BIOINFORMATICS

Vol. 18 no. 10 2002
Pages 1358–1364

Data mining of sequences and 3D structures of allergenic proteins

*Ovidiu Ivanciuc, Catherine H. Schein and Werner Braun**

Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX 77555-1157, USA

Received on January 17, 2002; revised on March 26, 2002; accepted on April 26, 2002

ABSTRACT

Motivation: Many sequences, and in some cases structures, of proteins that induce an allergic response in atopic individuals have been determined in recent years. This data indicates that allergens, regardless of source, fall into discreet protein families. Similarities in the sequence may explain clinically observed cross-reactivities between different biological triggers. However, previously available allergy databases group allergens according to their biological sources, or observed clinical cross-reactivities, without providing data about the proteins. A computer-aided data mining system is needed to compare the sequential and structural details of known allergens. This information will aid in predicting allergenic cross-responses and eventually in determining possible common characteristics of IgE recognition.

Results: The new web-based Structural Database of Allergenic Proteins (SDAP) permits the user to quickly compare the sequence and structure of allergenic proteins. Data from literature sources and previously existing lists of allergens are combined in a MySQL interactive database with a wide selection of bioinformatics applications. SDAP can be used to rapidly determine the relationship between allergens and to screen novel proteins for the presence of IgE or T-cell epitopes they may share with known allergens. Further, our novel similarity search method, based on five dimensional descriptors of amino acid properties, can be used to scan the SDAP entries with a peptide sequence. For example, when a known IgE binding epitope from shrimp tropomyosin was used as a query, the method rapidly identified a similar sequence in known shellfish and insect allergens. This prediction of cross-reactivity between allergens is consistent with clinical observations.

Availability: SDAP is available on the web at http://fermi.utmb.edu/SDAP/index.html.

Contact: werner@newton.utmb.edu.

*To whom correspondence should be addressed.

INTRODUCTION

Structural comparison of proteins that induce an allergic response in atopic individuals can be used to predict allergenic cross-responses (Ipsen and Lowenstein, 1997), and eventually to determine possible common characteristics of IgE recognition (Breiteneder and Ebner, 2000; Lascombe *et al.*, 2000; Midoro-Horiuti *et al.*, 1999; Soman *et al.*, 2000; Wellhausen *et al.*, 1996). Existing allergen databases are simple lists of information, with limited search capability. The IUIS website (http://www.allergen.org) provides a list of standardized allergen nomenclature. As the names are based solely on biological source, they do not indicate sequence and structural similarities shared by many allergens (Aalberse, 2000; Rouvinen *et al.*, 1999; Spangfort *et al.*, 1999). Other websites dedicated to allergy and asthma, such as the Allergy advisor (http://allergyadvisor.com/), suggest possible cross-reactions between foods that are based primarily on clinical observations. None of these sites provide an overview of the available information on allergenic proteins in a form useful for researchers in the field.

We have thus developed the Web tool SDAP (Structural Database of Allergenic Proteins), to integrate information about allergenic proteins with various bioinformatics tools that allow an easier overview of allergens, and where known, their epitopes. Sequence and structural data on allergens from SwissProt (Bairoch and Apweiler, 2000), PIR (Barker *et al.*, 2001), NCBI (Wheeler *et al.*, 2001), and PDB (Berman *et al.*, 2000) have been assembled in SDAP. The data is organized in tables that are maintained and cross-searched by the relational database management system MySQL. A user can search for structural neighbors of allergens or known epitopes using the built-in sequence search tools and the 3D structure comparison methods linked directly to SDAP. We show here that one can use the SDAP data lists and the integrated search tools to detect and predict cross-reactivities among common allergen triggers. This rapid search tool should be of immediate value to clinicians and researchers in allergy and asthma.

SYSTEM AND METHODS

Data selection

Our first aim is to build a systematic, searchable collection of structural and biological information about allergenic proteins. The current lists were assembled from major sequence (SwissProt, PIR and NCBI) and structure (PDB) databases, guided by the list of allergen names from the IUIS website. Where a novel allergen is described that is not yet included in the official list, the name given in the primary literature is used as long as it is consistent with the IUIS conventions. Supplemental information from the literature has been included, to both verify the entries and provide data on proteins that have not yet been incorporated in the general protein databases. As there is no openly available database summarizing information on known epitopes of allergenic proteins, the epitope list in SDAP is based solely on primary literature sources.

