The updated Structural Database of Allergenic Proteins (SDAP 2.0) provides 3D models for allergens and incorporated bioinformatics tools

Surendra S. Negi, PhD,[a,b,c] Catherine H. Schein, PhD,[b,c] and Werner Braun, PhD[a,b,c]   *Galveston, Tex*

Background: Allergenic proteins can cause IgE-mediated adverse reactions in sensitized individuals. Although the sequences of many allergenic proteins have been identified, bioinformatics data analysis with advanced computational methods and modeling is needed to identify the basis for IgE binding and cross-reactivity.

Objective: We aim to present the features and use of the updated Structural Database of Allergenic Proteins 2.0 (SDAP 2.0) webserver, a unique, publicly available resource to compare allergens using specially designed computational tools and new high-quality 3-D models for most known allergens.

Methods: Previously developed and novel software tools for identifying cross-reactive allergens using sequence and structure similarity are implemented in SDAP 2.0.

A comprehensive set of high-quality 3-D models of most allergens was generated with the state-of-the-art AlphaFold 2 software. A graphics tool enables the interactive visualization of IgE epitopes on experimentally determined and modeled 3-D structures.

Results: A user can search for allergens similar to a given input sequence with the FASTA algorithm or the window-based World Health Organization/International Union of Immunological Societies (WHO/IUIS) guidelines on safety concerns of novel food products. Peptides similar to known IgE epitopes can be identified with the property distance tool and conformational epitopes by the Cross-React method. The updated database contains 1657 manually curated sequences including all allergens from the IUIS database, 334 experimentally determined X-ray or NMR structures, and 1565 3-D models. Each allergen/isoallergen is classified according to its protein family.

Conclusions: SDAP provides access to the steadily increasing information on allergenic structures and epitopes with integrated bioinformatics tools to identify and analyze their similarities. In addition to serving the research and regulatory community, it provides clinicians with tools to identify potential coallergies in a sensitive patient and can help companies to design hypoallergenic foods and immunotherapies. (J Allergy Clin Immunol Global 2023;2:100162.)

Key words: Protein evaluation for allergenicity, database of allergens, IgE binding, cross-reactivity, physical-chemical property distance (PD)

Proteins from different foods, plants, fungi, and animals can sensitize individuals and stimulate IgE-mediated allergic responses.[1] After binding to IgE antibodies on mast cells,[2] allergens can trigger an immune response ranging from rashes to difficulty breathing to potentially deadly anaphylaxis.[3] It is known that allergens from many different sources can cause dangerous cross-reactions, which may be revealed by sequence and structural relationships.[4,5] The Structural Database of Allergenic Proteins (SDAP), the first publicly available, cross-referenced database of allergenic proteins,[6] allowed users to rapidly determine relationships between allergens or to any input sequence using specifically designed bioinformatic tools.[7,8] Other websites (recently summarized[9]) with different goals complement this information.

SDAP 2.0 is a comprehensive update that includes sequences and structures for all proteins classified by the Allergen Nomenclature Sub-Committee of the World Health Organization and International Union of Immunological Societies (WHO/IUIS) according to general clinical and biochemical guidelines as allergens.[10] The WHO/IUIS database lists proteins that conform to their clinical criteria for allergenicity with respect to the number of patients with documented IgE reactions. In addition to sequences from the WHO/IUIS database (which are marked as such), additional proteins described as having allergenic characteristics from the literature are included for comparison purposes. Bioinformatic tools developed by us and other groups, functioning "on the fly," allow rapid comparison of allergens according to protein families (Pfams),[11] physicochemical property motifs,[12] physical relationships, and biological functions.[4,6,13-20] The tools now integrated into SDAP 2.0 and, most significantly, 3-dimensional (3D) structures of more than 1600 distinct allergens and their isoforms are a valuable resource for identifying cross-reactive allergens. Recently, a major advance in determining a conformational IgE epitope of the house dust mite allergen Der p 2 in a high-resolution X-ray crystal structure was achieved,[21] which suggests that experimental data as input for our bioinformatics tools for predicting cross-reactivities will be available in increasing numbers.

SDAP was first publicly available in 2002,[7,22] with a major update in 2008[23] that included 3-D models of allergens for which no experimental structure was available.[24] New or updated features in SDAP 2.0 are the interactive graphics tool for the display of experimentally determined 3-D structures, an updated list of

---

Abbreviations used
3-D: 3-dimensional
PD: Property distance
Pfams: Protein families
pLDDT: Predicted local distance difference test
PR: Pathogenesis-related
SDAP: Structural Database of Allergenic Proteins
WHO/IUIS: World Health Organization and International Union of Immunological Societies

---

allergenic proteins classified by Pfam, a version of the previously validated Cross-React software[14] integrated into the webserver, and a comprehensive set of highly reliable 3-D models of most allergen sequences generated with the highly precise and validated AlphaFold program.[25] These models and the updated software tools now available in the webserver SDAP 2.0, can help in interpreting the experimental results and determining structural similarities in cross-reactive proteins even in the face of extremely variant sequences.[5,26,27] We show the capabilities of the webserver with the identification of food allergens cross-reactive to the birch pollen allergen, Bet v 1, a linear peptide search in the database for an IgE epitope of the *Juniper ashei* pollen allergen, Jun a 1, with the PD tool, which identified other pectate lyases from Cupressaceae tree pollen, but not other pectate lyases from the Asterales plant family of mugwort and ragweed.[28,29] New in SDAP 2.0, the integrated Cross-React software can be used to identify other potential conformational IgE epitopes, as illustrated by the conservation of a previously identified IgE epitope in other pathogenesis-related (PR)-5 thaumatin-like proteins.[30,31] Finally, we document the reliability of the new 3-D models in SDAP 2.0 generated by AlphaFold.

