JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Detecting Potential IgE-Reactive Sites on Food Proteins Using a Sequence and Structure Database, SDAP-Food

OVIDIU IVANCIUC,† VENKATARAJAN MATHURA,† TERUMI MIDORO-HORIUTI,‡ WERNER BRAUN,† RANDALL M. GOLDBLUM,‡ AND CATHERINE H. SCHEIN*,†

Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-1157 and Child Health Research Center, Department of Pediatrics, University of Texas Medical Branch, 310 University Boulevard, Galveston, Texas 77555-0366

The high incidence of food allergies, including oral allergy syndrome, represent major considerations when introducing new crops and foods. A new structural database of allergenic proteins, SDAP-Food, http://fermi.utmb.edu/SDAP/, has been developed to aid in predicting the IgE-binding potential of novel food proteins and cross-reactivities among known allergens. The site is designed to facilitate the first steps of a decision tree approach to determine the allergenicity of a given protein, based on the sequence and structural similarity to known allergens and their IgE binding sites. Immunological tests can then be used to confirm the predictions. A hierarchical procedure for identifying potential allergens, using a physical property-based sequence similarity index, has been designed to identify regions that resemble known IgE binding sites. As an example, SDAP tools were used to find food allergen sequences similar to an IgE binding site of the Jun a 3 allergen from mountain cedar pollen. The SDAP sequence similarity search matched the Jun a 3 epitope to regions in several food allergens, including cherry (Pru av 2), apple (Mal d 2) and pepper (Cap a 1), which are, like Jun a 3, members of the plant pathogenesis-related (PR-5) protein family. Homology modeling, using our EXDIS/DIAMOD/FANTOM program suite, indicated a similar surface location and structure for the potential epitope region on all of these allergens. The quantitative approach presented here can be used as part of a screening process for potential allergenicity of recombinant food products.

KEYWORDS: Allergen structure; pathogenesis related plant proteins; PR-5; Jun a 3; oral allergy syndrome; Pru av 2; Mal d 2; Cap a 1

INTRODUCTION

Food allergies (*1*, *2*), which affect approximately 8% of children and 2% of adults, including oral allergy syndrome (OAS), are factors that may limit the introduction of new dietary proteins. Sensitivity arises from formation of IgE antibodies against a few allergens that may have homologues in several different foods. In OAS, a form of contact allergy with oropharynx symptoms, tingling and angioedema of lips, tongue, palate and throat, patients sensitive to inhaled triggers, such as pollen, dust, or insect residue, may also react to foods containing similar proteins (*3*−*5*). OAS probably accounts for the simultaneous worldwide increase in seasonal pollen hypersensitivity and diagnosed food allergies (*6*, *7*). As more novel proteins are introduced into foods, medications, and other products in our environment, distinguishing allergens from other proteins becomes a more pressing issue (*8*, *9*).

The prediction of potential IgE epitopes in food produced by recombinant techniques is highly challenging, as they can include novel proteins not previously present in our food chain (*10*, *11*). Tomatoes with improved ripening characteristics that provide enhanced flavor, canola oil enriched with high levels of oleic acid and a monounsaturated fatty acid, "golden-rice" with enhanced levels of Vitamin A, virus-resistant squash and papaya have reached the market place (*12*). Recent concern over the possibility that a genetically modified Starlink corn might be allergenic pointed out the limitations of current criteria in assessing the allergenic potential of new food products (*13*). An expert panel judged the Cry9C protein to have "moderate capacity to be an allergen" due to its size (between 10 and 70 kD), resistance to acid and protease digestion, and limited results in animal models (*14*).

Analysis of allergen sequences and structures has clarified the physical basis of cross-reactivity among some allergens from different sources (*15*, *16*) and for OAS. Many of the allergenic proteins identified in foods have sequence and structural homology to those causing hypersensitivity to tree and grass pollen. For example, Bet v 1 (the major allergen of birch pollen) and Mal d 1 from apple are 56.3% identical in their amino acid sequences (*17*). It is estimated that 40−70% of patients allergic to pollen from birch and related pollens have OAS when they

* Corresponding Author. Tel.: 409 747-6843. Fax: 409 747-6850. E-mail: cathy@newton.utmb.edu.
† Sealy Center for Structural Biology.
‡ Child Health Research Center.

10.1021/jf034218r CCC: $25.00
Published on Web 06/25/2003

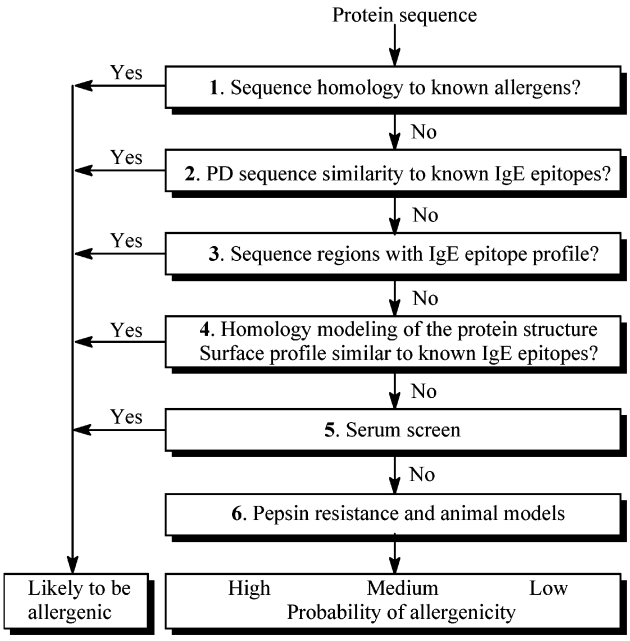

Figure 1. Decision tree for determining the potential allergenicity of a protein. SDAP is designed to aid in the first 4 steps, which measure to what degree the sequence and 3D structure of the protein match those of known allergens.

eat apples (*18*). There are several reports that patients allergic to Japanese cedar pollen develop sensitivity to apple, cherry and tomato (*19*−*21*). In addition to the demonstrated cross reactivity of the tomato protein polygalacturonase 2A and Cry j 2 from Japanese cedar pollen with patient IgE (*5*), recent results from this laboratory indicate that cross reactivity between the PR5 proteins, P23 in tomato and Jun a 3 from Texas mountain cedar, may also contribute to OAS in patients with cedar pollinosis (manuscript in preparation). These results suggest that some proteins, regardless of source and mechanism of exposure, share common features that make them particularly effective sensitizers and elicitors of allergic responses (*22*−*24*). Identifying these features could aid in the selection of crops and foods with reduced tendency to trigger an IgE-based response.