SDAP structure

The database component of SDAP is implemented with MySQL under Linux. The information is collected in tables according to: allergen type; species; systematic name and brief description; and sequence accession numbers from SwissProt, PIR, NCBI and, where available, the PDB file name. The data files retrieved from SwissProt, PIR, NCBI, or the PDB are stored within the main SDAP database, and can be accessed directly to provide additional information.

SDAP is an Internet service that uses a web browser to communicate with the MySQL database and various software tools, as depicted in Figure 1. The user fills an HTML form query by selecting the search domains, various key words and parameters. Subsequently, the information from the HTML form query is encoded, sent to the server that runs SDAP, and interpreted by specific CGI scripts written in C. These CGI scripts represent the interface between the web browser and the various services offered by SDAP, including the search of the MySQL database, bioinformatics tools, and links to other web databases and programs. After an MySQL database interrogation, a CGI script interprets the results and generates on-the-fly an HTML page that is sent back to the web browser that initiated the query.

Current status

In its present release, SDAP allows searches restricted to the following fields: allergen name (according to the IUIS website listing), scientific and common name for the species, origin, and description. All queries implement an approximate, case insensitive search; for example, the allergen name 'Ves v' returns results for three allergens from *Vespula vulgaris* (yellowjacket), Ves v 1, Ves v 2, and Ves v 5. Subsequently, the user can narrow the search and select a particular allergen, to obtain specific information, such as sequence and accession numbers for protein databases. For example, selecting the Ves v 5 allergen returns the SwissProt accession number VAS_VESVU and the PDB accession number 1QNX. A further selection of the SwissProt accession number will open the SwissProt file for Ves v 5, while selecting the PDB accession number returns a page with links to the 1QNX file from PDB. The user can then compare this structure to related ones using the tools provided within the PDB. Alternatively, SDAP is also directly connected 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). All links to the web servers are formatted for the specific allergen selected by the user.

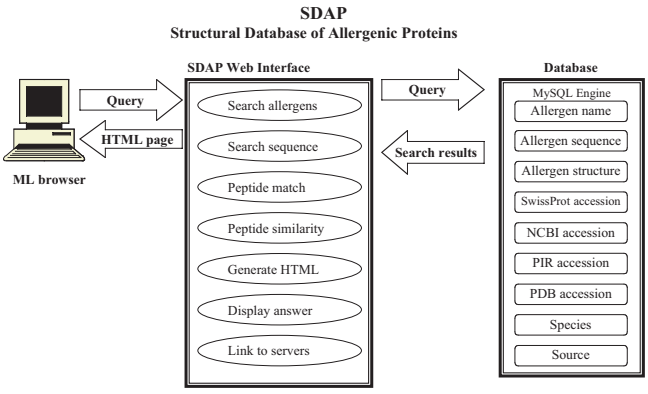

Fig. 1. Structural blocks and main functions of SDAP (see text for details).

Implemented sequence comparison tools

SDAP is not just an information storage and retrieval system. Various tools for performing structural studies of allergens or characterizing their epitopes are directly accessible from the site. SDAP is the first allergen database that allows a user to retrieve and characterize IgE epitopes of allergenic proteins. Epitope sequences, taken from the literature, are being added continuously to the database. Two SDAP tools are available for comparing the sequences of allergenic epitopes with those of all the proteins in the database: a peptide exact match and a peptide similarity search using a property distance

Table 1. SDAP search results for allergenic tropomyosin that contain the Pen i 1 IgE epitope FLAEEADRK

SwissProt accession	Allergen	Species		Sequence length	Epitope position
TPM1_METEN	Met e 1	*Metapenaeus ensis*	Shrimp	274	143–151
TPM1_PANST	Pan s 1	*Panulirus stimpsoni*	Lobster	274	143–151
TPM1_CHAFE	Cha f 1	*Charybdis feriatus*	Crab	264	153–161
TPM1_HOMAM	Hom a 1	*Homarus americanus*	American lobster	284	153–161

function developed in this group. While the first type of search will identify only sequences that match the query exactly, the second method can identify allergens that may not have significant overall sequence similarity, but contain a peptide with matching physical properties. This original procedure implements a protein sequence similarity search based on the five dimensional descriptors E_1–E_5 of amino acids properties derived from a pool of 237 physico-chemical properties (Venkatarajan and Braun, 2001). The similarity between two sequences A and B, each one consisting of N residues, is computed with the property distance function PD:

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

where λ_j is the eigenvalue of the jth E component, $E_j(A_i)$ is the E_j value for the amino acid in the ith position from sequence A, and$E_j(B_i)$ is the E_j value for the amino acid in the ith position from sequence B. Starting from a peptide from the SDAP epitope database or provided by the user, the SDAP tool calculates a similarity index between the query sequence and each sequence-window with the same length from the 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. Besides epitope identification, this tool can be used to find conserved sequences in the allergens from the SDAP protein database.

