	3.8173	159	VRABEREQ	166
26	Solat 4	P30941	8.17	3.7933	39	VAGKELDS	46
27	Eur m 14	6492307	8.18	3.7897	1595	VQKPRKA	1602