To aid in identifying the sequence and structural determinants of protein allergenicity (*23*, *24*), we implemented the web-based server SDAP (Structural Database of Allergenic Proteins, http://fermi.utmb.edu/SDAP), which allows easy access to the sequences, structures, and IgE epitopes of allergenic proteins (*25*). A new section of this site, SDAP-Food, has been designed to aid in the first steps of a sequence and structure based "decision tree" approach (**Figure 1**) to determine potential cross reactions between allergens. Previous results have shown that SDAP tools can rapidly identify cross-reactive allergens from sequence data (*26*). In the first section of this report, we describe the use of SDAP-Food and the integrated sequence search methods available in SDAP to identify potentially allergenic proteins in a given food source. We then show how the SDAP *PD* tool, coupled with structural models, can identify food sources that might trigger OAS in individuals who are sensitive to mountain cedar pollen.

MATERIALS AND METHODS

The *PD*-Value Tool. In the present release of SDAP, users can compare a given peptide to all the sequences in the SDAP allergen database, using either an exact match or a similarity search based on physicochemical properties, with the *PD*-value tool. As many food allergens are related to proteins that trigger inhalation or contact allergies, similarity searches initiated from the SDAP-Food section will also detect related sequences in the complete SDAP database of allergenic proteins. The SDAP peptide exact match function is useful for identifying closely related allergens that have identical IgE epitopes. To identify more distantly related sequences, the SDAP *PD*-value tool can be used. This tool uses the E_1-E_5 vectors derived by multidimensional scaling of 237 physical-chemical properties for all 20 naturally occurring amino acids (*27*) to describe each position of a sequence in terms of its chemical properties. The query peptide, of length N, is translated into an $N \times 5$ numerical matrix containing the values of the 5 physical chemical descriptors at each sequence position. This matrix is then compared to that for all the sequence windows of equal length in SDAP. The property distance (*PD*) value, a measure of the similarity between two sequences A and B, of length N residues, is (*26*, *25*)

$$PD(\mathrm{A,B}) = \frac{1}{N}\sum_{i=1}^{N}[\sum_{j=1}^{5}\lambda_j(E_j(\mathrm{A}_i) - E_j(\mathrm{B}_i))^2]^{1/2}$$

where λ_j is the eigenvalue of the j-th E component, $E_j(\mathrm{A}_i)$ is the E_j value for the amino acid in the i-th position from sequence A, and $E_j(\mathrm{B}_i)$ is the E_j value for the amino acid in the i-th position from sequence B.

The SDAP tool calculates the *PD* similarity index between the query sequence and each sequence window with the same length from all allergens collected in the SDAP protein database. The search result is a list of similar sequences identified in allergenic proteins, presented in decreasing order of similarity (increasing *PD*) with the query sequence. In addition to epitope identification, this tool can be used to find conserved regions in the allergens from the SDAP protein database. Because the *PD* sequence-similarity index is based on the E_1-E_5 scale of physical−chemical properties of amino acids, it can be successfully used in determining the homology and similarity of related allergens.

Homology Modeling. The best homology template for the sequences of the PR5 allergens Pru av 2 (28) (SwissProt P50694), Mal d 2 (29) (GenBank 10334651), and the available fragment of Cap a 1 (30), which is missing both N- and C-termini (GenBank 11321159) were identified with the 3D-PSSM server (http://www.sbg.bio.ac.uk/∼3dpssm). The best template for all three food allergens was a high resolution (1.80 Å) crystal structure of a thaumatin-like PR-5D protein from tobacco (PDB file 1AUN). The sequence alignment (**Table 3**) was hand-modified to improve the overall identity. The 16 cysteine residues common to all thaumatin-like proteins (TLPs) form eight disulfide bridges that could be aligned between the allergens and the tobacco protein. Pru av 2 and Mal d 2 have eight disulfide bonds each, while the fragment of Cap a 1 modeled has only six bonds and two unlinked cysteines. The disulfide bonds in the final models are between the following cysteine pairs: Pru av 2, 10−222, 58−68, 73−80, 128−211, 133−194, 141−157, 161−170, 171−181; Mal d 2, 10−222, 58−68, 73−80, 128−211, 133−194, 141−157, 161−170, 171−181; Cap a 1, 45−55, 60−66, 119−169, 127−137, 141−150, 151−156.

In the second step of the homology modeling, our program EXDIS was used to extract interatomic distance and dihedral angle constraints from the structure of the template PR-5D. The segments of 1AUN used by EXDIS were as follows: Pru av 2, 2−22, 32−47, 50−73, 77−83, 87−121, 124−148, 157−172, 180−212, 215−223; Mal d 2, 2−22, 32−47, 50−73, 77−83, 87−121, 124−148, 157−172, 180−212, 215−223; Cap a 1, 2−179. The self-correcting distance geometry-based DIAMOD program was then used to generate protein structures to satisfy these constraints (*31*−*33*). Our FANTOM program (*34*) for energy minimization with the ECEPP/2 all-atom force field (*35*) was used to optimize the target structure, subject to the distance constraints and dihedral angle constraints. The program PROCHECK from the Biotech Validation Suite for Protein Structures (http://biotech.embl-heidelberg.de:8400/) was used to stereochemically and geometrically validate the models.