RESULTS

Predicting cross-reactivity by peptide matching

Patients sensitive to an allergen may also be sensitive to completely different food or environmental triggers. Especially in severe allergy, clinical evaluation of potential allergens through oral or skin testing may not be possible (Sicherer, 2001). Thus rapid prediction of potentially cross-reactive allergens is an essential clinical tool, as the patient can be warned to avoid related sources. Crustaceans, particularly shrimp, are among the common foods that induce hypersensitivity reactions (Leung *et al.*, 1994). Ingestion of crustaceans by shellfish-sensitive persons can produce urticaria, angioedema, laryngospasm, and even life-threatening anaphylaxis. A molecular basis for cross reactivity among crustacean species, based on the similarity of their tropomyosins, was previously suggested (Leung *et al.*, 1998). We show here how SDAP can be used to predict allergens that contain epitopes similar to those identified for the Pen i 1 protein, a major allergen of shrimp (*Penaeus indicus*;Shanti *et al.*, 1993).

First, the peptide matching tool of SDAP was used to detect all allergens containing the IgE epitope sequence FLAEEADRK. As Table 1 shows, tropomyosin allergens Met e 1 (from another shrimp species), and Pan s 1, Cha f 1, and Hom a 1 from lobster and crab also contain this sequence. These proteins are all about 90% identical to one another.

Sequence similarity search suggests a much wider range of possible cross-reactive allergens

To identify more distantly related allergens, the same epitope query was then used with the property-based, peptide similarity search. In addition to those in Table 1, the SDAP peptide similarity function detected similar sequences in about 15 entries (of the approximately 1800 sequences in the database). As Table 2 shows, tropomyosins of several insect and mollusk species (overall sequence identity is about 50–60%; (Aki *et al.*, 1995; Leung and Chu, 2001; Santos *et al.*, 1999) have similar regions with the FLAEEADRK shrimp IgE epitope. Cross-reactivity between crustacean, mollusk and insect allergens has also been confirmed by clinical tests with sera from patients sensitive to cockroach allergens (Leung and Chu, 2001; Santos *et al.*, 1999). Indeed, it has been suggested that allergy to snail consumption may originate in sensitization to house dust mites (Sicherer, 2001) and there is a report that allergen immunotherapy with European house dust mite (*Dermatophagoides pteronyssinus*) may trigger severe reactions to mollusks and Crustacea (vanRee *et al.*, 1996).

Table 2. SDAP search results for allergens that contain regions with a high similarity towards the Pen i 1 IgE epitope FLAEEADRK

NCBI accession	Allergen	Species	Seq. Len.	Position	Sequence	*PD*
TPMM_ANISI	Ani s 3	*Anisakis simplex*	284	153–161	MLAEEADRK	1.041
TPM_PERAM	Per a 7	*Periplaneta americana* American cockroach	284	153–161	FMAEEADKK	2.067
AF216519_1	Per v 1	*Perna viridis* mussel	284	153–161	WIAEEADKK	2.657
AAK96889	Cra g 1	*Crassostrea gigas*, Pacific oyster	233	102–120	LIAEEADKK	2.713
BAA04557	Der f 10	*Dermatophagoides farinae*, American house dust mite	299	168–176	MMAEDADRK	3.461
TPM_LEPDS	Lep d 10	*Lepidoglyphus destructor* storage mite	284	153–161	MMAEDADRK	3.461
TPM_DERPT	Der p 10	*Dermatophagoides pteronyssinus* European house dust mite	284	153–161	MMAEDADRK	3.461
AF216518_1	Hal d 1	*Haliotis diversicolor* abalone	284	153–161	YIAEDAERK	4.576
JE0229	Tur c 1	*Batillus cornutus*	146	49–57	YIAEDAERK	4.576
MMAL_DERPT	Der p 1	*Dermatophagoides pteronyssinus* European house dust mite	320	40–48	FEDEEAARK	6.763

The last allergen listed, Der p 1, is not a tropomyosin, but a fecal allergen from the European house dust mite with similarity to the cysteine proteases (van Oort *et al.*, 2002). This result suggests that the actual sensitizing sequence of the tropomyosins may be common to allergens from other protein families.