METHODS
Data collection

All allergen sequences available in SDAP were manually identified using existing databases. The curated information is stored in MySQL tables according to a unique allergen and sequence ID linked to major databases, SWISS-PROT and NCBI. Any sequence annotated by WHO/IUIS as an allergen is labeled as "1" in SDAP, otherwise labeled a "0." In addition, existing scientific literature related to any allergen sequence is added and linked to the NCBI PubMed database. Epitope information collected from the literature search is also available in SDAP.

Sequence search

Internal search utilities in SDAP are implemented with MySQL. For example, users can search allergen name, species, and source. Basic sequence analysis within SDAP is performed using the FASTA sequence alignment program.[32] In addition, allergen sequences from SDAP can be directly searched against protein sequences in the NR and PDB databases at the NCBI using a BLAST search[33] implemented in SDAP.

Peptide search

Allergenic proteins are known to be cross-reactive if they have high sequence or structural similarities.[34,35] Therefore, a peptide search utility is implemented in SDAP to find similar sequences in other allergenic proteins. Users can compare a given peptide sequence against all allergenic protein sequences in SDAP using either an exact sequence match or peptide similarity with a property distance (PD) search.[7,36,37] The PD search method has been described earlier,[8,37-40] and is very useful to find small peptide sequences that share common amino acid physicochemical properties.[38,41-43]

3-D model structures

We used the AlphaFold 2 software,[25,44] running locally on a Linux computer, to generate 3-D model structures of all allergenic proteins in SDAP 2.0. AlphaFold generates a high-quality 3-D structure of a protein, when the sequence similarity between the target sequence and template structures varies from moderate to very high. The quality of the modeled structures is measured in terms of predicted local distance difference test (pLDDT) score, which is a reliable indicator of the quality of the models.[45,46] All model structures of the allergenic proteins are shown in the SDAP website interactively with the user-friendly 3DJmol molecule viewer plugin,[47] and are freely available to download with proper citation of their SDAP-2 origin. In addition, epitope information, if available, can be visualized on the 3-D structure of a protein.

Pfam classification

Despite the rapid increase in the number of sequences, most allergenic proteins belong to a small set of Pfams.[48,49] Using the Pfamscan method,[50] all allergenic proteins in SDAP are classified on the basis of their Pfam nomenclature,[51] and users can search allergens in SDAP based on their Pfam name.

Food and Allergy Organization/WHO guidelines

We have implemented the Food and Allergy Organization/WHO guidelines for allergenic proteins as a web server. Users can search any peptide or a protein sequence against all allergenic proteins in SDAP. According to the FAO/WHO guidelines, a query protein and a known allergen should be considered potentially cross-reactive if there is (a) more than 35% identity in the amino acid sequence of the expressed protein, using a window of 80 amino acids and a suitable gap penalty, or (b) a sequence identity of 6 contiguous amino acids. Because condition (b) is rather liberal and generates potentially many false positives, the user may choose another cutoff value for the length of the exact sequence match.

RESULTS
Main functionalities in SDAP 2.0

An overview of the integrated bioinformatics tools for storing and accessing the allergenic proteins in SDAP 2.0 is shown in Fig 1. Buttons, clearly visible to the user, allow access to the allergen data entries and search tools according to keywords, sequence similarity searches, peptide similarities by PDs,[37] allergenic risks factors according to the WHO/IUIS rules,[1] and graphical display of experimentally determined and 3-D model structures. In addition, existing scientific literature related to any allergen sequence is added and linked to the NCBI PubMed database. Epitope information from periodically updated
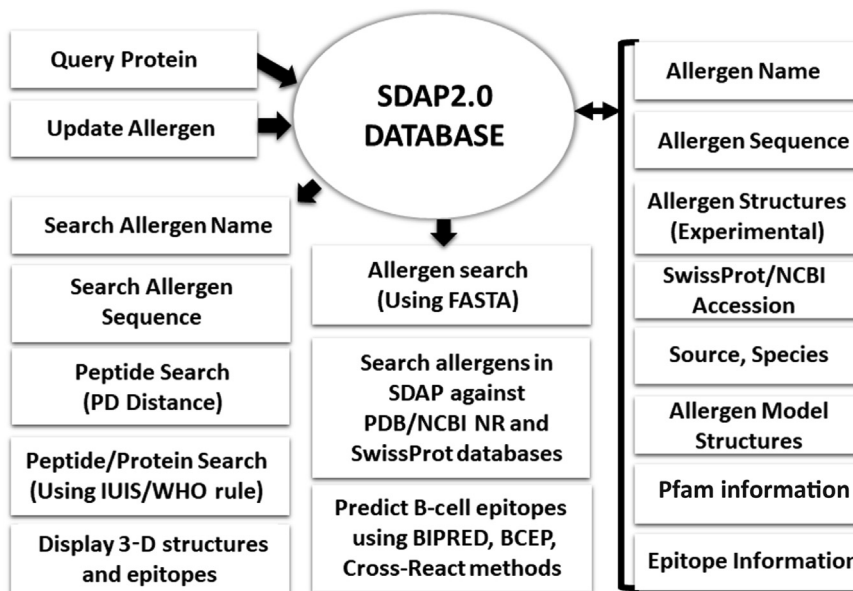
FIG 1. Main functions of SDAP and internal data storage of allergens: Left and middle columns contain the list of integrated search tools, display functions, and bioinformatics tool for epitope predictions; the right column lists the stored allergen information in the database.

literature searches is available in the SDAP database. All sequences are annotated on the basis of their Pfam 35.0 database (November 2021 release) domains. We obtained 223 protein families of allergen proteins compared with 19,632 different families in the Pfam database.