RESULTS AND DISCUSSION

Structural Blocks and Main Functions of SDAP-Food Allergens. SDAP (*26*, *25*) is an Internet service that uses a web

Table 1. Other Websites Provide Lists of Information about Allergens, but Lack the Onsite Data and Tools of SDAP

website name	address	information available
All Allergy	http://allallergy.net/	a portal to allergy information, useful for the general public
IUIS (International Union of Immunological Societies)	http://www.allergen.org	lists official names, grouped by source, and Genbank accession numbers of allergens
National Center for Food Safety and Technology	http://www.iit.edu~sgendel/fa.htm	lists official names of food allergens with links to Genbank
CSL (Central Science Laboratory, UK)	http://www.csl.gov.uk/allergen/index.htm	lists official names of allergens with sequence links to Genbank
Allergome	http://www.allergome.org	lists the official names of allergens, and links to PubMed & sequence databases
Swiss-Prot	http://us.expasy.org/cgi-bin/lists?allergen.txt	list of allergens, with sequence data
Farrp	http://allergenonline.com/asp/public/login.asp	lists official names of allergens, sequence links to Genbank, and a FASTA search for related sequences
Protall	http://www.ifrn.bbsrc.ac.uk/protall/	allergen names, plus links to detailed biochemical, structural, and clinical data
Modbase	http://alto.rockefeller.edu/modbase-cgi/index.cgi	automatically generated models for Swiss-Prot and TrEMBL sequences
SDAP	http://fermi.utmb.edu/SDAP	allergens sequences, on-site and cross-referenced by source and protein type, with links to all major sequence and structural databases, IgE epitopes collection, tools for sequence and epitope comparison, on site information about experimental structures of allergens, and high-quality protein models

browser to communicate with the MySQL database and various software tools. The assembly of the current database of allergens was guided by the list of allergen names from the IUIS website (see **Table 1**). The major food allergens are casein and lactoglobulin from milk, ovomucoid from egg, tropomyosin from shrimp and related species, albumin from nuts and mustard, α-amylase inhibitors from rice, wheat and barley, or profilins and pathogenesis-related proteins from various plants (*36*). The sequences of the allergenic proteins isolated from these and other foods have been compiled in a separate section, SDAP-Food. The information is collected in tables according to the following: allergen type; species; systematic name and brief description; and sequence accession numbers from SwissProt (*37*), PIR (*38*), NCBI (*39*), and where available, the PDB (*40*) file name. A user can now go directly to "SDAP-food allergens" and determine, for example, known allergenic proteins in a given foodstuff (e.g., peanut, apple, shrimp, etc.). The user can click on the block for the individual allergen to obtain specific information, such as amino acid sequence, experimentally determined 3D structure (PDB file), 3D model, or IgE epitopes within SDAP and to connect to other bioinformatics servers.

Many food allergens cross-react with nonfood allergens. For example, the known allergens of banana resemble other allergenic proteins of the Hevea family, including latex (*41*−*44*). The tropomyosins in shell fish are closely related to similar proteins that induce an IgE reaction to dust mites (*45*−*47*). There are numerous clinical reports linking allergies to foods and pollen proteins (*3*, *48*−*50*). Thus, the food allergy section of the database is completely integrated with the rest of SDAP, and epitopes matching those of a food allergen can come from any allergen subgroup in SDAP. This makes the database particularly useful for discerning possible OAS triggers.

Unique Features of SDAP. Table 1 lists other sites on the web that provide information about allergenic proteins. Most of these provide only clinical information or lists of names, sometimes with links to sequence or structural databases. SDAP is unique in providing cross-referenced lists of sequence, epitope, and structural data of allergenic proteins stored at one site. We are also working to provide high quality models for all allergens of known sequence. As noted in the table, the MODBASE (*51*) site provides structures for allergens, generated automatically with the Modeller (http://salilab.org/modeller/modeller.html) (*52*) program. Our modeling procedures (described below) yield very accurate structures, as evidenced by the performance of our models in the Critical Assessment of Techniques for Protein Structure Prediction competitions CASP4 (*53*) and CASP5 (manuscript in preparation). There is a direct link to our server GETAREA (http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl) (*54*), so that users can rapidly determine the surface exposed area of each residue in a protein from a structure file in PDB format. This will greatly aid in determining amino acids in known epitopes that are likely to form the IgE binding site.

Evaluating Allergenicity with a Decision Tree Approach. A decision tree, shown in **Figure 1**, combines computational and experimental tests to determine whether a protein is a potential allergen. This decision tree is a modification of the one proposed by IFBC (the International Food Biotechnology Council) and ILSI (Allergy and Immunology Institute of the International Life Sciences Institute) (*55*) and adapted by FAO/WHO (*56*, *57*). The SDAP-Food website is designed to provide the computational tests at the top of this tree, while the remaining steps from the decision tree (**Figure 1**) comprise laboratory and clinical tests. Steps 1, 2 and 3 from the decision tree use a hierarchical approach to determine whether a protein sequence, provided by the user, contains elements similar to known allergenic proteins. These steps rely on the comprehensive list of information about known allergenic proteins that is summarized in SDAP and the SDAP tools for determining sequence similarity.

Steps 2 and 3 in the decision tree require a list of IgE epitopes in known allergens and a way to determine to what degree the sequence supplied by the user matches these epitopes. The PD tool was developed for this step. SDAP has a list of IgE epitopes, which have been collected from the literature. The user may consult this list, or provide the sequence of another epitope, to determine which allergens in the database match the given sequence. This procedure will be described in detail below. For Step 4, an additional structural filter is applied to the sequence analysis. This allows screening for sequences that are expected to adapt a similar three-dimensional fold to the test epitope. As there are relatively few experimentally determined structures for allergenic proteins, we are in the process of adding structural models, based on homology to a known structure, to the database. Our MASIA/EXDIS/DIAMOD/FANTOM (*58*, *34*, *32*) suite is being adapted for this work, as described below. In the

Table 2. Allergen Sequences from SDAP with Regions Most Similar (*PD* value < 9.5) to DNCPATNYSK from Jun a 3[a]

Allergen	PD	Matching region		
Jun a 3	0.00	182	DNCPATNYSK	191
Cup a 3	0.59	156	NNCPATNYSK	165
Pru av 2	5.53	201	ETCPPTNYSE	210
Mal d 2	7.00	202	ETCPPTEYSE	211
Equ c 2	7.65	3	DPQSETDYSQ	12
Amb a 1	7.96	233	SNCKFTQQSK	242
Jun o 2	8.37	2	DEVPSSDGSK	11
Asp f 2	8.68	54	SSCNATQRRQ	63
Aed a 3	8.73	133	KNDPADTYRQ	142
Asp f 13	8.93	285	ENSDASNTSP	294
Cla h 3	9.05	105	DNGKATSMAR	114
Dol m 5	9.08	139	SSTTATQFDR	148
Equ c 2	9.11	3	DPQSETDYSQ	12
Sol i 3	9.12	18	ANTLATNYCN	27
Ves v 1	9.46	290	SSQPISSCTK	299

[a] Cup a 3, a close relative of Jun a 3, was isolated from pollen of a cypress species.