Statistical significance of the *PD* index

The user may determine what threshhold value of the *PD* sequence similarity index to use as a quantitative characterization of IgE epitope similarity. As an illustration of a typical distribution of *PD* values, the IgE epitope sequence from Pen i 1 was extended to include six more C-terminal residues (the N-terminal residues of this IgE epitope of Pen I 1 were not determined). Starting with FLAEEADRKYDEVAR, a set of seven peptides was generated by cutting one residue at a time from the C-terminus, until the nine-residue IgE epitope was left. For each of these sequences, the SDAP peptide similarity function was used to determine the most similar *PD* value in the SDAP database. Table 3 summarizes the maximum value $PD_{\max}$, average value PD_{ave}, standard deviation $\mathbf{SD}(PD)$, and distribution score $z(PD)$ determined for each sequence. In this experiment the minimum values $PD_{\min}$ where always zero, because an exact match was found in the database. The distribution score $z(PD)$ was computed with the equation:

$$z(PD) = \frac{|PD_{\min} - PD_{\text{ave}}|}{\mathbf{SD}(PD)} \qquad (2)$$

The statistical values presented in Table 3 vary little with the sequence length, a fact that allows us to generalize the findings obtained for specific IgE epitopes. The histogram of the *PD* values recorded for the Pen i 1 IgE epitope FLAEEADRK are presented in Figure 2. A comparison of the *PD* values from Table 2 with the *PD* distribution for this peptide in the SDAP allergen database clearly demonstrates that we have retained only hits that display a high similarity with the query sequence. The $z(PD)$-score for the sequence FLAEEADRK with the value 4.904 is a clear cutoff between the Tur c 1 tropomyosin ($z = 4.576$) and Der p 1 cysteine protease ($z = 6.763$). When a peptide is submitted to the similarity search, a histogram of the format of Figure 2 is printed out of results over the whole SDAP database, allowing the user to choose an appropriate threshold value for significance.

DISCUSSION

Allergic diseases, including rhinitis, asthma and conjunctivitis, affect up to 30% of the population of industrialized

Table 3. Statistical results for the distribution of *PD* values in the SDAP database of allergenic proteins (see text for details)

Sequence	PD_{max}	PD_{ave}	**SD**(*PD*)	*z*(*PD*)
FLAEEADRK	21.729	11.957	2.438	4.904
FLAEEADRKY	21.202	12.615	2.373	5.317
FLAEEADRKYD	22.341	12.778	2.399	5.327
FLAEEADRKYDE	21.450	13.084	2.533	5.164
FLAEEADRKYDEV	21.762	12.737	2.142	5.946
FLAEEADRKYDEVA	21.137	12.776	2.071	6.168
FLAEEADRKYDEVAR	22.549	13.047	2.112	6.176

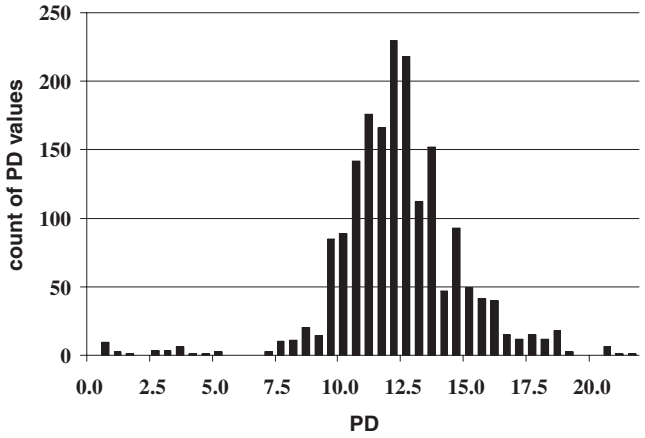

Fig. 2. Histogram of the *PD* values obtained for the Pen i 1 IgE epitope FLAEEADRK.

nations. About 40 million people in the US suffer from allergic rhinitis, which accounts for over 10 million doctor visits per year (Malone *et al.*, 1997; Platts-Mills *et al.*, 1998). Given the limited treatment options for allergy, a large amount of research in recent years has been dedicated to determining the structural aspects of proteins that trigger an allergic response. SDAP brings all this data into one easily accessible site. As we demonstrate for a major allergenic epitope from shrimp, searching in a database dedicated to allergens can suggest regions of similarity in proteins with no overall sequence similarity.