Sequence and 3-D structural similarity searches in the database

FASTA search. Allergenic proteins are considered to be potentially cross-reactive if they have high sequence or structural similarities.[34,35] Related allergens can be identified in SDAP 2.0 by a sequence search. For example, a FASTA search identified many food allergens in SDAP 2.0 with high sequence similarity to the birch pollen allergen Bet v 1 (Table I), many of which have been implicated in the pollen food syndrome.[52,53] These sequence-related allergens are PR-10 proteins, activated in plants in response to stress. Their models also show 3-D structural similarity to Bet v 1, with a 7-stranded β-sheet wrapped around an α-helix (Fig 2). Most hazelnut-allergic patients in a Dutch study[54] were sensitized to Bet v 1. Those with coallergy to peanut and high IgE levels to hazelnut Cor a 1 also had similar IgE levels to the peanut allergen Ara h 8 and pollen allergen Bet v 1. Cor a 1 and Ara h 8 sensitization was always accompanied by a Bet v 1 sensitization. The observed high correlation between IgE responses to Cor a 1 and Ara h 8 in the patient cohort, as well as the close correlation to IgE responses to Bet v 1, strongly suggest the cross-reactivity of these proteins despite their vastly different sources.

PD tool. When the IgE epitopes of an allergen have been experimentally determined, potential cross-reactivity to other allergens can be found by a peptide similarity search, using the (implemented) property distance (PD) tool. The PD tool calculates how 2 peptides differ based on their quantitative value of 5

TABLE I. List of food allergens obtained from a FASTA search in SDAP 2.0 starting with the pollen allergen Bet v 1

Allergen name*	Source	Percentage of amino acid sequence identity with Bet v 1
Cor a 1	Hazelnut	84
Mal d 1	Apple	68
Cas s 1	Chestnut	67
Jug r 5	English Walnut	67
Pru av 1	Sweet Cherry	61
Pru p 1	Peach	61
Pru ar 1	Apricot	61
Pru du 1	Almond	57
Pyr c 1	Pear	60
Fra a 1	Strawberry	57
Gly m 4	Soybean	51
Ara h 8	Peanut	47

*These allergens share high sequence and structural similarity with Bet v 1 but are from different sources known to induce pollen-food syndrome.

quantitative descriptors representing physical-chemical properties for each amino acid residue in each peptide. The PD value is 0 for identical sequences and assumes values up to 5-8, depending on length, for cross-reactive peptides. The PD search method has been validated,[8,37-40] where it was demonstrated to find peptide sequences of cross-reactive allergens that share common properties.[38,41,42]

In Fig 3 we give an example of a PD search in SDAP 2.0 for the pollen allergen Jun a 1. Using the Jun a 1 linear epitope 1,[55,56] we found several pectate lyase allergens from other Cupressaceae species: common cypress (Cup s 1), Japanese cedar (Cry j 1), and Japanese cypress (Cha o 1). These allergens each have sequences with a low PD score (<3.0) to the input IgE-binding peptide of Jun a 1. Consistent with this result, pollen from all 3 of
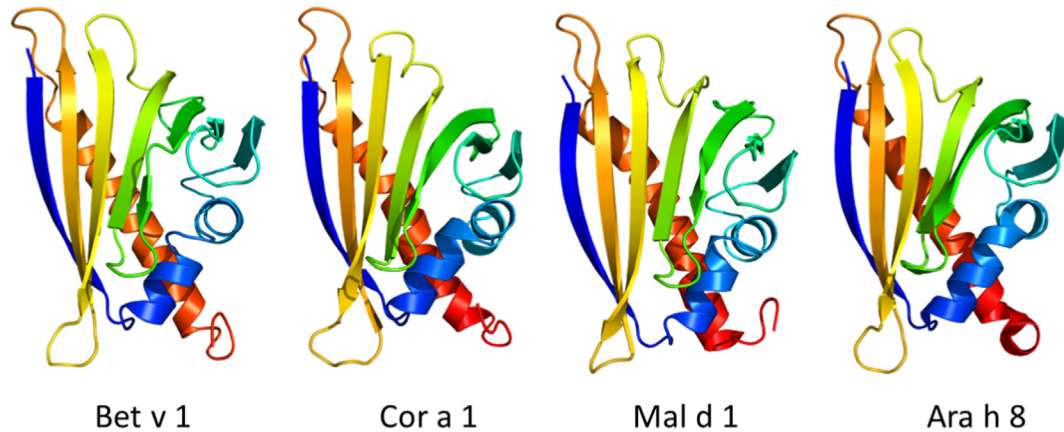
FIG 2. 3-D structures of cross-reacting food allergens to Bet v 1, Cor a 1 (hazelnut, PDB id: 6Y3I), Mal d 1 (apple, PDB id: 5MMU), and Ara h 8 (peanut, PDB id: 4M9B) were found by a FASTA search in SDAP-2 starting with the probable sensitizing allergen from birch pollen, Bet v 1 (PDB id: 1BV1). These allergens all group to the same Pfam, despite their diverse sources.

A

No.	Allergen name	Allergen species	PD value	Z(PD,min)	Z(PD,all)	Matching region
1	Jun a 1	*Juniperus ashei*	0.00	7.10	8.47	IFSQNMNIKLKMP
2	Jun v 1	*Juniperus virginiana*	0.00	7.10	8.47	IFSQNMNIKLKMP
3	Jun o 1	*Juniperus oxycedrus*	0.00	7.10	8.47	IFSQNMNIKLKMP
4	Cup s 1	*Cupressus sempervirens*	0.00	7.10	8.47	IFSQNMNIKLKMP
5	Cup a 1	*Cupressus arizonica*	0.66	6.71	8.16	IFSQNMNIKLQMP
6	Cry j 1	*Cryptomeria japonica*	1.65	6.71	7.70	IFSGNMNIKLKMP
7	Cha o 1	*Chamaecyparis obtusa*	2.57	5.61	7.27	IFSKNLNIKLNMP