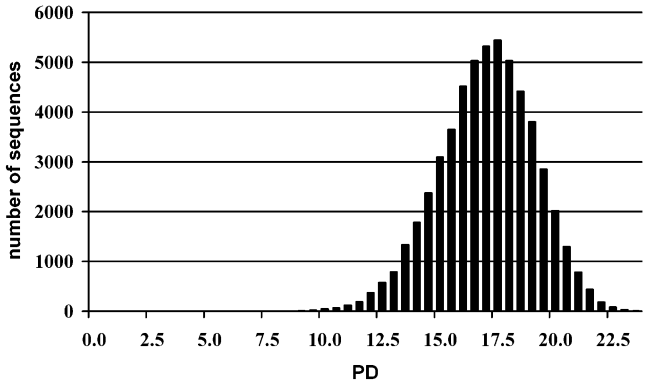

Figure 2. Histogram of *PD* values for the Jun a 3 IgE binding region DNCPATNYSK found by searching all 55873 sequence windows in the SDAP allergen database. Sequences with scores lower than ~9.5 would be below the maximum of the population distribution and considered significant. However, as noted in the text, a more generous cutoff value may be used to ensure that all related sequences are found.

final steps of the decision tree, immunological and biophysical assays are used to conclusively determine the allergenic potential of a protein.

The *PD* Tool Identifies Food Allergens Similar to a Cedar Pollen Allergen. Cedar pollen is a major cause of seasonal allergies in Texas and throughout the Northern Hemisphere, particularly in Japan and southern Europe (*5*, 59−*61*). We have identified two major allergens in pollen from mountain cedar trees (*62*, *63*). We modeled the structure of one of these, the PR-5 protein Jun a 3, and mapped the location of peptides that bound patient IgE to loop regions on the surface (*62*, *63*, *31*). A number of food allergens have since been identified to be PR5 proteins (*64*), including Pru av 2 from cherry, Mal d 2 from apple, and Cap a 1 from bell pepper.

We used the sequence of one of the Jun a 3 IgE binding sites and the SDAP *PD* tool to identify foods that might trigger OAS in patients sensitive to cedar pollen. First, we computed *PD* values between the IgE binding peptide sequence DNCPATNYSK (*31*) and all decapeptides from the 285 allergens in SDAP. A sequence window identical to the query sequence DNCPATNYSK has a *PD* value of 0, and the *PD* value for a given sequence increases with the dissimilarity of the two sequences. The distribution of the *PD* values for the 55873 sequence windows from the SDAP database of allergens shows that there are few sequences below *PD* 9.5, and the average value is 17.27 (**Figure 2**). Cross-reactive allergens should, according to our hypothesis, have *PD* values significantly less than the average of *PD* values for all decaptides in the database.

Table 3. Alignment of One Pollen and Three Food Allergens with the Template PDB File 1AUN Used to Prepare the Models of Figure 3[a]

```
Jun a 3   -VKFDIKNQCGYTVWAAGLP-GG-------GKRLDQGQTWTVNLAAGTASARFWGRTGCT
Mal d 2   AAKITFTNNCPNTVWPGTLTGDQKPQLSLTGFELASKASRSVDAPSPW-SGRFWGRTRCS
Pru av 2  AATISFKNNCPYMVWPGTLTSDQKPQLSTTGFELASQASFQLDTPVPW-NGRFWARTGCS
Cap a 1   -------NNCPYTVWAAATPVGG-------GKRLERGQSWWFWAPPGTKMARIWGRTNCN
1aun      SGVFEVHNNCPYTVWAAATPVGG-------GRRLERGQSWWFWAPPGTKMARIWGRTNCN
                 *:*:::**..  . ..        * .*  ..:.   .: ..  .* *:** *

Jun a 3   FDASGKGSCQTGDCG-GQLSCTVSGAVP-ATLAEYTQ---SDQDYYDVSLVDGFNIPLAI
Mal d 2   TDAAGKFTCETADCGSGQVACNGAGAVPPATLVEITIAANGGQDYYDVSLVDGFNLPMSV
Pru av 2  TDASGKFVCATADCASGQVMCNGNGAIPPATLAEFNIPAGGGQDFYDVSLVDGFNLPMSV
Cap a 1   FDGAGRGWCQTGDCG-GVLECKGWG-KPPNTLAEYALNQFSNLDFWDISVIDGFNIPMSF
1aun      FDGAGRGWCQTGDCG-GVLECKGWG-KPPNTLAEYALNQFSNLDFWDISVIDGFNIPMSF
          .* .* . *.*.**: * . * : *  ** **:*.     .  *. * *  ****.*::

Jun a 3   NPT---NAQCTAPACKADINAVCPSELKVDG------GCNSACNVFKTDQYCCRN--AYV
Mal d 2   APQ-GGTGECKPSSCPANVNKVCPAPLQVKAADGSVISCKSACLAFGDSKYCCTPPNNTP
Pru av 2  TPQ-GGTGDCKTASCPANVNAVCPSELQKKGSDGSVVACLSACVKFGTPQYCCTPPQNTP
Cap a 1   GPTKPGPGKCHPIQCVANINGECPGSLRVPG------GCNNPCTTFGGQQYCCTQ-----
1aun      GPTKPGPGKCHGIQCTANINGECPGSLRVPG------GCNNPCTTFGGQQYCCTQ-----
           *   * : *    * *: *  **  * : :       .*   *  *:  :***:

Jun a 3   DNCPATNYSKIFKNQCPQAYSYAKDDT-ATFACAS-GTDYSIVFCP
Mal d 2   ETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSG-GPDYVITFCP
Pru av 2  ETCPPTNYSEIFHNACPDAYSYAYDDKRGTFTCNG-GPNYAITFCP
Cap a 1   GPCGPTDLSRFFKQRCPDAYSYPQDDP------------------
1aun      GPCGPTELSRWFKQRCPDAYSYPQDDPTSTFTCTSWTTDYKVMFCP
            * :*  *  *   **.****  **   **:*     :*   ***
```

[a] The Jun a 3 epitope used for **Figures 3** and **4** and **Table 2** and the corresponding regions in each of the other proteins are shown in bold type. The symbols below the aligned sequences indicate the residues conservation relative to 1aun: * = all residues are conserved, : = all but one residues are conserved, . = all but two residues are conserved.