Although the sequence is known for many allergenic proteins, IgE binding areas have been determined for only a few (Banerjee *et al.*, 2000; Shin *et al.*, 1998; Soman *et al.*, 2000). The sequence searchable list of known IgE binding epitopes, assembled from the primary literature, should develop into one of the most important aspects of SDAP. Tools for comparing a sequence to this list should help in predicting possible allergenicity in novel proteins (Gendel, 1998). The user is cautioned here that the definition of an epitope differs with the assay system, and most of the known epitopes are limited to those that occur linearly in the protein sequence. Those currently in the list are known to react with IgE in the sera of atopic individuals, but not normal controls. Users can consult the accompanying references, via literature links to PubMed and Medline, for further details on the experimental validation.

The major challenge in allergy research at the current time is in looking for epitopes that are not linear in the sequence of the allergens. In order to understand and predict allergenic activity, one should know the structure of the epitopes on the protein surface. Only a few experimentally determined structures are available for allergenic proteins. As most known allergens belong to only a few sequence families, we plan to prepare high quality template based models using known structures. In the next development stage of SDAP, a database of 3D-model structures for allergens will be added. The predicted structures will be useful for visualization and computational studies to characterize epitope structure. For that reason, our major effort at the present time is in expanding the tools for structural comparison within SDAP and including high quality models, prepared with our EXDIS/DIAMOD/FANTOM suite (Soman *et al.*, 2001), for all allergens where no experimental structure has been determined.

FURTHER DEVELOPMENT

More data on allergenic proteins and their epitopes, as they become available, are being added to the lists in SDAP. Further developments will provide an expanded list of known epitopes and accompanying references, a database of homology models for allergens for which no experimental structure is available, and visualization tools for mapping epitopes on structures.

ACKNOWLEDGEMENTS

This work was supported by a Research Development Grant (#2535-01) from the John Sealy Memorial Endowment Fund for Biomedical Research.

REFERENCES

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

Aki,T., Kodama,T., Fujikawa,A., Miura,K., Shigeta,S., Wada,T., Jyo,T., Murooka,Y., Oka,S. and Ono,K. (1995) Immunochemical characterization of recombinant and native tropomyosins as a new allergen from the house-dust mite, *Dermatophagoides farinae*. *J. Allergy Clin. Immunol.*, **96**, 74–83.

Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.

Banerjee,B., Kanitpong,K., Fink,J.N., Zussman,M., Sussman,G.L., Kelly,K.J. and Kurup,V.P. (2000) Unique and shared IgE epitopes of Hev b 1 and Hev b 3 in latex allergy. *Mol. Immunol.*, **37**, 789–798.

Barker,W.C., Garavelli,J.S., Hou,Z.L., Huang,H.Z., Ledley,R.S., McGarvey,P.B., Mewes,H.W., Orcutt,B.C., Pfeiffer,F., Tsugita,A. *et al.* (2001) Protein Information Resource: a community resource for expert annotation of protein data. *Nucleic Acids Res.*, **29**, 29–32.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

Breiteneder,H. and Ebner,C. (2000) Molecular and biochemical classification of plant-derived food allergens. *J. Allergy Clin. Immunol.*, **106**, 27–36.

Conte,L.L., Ailey,B., Hubbard,T.J.P., Brenner,S.E., Murzin,A.G. and Chothia,C. (2000) SCOP: a structural classification of proteins database. *Nucleic Acids Res.*, **28**, 257–259.

Gendel,S.M. (1998) Sequence databases for assessing the potential allergenicity of proteins used in transgenic foods. *Adv. Food Nutrition Res.*, **42**, 63–92.

Gibrat,J.F., Madej,T. and Bryant,S.H. (1996) Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.*, **6**, 377–385.

Gilbert,D., Westhead,D., Nagano,N. and Thornton,J. (1999) Motif-based searching in TOPS protein topology databases. *Bioinformatics*, **15**, 317–326.

Holm,L. and Sander,C. (1996) Mapping the protein universe. *Science*, **273**, 595–602.

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

Lascombe,M., Gregoire,C., Poncet,P., Rosinski-Chupin,I., Rabillon,J., Goubran-Botros,H., Mazie,J., David,B. and Alzari,P. (2000) Crystal structure of the allergen Equ c 1: a dimeric lipocalin with restricted IgE-reactive epitopes. *J. Biol. Chem.*, **275**, 21572–21577.

Leung,P.S.C., Chu,K.H., Chow,W.K., Ansari,A., Bandea,C.I., Kwan,H.S., Nagy,S.M. and Gershwin,M.E. (1994) Cloning, expression, and primary structure of *Metapenaeus ensis* tropomyosin, the major heat-stable shrimp allergen. *J. Allergy Clin. Immunol.*, **94**, 882–890.