B **C** **D** **E**
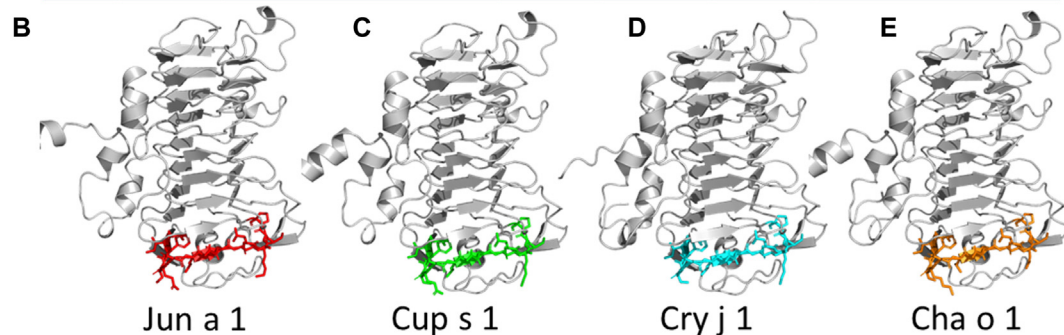

Jun a 1   Cup s 1   Cry j 1   Cha o 1

FIG 3. Peptide search with the PD tool implemented in SDAP 2.0, starting with the experimentally determined, linear IgE epitope 1 (IFSQNMNIKLKMP) of Jun a 1. **A,** List of allergens with similar peptides as indicated by their low PD value. A more complete list of the top 20 allergens in SDAP 2.0 found by this search is given in Table E1. **B-E,** The common epitope areas mapped on the 3-D structures of the similar pectate lyase allergens from the Cupressaceae species Mountain cedar (Jun a 1), common cypress (Cup s 1), Japanese cedar (Cry j 1), and Japanese cypress (Cha o 1). IgE from both Texas and Japanese patients bind to the same regions of Jun a 1.

these Cupressaceae trees cross-react with IgE pollen-allergic patient sera in ELISA experiments.[28] The PD scale clearly distinguished the pollen pectate lyases from Asteraceae species that have only limited cross-reactivity to Jun a 1, including Art v 6 from mugwort, Amb a 1 from ragweed, or Hel a 6 from sunflower that had PD values more than 7.0. In Table E1 (in the Online Repository available at www.jaci-global.org) we give a list of the top 20 allergens in the database that shows this drastic increase in PD value from Cupressaceae allergens to the Asteraceae pollen allergen Art v 6, illustrating how different the grass and tree pollen proteins are. Other allergens listed in Table E1 have high PD values, such as Chi t 6, Chi t 8, Bomb m 5, or Fag e 1, indicating their low degree of similarity to the query allergen. This indicates the power of the PD approach in distinguishing possible cross-reactive allergens using experimental IgE epitope data.

Cross-React. Our conformational epitope search method, Cross-React,[14] has now been implemented in SDAP 2.0. This tool identifies exposed surface patches on the SDAP proteins that best
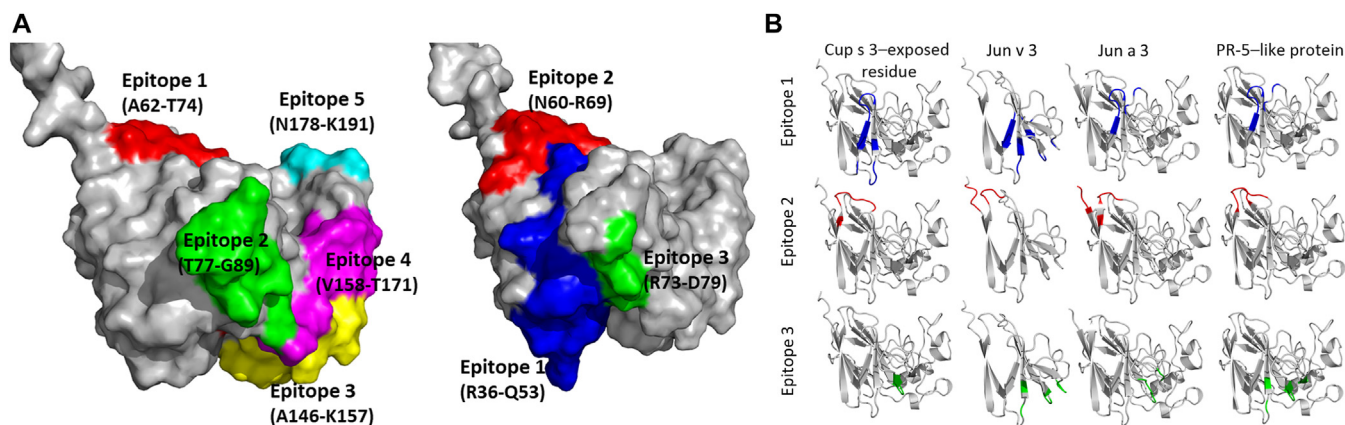
FIG 4. **A,** Linear epitopes of Jun a 3 (*left*) and Cup s 3 (*right*) mapped on the 3-D structures of the PR-5 allergens. **B,** Cross-React search in SDAP 2.0 with the linear epitopes of Cup s 3. We found similar surface-exposed patches on the pollen allergens Jun v 3, Jun a 3, and a thaumatin-like protein homolog from *Cupressus arizonica* with a high Pearson correlation value. N-terminals of proteins are not shown (from M1-L19 in Jun a 1, and M1-A16 in Cup s 3) because they are part of the signal peptides and not part of the mature protein. Epitope location is labeled on the basis of the original publication related to Jun a 1 and Cup s 3.