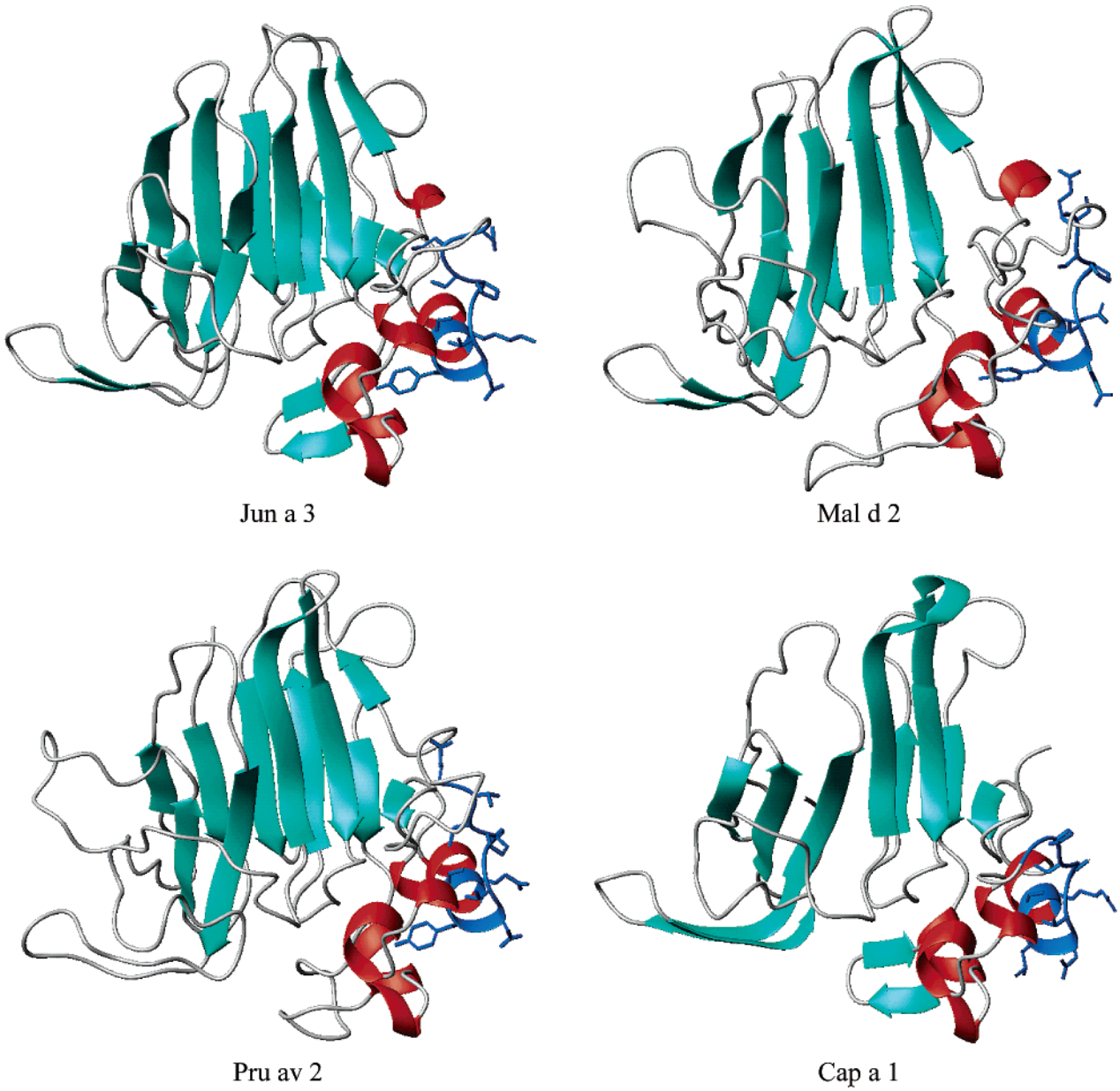

Figure 3. Ribbon diagrams of models for Jun a 3, Mal d 2, Pru av 2, and Cap a 1, prepared with our EXIDS/DIAMOD/FANTOM suite, based on the alignment with the template 1AUN, shown in **Table 1**. The region corresponding to the IgE binding site on Jun a 3, DNCPATNYSK, used for the calculations of **Figure 3** and **Table 2**, is shown in blue, with the side chains. The figure was prepared with the program MOLMOL (68).

We chose 9.5 as our threshold for **Table 2**, also based on our previous experiments with other IgE epitopes (*26*).

Table 2 lists all sequences found in SDAP with a *PD* value less than the threshold of 9.5. The sequence with the lowest *PD* value, 0.59, was from Cup a 3, an allergen from the cypress *Cupressus arizonica*, a close relative of mountain cedar. Two food proteins, Pru av 2 (*PD* 5.53) and Mal d 2 (*PD* 7.00), had *PD* values below the cutoff value of 9.5. The result is significant, as the proteins were chosen from a database in which all the proteins are known allergens. Three other PR-5 family allergens, Cap a 1 (*PD* 9.95) from bell pepper, the thaumatin-related tomato protein P23 (P12670; *PD* 10.03), and PR-5D (*PD* 10.82) from tobacco have *PD* values slightly higher than the threshold selected on statistical grounds. Immunological assays will be necessary to determine how accurately the *PD* value of these proteins reflects their true potential to cross react with cedar pollen.

The user of the *PD*-tool can choose the number of sequences to be printed out and the *PD* value cutoff. However, many of the finds will be random if the threshold is set too high. To help the user in selecting a proper threshold, the search result also gives a histogram of the *PD* distribution in the database, similar to **Figure 2**. Applying a structural filter to the results, as described in the next section, allows the user to select among entries with higher *PD* values for those likely to be significant.

Structural Modeling of PR-5 Food Allergens. In addition to high sequence similarity, as indicated by low *PD* values, the configuration and surface location of the suspected sequences should be compared to determine potential cross-reactivity between two allergens. SDAP provides access to the known structures of allergens in the protein database (PDB), and will eventually provide 3D models for all other allergens of known sequence. An experimental structure is not available for any of the proteins detected in our sequence search, but all of them have high homology to other PR-5 family members whose structures are available. Using methods described previously (*31*, *32*, *65*) we prepared 3D models for the three PR-5 food

proteins detected by the *PD* search, using the structure of 1AUN, a pathogenesis related protein from tobacco.