Leung,P.S.C., Chen,Y.C., Mykles,D.L., Chow,W.K., Li,C.P. and Chu,K.H. (1998) Molecular identification of the lobster muscle protein tropomyosin as a seafood allergen. *Mol. Mar. Biol. Biotechnol.*, **7**, 12–20.

Leung,P.S.C. and Chu,K.H. (2001) cDNA cloning and molecular identification of the major oyster allergen from the Pacific oyster *Crassostrea gigas*. *Clin. Exp. Allergy*, **31**, 1287–1294.

Malone,D.C., Lawson,K.A., Smith,D.H., Arrighi,H.M. and Battista,C. (1997) A cost of illness study of allergic rhinitis in the United States. *J. Allergy Clin. Immunol.*, **99**, 22–27.

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

van Oort,E., de Heer,P.G., van Leeuwen,W.A., Derksen,N.I.L., Muller,M., Huveneers,S., Aalberse,R.C. and van Ree,R. (2002) Maturation of *Pichia pastoris*-derived recombinant pro-Der p 1 induced by deglycosylation and by the natural cysteine protease Der p 1 from house dust mite. *Eur. J. Biochem.*, **269**, 671–679.

Pearl,F.M.G., Martin,N., Bray,J.E., Buchan,D.W.A., Harrison,A.P., Lee,D., Reeves,G.A., Shepherd,A.J., Sillitoe,I., Todd,A.E., Thornton,J.M. and Orengo,C.A. (2001) A rapid classification protocol for the CATH domain database to support structural genomics. *Nucleic Acids Res.*, **29**, 223–227.

Platts-Mills,T.A.E., Mueller,G.A. and Wheatley,L.M. (1998) Future directions for allergen immunotherapy. *J. Allergy Clin. Immunol.*, **102**, 335–343.

Rouvinen,J., Rautiainen,J., Virtanen,T., Zeiler,T., Kauppinen,J., Taivainen,A. and Mantyjarvi,R. (1999) Probing the molecular basis of allergy. Three-dimensional structure of the bovine lipcalin allergen Bos d 2. *J. Biol. Chem.*, **274**, 2337–2343.

vanRee,R., Antonicelli,L., Akkerdaas,J.H., Garritani,M.S., Aalberse,R.C. and Bonifazi,F. (1996) Possible induction of food allergy during mite immunotherapy. *Allergy*, **51**, 108–113.

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. and Arruda,L.K. (1999) 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.*, **104**, 329–337.

Shanti,K.N., Martin,B.M., Nagpal,S., Metcalfe,D.D. and Rao,P.V.S. (1993) Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes. *J. Immunol.*, **151**, 5354–5363.

Shin,D.S., Compadre,C.M., Maleki,S.J., Kopper,R.A., Sampson,H., Huang,S.K., Burks,A.W. and Bannon,G.A. (1998) Biochemical and structural analysis of the IgE binding sites on Ara h1, an abundant and highly allergenic peanut protein. *J. Biol. Chem.*, **273**, 13753–13759.

Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.

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

Soman,K., Midoro-Horiuti,T., Ferreon,J., Goldblum,R., Brooks,E., Kurosky,A., Braun,W. and Schein,C. (2000) Homology modeling and characterization of IgE epitopes of mountain cedar

allergen Jun a 3. *Biophys J.*, **79**, 1601–1609.

Soman,K.V., Schein,C.H., Zhu,H. and Braun,W. (2001) Homology modeling and simulations of nuclease structures. In Schein,C.H. (ed.), *Nuclease Methods and Protocols, Methods in Molecular Biology*, Vol. 160, Humana Press, Totowa, N.J., pp. 263–286.

Spangfort,M.D., Mirza,O., Holm,J., Larsen,J.N., Ipsen,H. and Lowenstein,H. (1999) The structure of major birch pollen allergens-epitopes, reactivity and cross-reactivity. *Allergy*, **50**, 23–26.

Venkatarajan,M.S. and Braun,W. (2001) New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical–chemical properties. *J. Mol. Model.*, **7**, 445–453.

Wellhausen,A., Schoning,B., Petersen,A. and Vieths,S. (1996) IgE binding to a new cross-reactive. *Zeit. Ernahrungswissenschaft*, **35**, 348–355.

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. and Rapp,B.A. (2001) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **29**, 11–16.