map to known conformational epitopes. For existing linear epitopes in SDAP, we calculated the solvent-accessible surface areas of the amino acids in the epitope region of the 3D structure using GETAREA.[57] The Cross-React program selects these surface-exposed residues and maps them on to the protein surfaces of all allergens in SDAP 2.0. Similar conformational epitopes are then found by a high correlation coefficient.[14] A high Cross-React correlation score identified the allergens of apple (Mal d 1), cherry (Pru av 1), carrots (Dau c 1), and celery (Api g 1) as being cross-reactive with the birch pollen allergen Bet v 1, confirming previous data.[14]

In another example of how structure can identify cross-reactive epitopes, our group previously identified a novel allergen in Juniper pollen, Jun a 3, and characterized a major area on its surface for IgE binding.[31] Five IgE-reactive tryptic fragments, separated by HPLC, were mapped on a model based on the nearest structural template in the PDB, for the sweet-tasting protein thaumatin. Subsequently, similar PR-5 homologs of Jun a 3 were identified as allergens in peach, cherry, bell pepper, and apple and could be related to observed pollen-related food sensitivities (oral allergy syndrome).[58] In Fig 4, *A*, we show the location of the IgE-reactive tryptic fragments on Jun a 3 and compare them, structurally, to recently detected epitopes in another PR-5 allergen, Cup s 3 from common cypress (*Cupressus sempervirens*).[30] Cup s 3 and Jun a 3 are known to be cross-reactive from earlier studies by ELISA inhibition assays.[38] The epitope 1 from Jun a 3 and epitope 2 of Cup s 3, and epitope 2 from Jun a3 and epitope 3 from Cup s 3, overlap despite the different methods used to identify them. The Cross-React program identified other structurally related proteins in SDAP with common conformational epitopes. Cross-React found 3 structurally similar proteins related to Cup s 3: Jun v 3, Jun a 3, and a thaumatin-like protein homolog from *Cupressus arizonica* (a non-WHO/IUIS allergen protein in SDAP).[59,60] All WHO/IUIS allergens are labeled as 1 in SDAP or 0 if they are not included in this list. These allergens share a high sequence identity with Cup s 3 and have a high correlation coefficient calculated by Cross-React (Table II) of 0.84 or higher. We suggest that the exposed surface area

overlapping with epitope 1 from Jun a 3 and epitope1/epitope 2 from Cup s 3 is a conserved conformational epitope among the PR-5 thaumatin-like allergens. The banana allergen Mus a 4, also a PR-5-TLP, also has a high correlation coefficient, but a lower sequence identity. Earlier structural studies of Mus a 4[61] detected a similar IgE-binding groove of the banana allergen and the pollen allergen Jun a 3, and suggested that this similarity could be the molecular basis for the cross-reactivity between aeroallergens and fruit allergens from the TLP family.

Accuracy of our 3-D models of allergenic proteins in SDAP 2.0

We used the AlphaFold 2 software[25,44] running locally on a Linux computer to generate 3-D model structures of all allergenic proteins in SDAP. The qualities of the modeled structures are measured in terms of pLDDT scores.[45] Based on the CASP experiment,[46] an AlphaFold model is considered reliable if the pLDDT score is above 70. Fig 5, *A*, shows the average pLDDT scores compared with the sequence identities of the top template structures obtained by a BLAST sequence search against the protein data bank. This indicates that AlphaFold can generate high-quality structures even for difficult targets with low sequence identities to potential template structures. Most of our 3-D model structures have very high average pLDDT scores (Fig 5, *B*), with 1283 modeled structures above 80%, 1404 structures above 70%, and only 120 allergen structures with average pLDDT score values less than 70%.

Fig 6 shows an example of a model structure of the major peanut allergen Ara h 2 obtained from the AlphaFold method. Although multiple IgE epitopes have been identified on the basis of peptide mapping,[62-65] Ara h 2 resisted experimental structure determination for years, due to high mobility especially in the flexible loop region, which is one of its most important IgE epitopes. The partial X-ray crystal structure of Ara h 2 was solved in 2011, but major disordered loop regions are missing from the coordinate list in the PDB.[66] The loop region with correct stereochemistry is included in the SDAP-2 model, permitting

TABLE II. Mapping of linear epitopes using the Cross-React* search

Allergen name	Predicted residues in the patch (epitope 1)	PCC	Sequence identity (%)
Jun v 3	T18V19W20G28K29R30G52T54G55C66L75	0.84	95.90
Jun a 3	L44P45G46G47G48K49S96T98	0.81	95.10
PR-5–like protein	L18P19G20G21G22K23S70T72	0.81	98.00
Mus a 4	W14G20G21G22R23W31L69S70	0.82	55.70

Allergen name	Predicted residues in the patch (epitope 2)	PCC	Sequence identity (%)
Jun v 3	G3A4G5A6A42A43G44T45	0.94	95.90
Jun a 3	A24G25V26N60A62A63G64T65	0.93	95.10
PR-5–like protein	L35A36A37G38T39A40Q86S87	0.88	98.00
Mus a 4	A1T2N34A37G190G191N193	0.91	55.70

Allergen name	Predicted residues in the patch (epitope 3)	PCC	Sequence identity (%)
Jun v 3	T18W20T54G55T57C66Q67T68	0.87	95.90
Jun a 3	G75C76T77F78D79G84S85	0.89	95.10
PR-5–like protein	G49C50T51F52G58S59	0.92	98.00
Mus a 4	T12W14T48G49C50S51Q61G137	0.84	55.70