The alignment used for the modeling is shown in **Table 3** and the resulting models in **Figure 3**. All three models had acceptably low ECEPP energies (Mal d 2, −355 kcal/mol; Pru av 2, −230 kcal/mol; Cap a 1, −1017 kcal/mol) and few or no residues outside the generously allowed regions in Ramachandran plots (Mal d 2, 8; Pru av 2, 5; Cap a 1, 0). The backbone root-mean-square deviations (RMSD) of the final models to the template structure are low (Pru av 2, 0.493 Å; Mal d 2, 0.464 Å; Cap a 1, 0.187 Å), indicating a high degree of structural similarity, as was expected from the high sequence identity (**Table 3**).

Similarity of the Surface Location and Structure for the Identified Sequences. Figure 3 shows the modeled structures of the three food allergens in comparison to that of the pollen protein Jun a 3. The region corresponding to the IgE-binding peptide from Jun a 3, used for the PD-search of **Figure 2** and **Table 2**, is highlighted. The region in Cap a 1, which is not from a tree product, has a higher *PD* value. The structure and surface location of the region is similar in all four allergen models. These results suggest that patients with high sensitivity to cedar pollen may also react to apples and cherries, but not necessarily to bell peppers.

The indicated sequences will be tested in dot-spot immunoblotting or in solution assays for their ability to cross-react with IgE from sera of patients suffering from cedar pollen allergy. Results from these studies can be used, in line with the decision tree, to further refine the description of characteristics that define the IgE binding sites of PR-5 family members. For example, the PR-5 protein thaumatin is of interest as a sweetener (*66*). Depending on future results from these studies, we may select isoforms with reduced potential allergenicity. This can aid in the design of novel proteins and selection of crops that are less likely to evoke allergic responses.

CONCLUSIONS

Food allergens affect a significant proportion of the population, and new food proteins can be health hazards if they cross-react with other food or pollen allergens (*55*, *56*, *67*). The results shown here demonstrate how one can quickly identify potentially important similarities between allergenic proteins with SDAP and the incorporated tools. SDAP-Food is the first cross-referenced, interactive website of sequences and structures of food proteins. SDAP also contains a list of known IgE epitopes and novel tools for comparing sequences. In the example shown here, the *PD*-value tool of SDAP quickly identified food allergens that contain an epitope similar in sequence and predicted structure to one from cedar pollen. A patient with pollen allergy could be warned that fruits containing these proteins might induce OAS.

Because almost all known allergens belong to a small number of sequence families, we plan to generate and include homology models in SDAP, like those shown in **Figure 3**, for all food allergens of known sequence but unknown structure. The predicted structures will be useful for visualization and computational studies to characterize epitope structure. These data will eventually allow us to direct the design of proteins and select crops with reduced allergenicity.

LITERATURE CITED

(1) Sampson, H. A. Food allergy. Part 2: Diagnosis and management. *J. Allergy Clin. Immunol.* **1999**, *103*, 981−989.

(2) Sampson, H. A. Food allergy. Part 1: Immunopathogenesis and clinical disorders. *J. Allergy Clin. Immunol.* **1999**, *103*, 717−728.

(3) Yagami, T. Allergies to cross-reactive plant proteins − Latex-fruit syndrome is comparable with pollen-food allergy syndrome. *Int. Arch. Allergy Immunol.* **2002**, *128*, 271−279.

(4) Petersen, A.; Vieths, S.; Aulepp, H.; Schlaak, M.; Becker, W. M. Ubiquitous structures responsible for IgE cross-reactivity between tomato fruit and grass pollen allergens. *J. Allergy Clin. Immunol.* **1996**, *98*, 805−815.

(5) Kondo, Y.; Tokuda, R.; Urisu, A.; Matsuda, T. Assessment of cross-reactivity between Japanese cedar (Cryptomeria japonica) pollen and tomato fruit extracts by RAST inhibition and immunoblot inhibition. *Clin. Exp. Allergy* **2002**, *32*, 590−594.

(6) Sloane, D.; Sheffer, A. Oral allergy syndrome. *Allergy Asthma Proc.* **2001**, *22*, 321−325.

(7) Wuthrich, B. Food-induced cutaneous adverse reactions. *Allergy* **1998**, *53*, 131−135.

(8) Kimber, I.; Dearman, R. J. Approaches to assessment of the allergenic potential of novel proteins in food from genetically modified crops. *Toxicol. Sci.* **2002**, *68*, 4−8.

(9) Kimber, I.; Dearman, R. J. Food allergy: what are the issues? *Toxicol. Lett.* **2001**, *120*, 165−170.

(10) Lack, G. Clinical risk assessment of GM foods. *Toxicol. Lett.* **2002**, *127*, 337−340.

(11) Lack, G.; Chapman, M.; Kalsheker, N.; King, V.; Robinson, C.; Venables, K. Report on the potential allergenicity of genetically modified organisms and their products. *Clin. Exp. Allergy* **2002**, *32*, 1131−1143.

(12) Gendel, S. M. Sequence Databases for Assessing the Potential Allergenicity of Proteins Used in Transgenic Foods. *Adv. Food Nutrition Res.* **1998**, *42*, 63−92.

(13) Bucchini, L.; Goldman, L. R. Starlink corn: A risk analysis. *Environ. Health Perspect.* **2002**, *110*, 5−13.

(14) FIFRA "A set of scientific issues being considered by the environmental protection agency regarding: assessment of scientific information concerning StarLink Corn.," FIFRA, 2000.

(15) Ipsen, H.; Lowenstein, H. Basic features of crossreactivity in tree and grass pollen allergy. *Clin. Rev. Allergy Immunol.* **1997**, *15*, 389−396.

(16) Sicherer, S. H. Clinical implications of cross-reactive food allergens. *J. Allergy Clin. Immunol.* **2001**, *108*, 881−890.

(17) Scheurer, S.; Son, D. Y.; Boehm, M.; Karamloo, F.; Franke, S.; Hoffmann, A.; Haustein, D.; Vieths, S. Cross-reactivity and epitope analysis of Pru a 1, the major cherry allergen. *Mol. Immunol.* **1999**, *36*, 155−167.

(18) Vanek-Krebitz, M.; Hoffmannsommergruber, K.; Machado, M. L. D.; Susani, M.; Ebner, C.; Kraft, D.; Scheiner, O.; Breiteneder, H. Cloning and Sequencing of Mal-D-1, the Major Allergen from Apple (Malus-Domestica), and Its Immunological Relationship to Bet-V-1, the Major Birch Pollen Allergen. *Biochem. Biophys. Res. Commun.* **1995**, *214*, 538−551.