*The Cross-React method uses all surface-exposed residues in the linear peptide and locates surface patches in all 3-D structures of allergens in the SDAP 2.0 database with a similar amino acid composition. Using the 3 linear epitopes of Cup s 3 (RYTVWAAGLPGGGKRLDQ, NLAAGTASAR, and RTGCTFD) we found Jun v 3, Jun a 3, and a thaumatin-like protein homolog from *Cupressus arizonica* (labeled as PR-5–like protein in the table) and Mus a 4 allergens as potential cross-reactive allergens. All amino acid labels are based on the sequence information in the database.
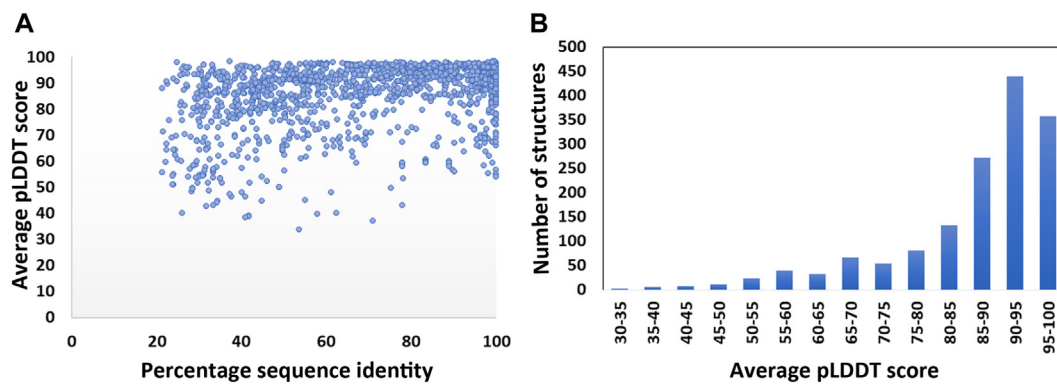

FIG 5. **A,** Average values of pLDDT scores of 3-D model structures obtained with the AlphaFold program vs the sequence identities of the top template structures in the protein data bank using a BLAST search. **B,** Histogram of the pLDDT scores. Most of the 3-D models in SDAP 2.0 have average pLDDT scores greater than 70, indicating they are highly reliable 3-D models.

visualization of this area and rational interpretation of clinical data. The model can be used as a starting structure in molecular dynamics calculations, allowing interpretation of mutations affecting IgE binding that are consistent with experimental data (unpublished data). In the current version of SDAP 2.0, we have also provided snapshots of the sequence alignment used to generate each 3-D model structure and its pLDDT score profile plot, allowing easily identification of the reliability of individual segments.

DISCUSSION

Allergenic diseases including asthma, allergic rhinitis (also known as hay fever), eczema, hives, and food allergies affect more than 100 million people worldwide. The societal and economic costs of food allergy mentioned in a recent publication[67] vary considerably depending on the methods and parameters used and the different geographical locations. However, the studies indicate that the costs are high, exceeding several billions of dollars for individual countries. For example, a large fraction of

these costs, the direct medical costs of food allergy for the population in the United States, was estimated to be $4.3 billion.

There is a significant rise in the world population suffering from allergies according to a statistic from the American Academy of Allergy, Asthma & Immunology, which estimated that more than 50 million people alone in the United States have some form of allergies. In addition, allergenic proteins are known to be cross-reactive, increasing the burden to the health care system. Data on allergenic proteins, available from multiple sources, have rapidly expanded with more efficient protein sequencing and experimental data available from clinical trials.[68,69] A summary of some of the widely used allergen databases[70] is shown in Table E2 (in the Online Repository available at www.jaci-global.org). The specific advantages of SDAP 2.0 are the new 3-D models of all allergens and bioinformatics tools, such as the PD tool and Cross-React, which can be used to find related allergens in the database and to generate hypotheses on potential cross-reactivities. For example, we illustrated how the PD tool can be used to identify potential cross-reactive allergens to the pollen allergen Jun a 1. We showed
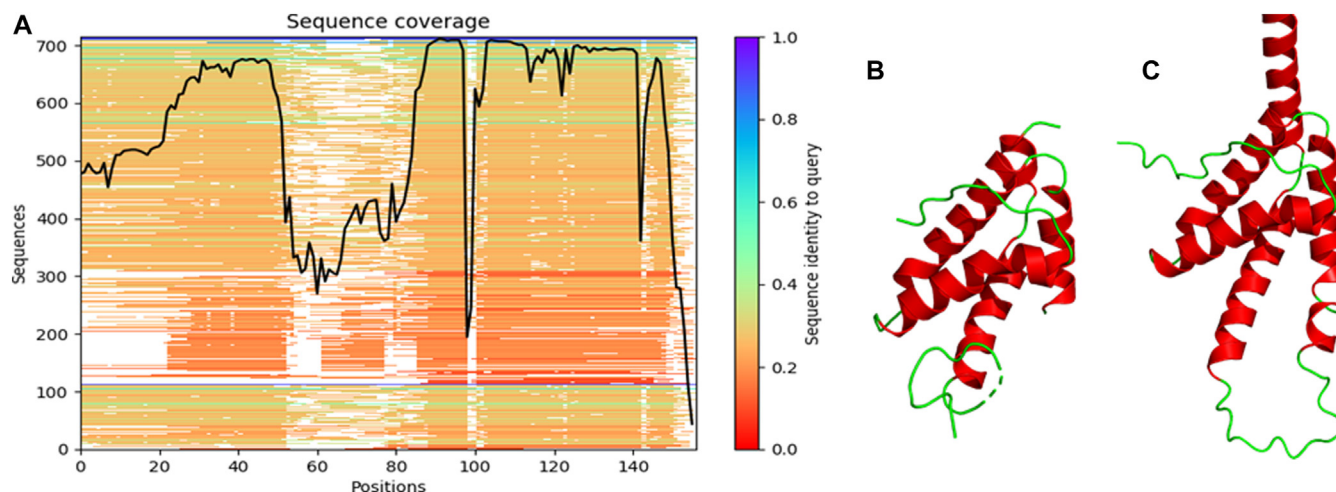
FIG 6. Quality profile and experimental and 3-D model of the peanut allergen Ara h 2. **A,** Sequence coverage of Ara h 2 obtained from a database search. Sequence coverage plot (or head-map representation of the multiple sequence alignment) shows the number of homologous sequences identified. These are colored according to sequence identity with Ara h 2. Red color shows low sequence identity, while yellow to blue color shows moderate to high sequence identity. The gap (white space) in the alignment shows regions not covered, and bold black line shows the relative coverage of the query sequence with respect to the total number of aligned sequences. **B,** X-ray-crystal structure of Ara h 2 in the Protein Data Bank. **C,** 3-D model of Ara h 2 obtained from AlphaFold 2. Although the X-ray crystal structure has low sequence coverage in the disordered loop region (residue number ~45-85), AlphaFold builds a good quality complete model of this area, allowing visualization of IgE epitopes, 59-64 RDPYSP and 65-72 SQDPYSPD. Quality of the predicted model is measured by the pLDDT score. A model structure with pLDDT score greater than 70 is modeled well, whereas those with less than 70 need more attention.

that several pectate lyase allergens from the common cypress (Cup s 1), Japanese cedar (Cry j 1), and Japanese cypress (Cha o 1) have a low PD value to Jun a 1, which is consistent with experimental observations from ELISA experiments. We also illustrated the usefulness of the 3-D structure to find potential conformational IgE epitopes indicating cross-reactivity by Cross-React. The Cross-React program identified 3 structurally similar proteins related to the allergen Cup s 3: Jun v 3, Jun a 3, and a thaumatin-like protein homolog from *Cupressus arizonica*. Cup s 3 shares sequence identity with these allergens and has a high correlation coefficient, calculated by Cross-React. We also showed that these findings from Cross-React can suggest potential conserved conformational epitopes among the PR-5 thaumatin-like allergens.

Conclusions

Our webserver SDAP 2.0 provides previously developed and novel bioinformatics tools for allergen researchers and regulatory agencies to assess the allergenic risk of novel proteins. The updated sequence and 3-D structure information allows rapid comparison of all allergens in the WHO/IUIS Allergen Nomenclature database, and many others described in the literature. SDAP 2.0 will be updated on a regular basis, especially in response to user comments and corrections. More practical examples how to use SDAP 2.0 can be found in a recent publication.[71]

DISCLOSURE STATEMENT

Key messages

- **SDAP 2.0 is a novel multifunctional webserver for allergen researchers and regulatory agencies.**
- **The SDAP 2.0 database contains more than 1600 allergenic proteins, including all those validated by the WHO/IUIS Allergen Nomenclature Sub-Committee.**
- **Novel resources include an updated list of experimentally determined 3-D structures and a complete set of high-quality 3-D models of allergenic proteins.**
- **Novel sequence and structure-based search tools are implemented in the web server.**

REFERENCES

1. Pomes A, Davies JM, Gadermaier G, Hilger C, Holzhauser T, Lidholm J, et al. WHO/IUIS Allergen Nomenclature: providing a common language. Mol Immunol 2018;100:3-13.
2. Kolkhir P, Elieh-Ali-Komi D, Metz M, Siebenhaar F, Maurer M. Understanding human mast cells: lesson from therapies for allergic and non-allergic diseases. Nat Rev Immunol 2022;22:294-308.
3. Ring J. History of allergy: clinical descriptions, pathophysiology, and treatment. Handb Exp Pharmacol 2022;268:3-19.

4. Dreskin SC, Koppelman SJ, Andorf S, Nadeau KC, Kalra A, Braun W, et al. The importance of the 2S albumins for allergenicity and cross-reactivity of peanuts, tree nuts, and sesame seeds. J Allergy Clin Immunol 2021;147:1154-63.

5. Nesbit JB, Schein CH, Braun BA, Gipson SAY, Cheng H, Hurlburt BK, et al. Epitopes with similar physicochemical properties contribute to cross reactivity between peanut and tree nuts. Mol Immunol 2020;122:223-31.

6. Schein CH, Negi SS, Braun W. Still SDAPing along: 20 years of the Structural Database of Allergenic Proteins. Front Allergy 2022;3:863172.

7. Ivanciuc O, Schein CH, Braun W. Data mining of sequences and 3D structures of allergenic proteins. Bioinformatics 2002;18:1358-64.

8. Ivanciuc O, Schein CH, Garcia T, Oezguen N, Negi SS, Braun W. Structural analysis of linear and conformational epitopes of allergens. Regul Toxicol Pharmacol 2009;54:S11-9.

9. van Ree R, Sapiter Ballerda D, Berin MC, Beuf L, Chang A, Gadermaier G, et al. The COMPARE Database: a public resource for allergen identification, adapted for continuous improvement. Front Allergy 2021;2:700533.

10. Sudharson S, Kalic T, Hafner C, Breiteneder H. Newly defined allergens in the WHO/IUIS Allergen Nomenclature Database during 01/2019-03/2021. Allergy 2021;76:3359-73.

11. Schein CH, Ivanciuc O, Midoro-Horiuti T, Goldblum RM, Braun W. An allergen portrait gallery: representative structures and an overview of IgE binding surfaces. Bioinform Biol Insights 2010;4:113-25.

12. Lu W, Negi SS, Schein CH, Maleki SJ, Hurlburt BK, Braun W. Distinguishing allergens from non-allergenic homologues using Physical-Chemical Property (PCP) motifs. Mol Immunol 2018;99:1-8.

13. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The immune epitope database and analysis resource: from vision to blueprint. PLoS Biol 2005;3:e91.

14. Negi SS, Braun W. Cross-React: a new structural bioinformatics method for predicting allergen cross-reactivity. Bioinformatics 2017;33:1014-20.

15. Fiers MW, Kleter GA, Nijland H, Peijnenburg AA, Nap JP, van Ham RC. Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC Bioinformatics 2004;5:133.

16. Mittag D, Batori V, Neudecker P, Wiche R, Friis EP, Ballmer-Weber BK, et al. A novel approach for investigation of specific and cross-reactive IgE epitopes on Bet v 1 and homologous food allergens in individual patients. Mol Immunol 2006;43: 268-78.