(19) Ishida, T.; Murai, K.; Yasuda, T.; Satou, T.; Sejima, T.; Kitamura, K. Oral allergy syndrome in patients with Japanese cedar pollinosis. *Nippon Jibiinkoka Gakkai Kaiho* **2000**, *103*, 199−205.

(20) Arai, Y.; Ogawa, C.; Ohtomo, M.; Sano, Y.; Ito, K. Food and food additives hypersensitivity in adult asthmatics. II. Oral allergy syndrome in adult asthmatics with or without Japanese hay fever. *Arerugi* **1998**, *47*, 715−719.

(21) Kondo, Y.; Urisu, A.; Tokuda, R. Identification and characterization of the allergens in the tomato fruit by immunoblotting. *Int. Arch. Allergy Immunol.* **2001**, *126*, 294−299.

(22) Aalberse, R. C. Structural biology of allergens. *J. Allergy Clin. Immunol.* **2000**, *106*, 228−238.

(23) Oehlschlager, S.; Reece, P.; Brown, A.; Hughson, E.; Hird, H.; Chisholm, J.; Atkinson, H.; Meredith, C.; Pumphrey, R.; Wilson, P.; Sunderland, J. Food allergy − towards predictive testing for novel foods. *Food Addit. Contam.* **2001**, *18*, 1099−1107.

(24) Huby, R. D. J.; Dearman, R. J.; Kimber, I. Why are some proteins allergens? *Toxicol. Sci.* **2000**, *55*, 235−246.

(25) Ivanciuc, O.; Schein, C. H.; Braun, W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res.* **2003**, *31*, 359−362.
(26) Ivanciuc, O.; Schein, C. H.; Braun, W. Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics* **2002**, *18*, 1358−1364.
(27) Venkatarajan, M. S.; Braun, W. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. *J. Mol. Model.* **2001**, *7*, 445−453.
(28) Inschlag, C.; Hoffmann-Sommergruber, K.; O'Riordain, G.; Ahorn, H.; Ebner, C.; Scheiner, O.; Breiteneder, H. Biochemical characterization of Pru a 2, a 23-kD thaumatin-like protein representing a potential major allergen in cherry (Prunus avium). *Int. Arch. Allergy Immunol.* **1998**, *116*, 22−28.
(29) Hsieh, L. S.; Moos, M.; Lin, Y. Characterization of apple 18 and 31 kd allergens by microsequencing and evaluation of their content during storage and ripening. *J. Allergy Clin. Immunol.* **1995**, *96*, 960−970.
(30) Jensen-Jarolim, E.; Santner, B.; Leitner, A.; Grimm, R.; Scheiner, O.; Ebner, C.; Breiteneder, H. Bell peppers (Capsicum annuum) express allergens (profilin, pathogenesis-related protein P23 and Bet v 1) depending on the horticultural strain. *Int. Arch. Allergy Immunol.* **1998**, *116*, 103−109.
(31) Soman, K.; Midoro-Horiuti, T.; Ferreon, J.; Goldblum, R.; Brooks, E.; Kurosky, A.; Braun, W.; Schein, C. H. Homology modeling and characterization of IgE epitopes of mountain cedar allergen Jun a 3. *Biophys. J.* **2000**, *79*, 1601−1609.
(32) Soman, K. V.; Schein, C. H.; Zhu, H.; Braun, W. Homology modeling and simulations of nuclease structures. In *Nuclease Methods and Protocols, Methods in Molecular Biology*; Schein, C. H., Ed.; Humana Press: Totowa, NJ, 2001; pp 263−286.
(33) Schein, C. H.; Nagle, G. T.; Page, J. S.; Sweedler, J. V.; Xu, Y.; Painter, S. D.; Braun, W. Aplysia attractin: Biophysical characterization and modeling of a water-borne pheromone. *Biophys. J.* **2001**, *81*, 463−472.
(34) Schaumann, T.; Braun, W.; Wuthrich, K. The program FANTOM for energy refinement of polypeptides and proteins using a Newton−Raphson minimizer in torsion angle space. *Biopolymers* **1990**, *29*, 679−694.
(35) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. Energy Parameters in Polypeptides 0.7. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen-Bond Interactions, and Intrinsic Torsional Potentials for Naturally Occurring Amino-Acids. *J. Phys. Chem.* **1975**, *79*, 2361−2381.
(36) Breiteneder, H.; Ebner, C. Molecular and biochemical classification of plant-derived food allergens. *J. Allergy Clin. Immunol.* **2000**, *106*, 27−36.
(37) Bairoch, A.; Apweiler, R. The SWISS−PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28*, 45−48.
(38) Wu, C. H.; Huang, H. Z.; Arminski, L.; Castro-Alvear, J.; Chen, Y. X.; Hu, Z. Z.; Ledley, R. S.; Lewis, K. C.; Mewes, H. W.; Orcutt, B. C.; Suzek, B. E.; Tsugita, A.; Vinayaka, C. R.; Yeh, L. S. L.; Zhang, J.; Barker, W. C. The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res.* **2002**, *30*, 35−37.
(39) Wheeler, D. L.; Church, D. M.; Lash, A. E.; Leipe, D. D.; Madden, T. L.; Pontius, J. U.; Schuler, G. D.; Schriml, L. M.; Tatusova, T. A.; Wagner, L.; Rapp, B. A. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2001**, *29*, 11−16.
(40) Westbrook, J.; Feng, Z. K.; Jain, S.; Bhat, T. N.; Thanki, N.; Ravichandran, V.; Gilliland, G. L.; Bluhm, W.; Weissig, H.; Greer, D. S.; Bourne, P. E.; Berman, H. M. The Protein Data Bank: Unifying the archive. *Nucleic Acids Res.* **2002**, *30*, 245−248.
(41) Wagner, S.; Breiteneder, H. The latex-fruit syndrome. *Biochem. Soc. Trans.* **2002**, *30*, 935−940.
(42) Grob, M.; Reindl, J.; Vieths, S.; Wuthrich, B.; Ballmer-Weber, B. K. Heterogeneity of banana allergy: characterization of allergens in banana-allergic patients. *Ann. Allergy Asthma Immunol.* **2002**, *89*, 513−516.
(43) Crespo, J. F.; Rodriguez, J.; James, J. M.; Daroca, P.; Reano, M.; Vives, R. Reactivity to potential cross-reactive foods in fruit-allergic patients: implications for prescribing food avoidance. *Allergy* **2002**, *57*, 946−949.
(44) Perkin, J. E. The latex and food allergy connection. *J. Am. Diet. Assoc.* **2000**, *100*, 1381−1384.
(45) Ayuso, R.; Reese, G.; Leong-Kee, S.; Plante, M.; Lehrer, S. B. Molecular basis of arthropod cross-reactivity: IgE-binding cross-reactive epitopes of shrimp, house dust mite, and cockroach tropomyosins. *Int. Arch. Allergy Immunol.* **2002**, *129*, 38−48.
(46) Santos, A. B. R.; Chapman, M. D.; Aalberse, R. C.; Vailes, L. D.; Ferriani, V. P. L.; Oliver, C.; Rizzo, M. C.; Naspitz, C. K.; Arruda, L. K. 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.* **1999**, *104*, 329−337.
(47) vanRee, R.; Antonicelli, L.; Akkerdaas, J. H.; Garritani, M. S.; Aalberse, R. C.; Bonifazi, F. Possible induction of food allergy during mite immunotherapy. *Allergy* **1996**, *51*, 108−113.
(48) Grote, M.; Stumvoll, S.; Reichelt, R.; Lidholm, J.; Valenta, R. Identification of an allergen related to Phl p 4, a major timothy grass pollen allergen, in pollens, vegetables, and fruits by immunogold electron microscopy. *Biol. Chem.* **2002**, *383*, 1441−1445.
(49) Ballmer-Weber, B. K. Lipid transfer protein as a potential panallergen? *Allergy* **2002**, *57*, 873−875.
(50) Vieths, S.; Scheurer, S.; Ballmer-Weber, B. Current understanding of cross-reactivity of food allergens and pollen. In *Genetically Engineered Foods Assessing Potential Allergenicity*; Annals of the New York Academy of Sciences; New York Academy of Sciences: New York, 2002; pp 47−68.
(51) Pieper, U.; Eswar, N.; Stuart, A. C.; Ilyin, V. A.; Sali, A. MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res.* **2002**, *30*, 255−259.
(52) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomolec. Struct.* **2000**, *29*, 291−325.
(53) Tramontano, A.; Leplae, R.; Morea, V. Analysis and assessment of comparative modeling predictions in CASP4. *Proteins* **2001**, 22−38.
(54) Fraczkiewicz, R.; Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **1998**, *19*, 319−333.
(55) Metcalfe, D. D.; Astwood, J. D.; Townsend, R.; Sampson, H. A.; Taylor, S. L.; Fuchs, R. L. Assessment of allergenic potential of foods derived from genetically engineered crop plants. *Crit. Rev. Food Sci. Nutr.* **1996**, *36*, S165−S186.
(56) WHO Evaluation of Allergenicity of Genetically Modified Foods. Report of a Joint FAO/WHO Expert Consultation. World Health Organization, 2001.
(57) Lamb, I. C. Application of the IFBC/ILSI allergenicity assessment decision tree − An industry perspective. In *Genetically Engineered Foods Assessing Potential Allergenicity*, Annals of the New York Academy of Sciences; New York Academy of Sciences: New York, 2002; pp 117−127.
(58) Zhu, H.; Schein, C. H.; Braun, W. MASIA: recognition of common patterns and properties in multiple aligned protein sequences. *Bioinformatics* **2000**, *16*, 950−951.
(59) Kawamoto, S.; Fujimura, T.; Nishida, M.; Tanaka, T.; Aki, T.; Masubuchi, M.; Hayashi, T.; Suzuki, O.; Shigeta, S.; Ono, K. Molecular cloning and characterization of a new Japanese cedar pollen allergen homologous to plant isoflavone reductase family. *Clin. Exp. Allergy* **2002**, *32*, 1064−1070.
(60) Di Felice, G.; Barletta, B.; Tinghino, R.; Pini, C. Cupressaceae pollinosis: Identification, purification, and cloning of relevant allergens. *Int. Arch. Allergy Immunol.* **2001**, *125*, 280−289.

(61) Yokoyama, M.; Miyahara, M.; Shimizu, K.; Kino, K.; Tsunoo, H. Purification, identification, and cDNA cloning of Jun a 2, the second major allergen of mountain cedar pollen. *Biochem. Biophys. Res. Commun.* **2000**, *275*, 195−202.

(62) Midoro-Horiuti, T.; Goldblum, R.; Kurosky, A.; Wood, T.; Schein, C. H.; Brooks, E. Molecular cloning of the mountain cedar (Juniperus ashei) pollen major allergen, Jun a 1. *J. Allergy Clin. Immunol.* **1999**, *104*, 613−617.

(63) Midoro-Horiuti, T.; Goldblum, R.; Kurosky, A.; Goetz, D.; Brooks, E. Isolation and characterization of the mountain cedar (Juniperus ashei) pollen major allergen, Jun a 1. *J. Allergy Clin. Immunol.* **1999**, *104*, 608−612.

(64) Midoro-Horiuti, T. M.; Brooks, E. G.; Goldblum, R. M. Pathogenesis-related proteins of plants as allergens. *Annals Allergy Asthma Immunol.* **2001**, *87*, 261−271.

(65) Murtazina, D.; Puchkaev, A. V.; Schein, C. H.; Oezguen, N.; Braun, W.; Nanavati, A.; Pikuleva, I. A. Membrane protein interactions contribute to efficient 27-hydroxylation of cholesterol by mitochondrial cytochrome P450−27a1. *J. Biol. Chem.* **2002**, *277*, 37582−37589.

(66) Zemanek, E. C.; Wasserman, B. P. Issues and Advances in the Use of Transgenic Organisms for the Production of Thaumatin, the Intensely Sweet Protein from Thaumatococcus-Danielli. *Crit. Rev. Food Sci. Nutr.* **1995**, *35*, 455−466.

(67) WHO Safety Aspects of Genetically Modified Foods of Plant Origin. Report of a Joint FAO/WHO Expert Consultation. World Health Organization, 2000.

(68) Koradi, R.; Billeter, M.; Wuthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **1996**, *14*, 51−55.

Received for review March 5, 2003. Revised manuscript received May 13, 2003. Accepted May 14, 2003. This work was supported by grants from the John Sealy Memorial Endowment Fund (#2535-01), the U.S. Food and Drug Administration (FD-U-002249-1), and the Texas Higher Education Coordinating Board ATP(004952-0060-2001).

JF034218R