17. Sharma N, Patiyal S, Dhall A, Pande A, Arora C, Raghava GPS. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. Brief Bioinform 2021;22.

18. Maurer-Stroh S, Krutz NL, Kern PS, Gunalan V, Nguyen MN, Limviphuvadh V, et al. AllerCatPro—prediction of protein allergenicity potential from the protein sequence. Bioinformatics 2019;35:3020-7.

19. Dimitrov I, Bangov I, Flower DR, Doytchinova I, AllerTOP v. 2–a server for in silico prediction of allergens. J Mol Model 2014;20:2278.

20. Negi SS, Braun W. Automated detection of conformational epitopes using phage display peptide sequences. Bioinform Biol Insights 2009;3:71-81.

21. Smith SA, Chruszcz M, Chapman MD, Pomes A. Human monoclonal IgE antibodies—a major milestone in allergy. Curr Allergy Asthma Rep 2023;23:53-65.

22. Ivanciuc O, Schein CH, Braun W. SDAP: database and computational tools for allergenic proteins. Nucleic Acids Res 2003;31:359-62.

23. Oezguen N, Zhou B, Negi SS, Ivanciuc O, Schein CH, Labesse G, et al. Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes. Mol Immunol 2008;45:3740-7.

24. Power TD, Ivanciuc O, Schein CH, Braun W. Assessment of 3D models for allergen research. Proteins 2013;81:545-54.

25. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583-9.

26. Nesbit JB, Hurlburt BK, Schein CH, Cheng H, Wei H, Maleki SJ. Ara h 1 structure is retained after roasting and is important for enhanced binding to IgE. Mol Nutr Food Res 2012;56:1739-47.

27. Maleki SJ, Teuber SS, Cheng H, Chen D, Comstock SS, Ruan S, et al. Computationally predicted IgE epitopes of walnut allergens contribute to cross-reactivity with peanuts. Allergy 2011;66:1522-9.

28. Barre A, Sénéchal H, Nguyen C, Granier C, Rougé P, Poncet P. Identification of potential IgE-binding epitopes contributing to the cross-reactivity of the major cupressaceae pectate-lyase pollen allergens (group 1). Allergies 2022;2:106-18.

29. Pichler U, Hauser M, Wolf M, Bernardi ML, Gadermaier G, Weiss R, et al. Pectate lyase pollen allergens: sensitization profiles and cross-reactivity pattern. PLoS One 2015;10:e0120038.

30. Barre A, Sénéchal H, Nguyen C, Granier C, Poncet P, Rougé P. Structural basis for the IgE-binding cross-reacting epitopic peptides of Cup s 3, a PR-5 thaumatin-like protein allergen from common cypress (*Cupressus sempervirens*) pollen. Allergies 2023;3:11-24.

31. Soman KV, Midoro-Horiuti T, Ferreon JC, Goldblum RM, Brooks EG, Kurosky A, et al. Homology modeling and characterization of IgE binding epitopes of mountain cedar allergen Jun a 3. Biophys J 2000;79:1601-9.

32. Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics 1991;11:635-50.

33. Altschul S, Madden T, Schaffer A, Zhang JH, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Faseb J 1998;12:A1326-A.

34. Aalberse RC. Assessment of sequence homology and cross-reactivity. Toxicol Appl Pharmacol 2005;207:149-51.

35. Breiteneder H, Mills C. Structural bioinformatic approaches to understand cross-reactivity. Mol Nutr Food Res 2006;50:628-32.

36. Mathura VS, Schein CH, Braun W. Identifying property based sequence motifs in protein families and superfamilies: application to DNase-1 related endonucleases. Bioinformatics 2003;19:1381-90.

37. Ivanciuc O, Midoro-Horiuti T, Schein CH, Xie L, Hillman GR, Goldblum RM, et al. The property distance index PD predicts peptides that cross-react with IgE antibodies. Mol Immunol 2009;46:873-83.

38. 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. J Agr Food Chem 2003;51:4830-7.

39. Schein CH, Ivanciuc O, Braun W. Common physical-chemical properties correlate with similar structure of the IgE epitopes of peanut allergens. J Agric Food Chem 2005;53:8752-9.

40. Schein CH, Ivanciuc O, Braun W. Bioinformatics approaches to classifying allergens and predicting cross-reactivity. Immunol Allergy Clin North Am 2007;27:1-27.

41. Ivanciuc O, Oezguen N, Mathura VS, Schein CH, Xu Y, Braun W. Using property based sequence motifs and 3D modeling to determine structure and functional regions of proteins. Curr Med Chem 2004;11:583-93.

42. Schein CH, Zhou B, Oezguen N, Mathura VS, Braun W. Molego-based definition of the architecture and specificity of metal-binding sites. Proteins Struct Funct Bioinform 2005;58:200-10.

43. Braun BA, Schein CH, Braun W. DGraph clusters flaviviruses and beta-coronaviruses according to their hosts, disease type, and human cell receptors. Bioinform Biol Insights 2021;15:11779322211020316.

44. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. Colab-Fold: making protein folding accessible to all. Nat Methods 2022;19:679-82.

45. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics 2013;29:2722-8.

46. Ozden B, Kryshtafovych A, Karaca E. Assessment of the CASP14 assembly predictions. Proteins 2021;89:1787-99.

47. Rego N, Koes D. 3Dmol.js: molecular visualization with WebGL. Bioinformatics 2015;31:1322-4.

48. 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. J Allergy Clin Immunol 2008;121:847-52.e7.

49. Ivanciuc O, Garcia T, Torres M, Schein CH, Braun W. Characteristic motifs for families of allergenic proteins. Mol Immunol 2009;46:559-68.

50. Mistry J, Bateman A, Finn RD. Predicting active site residue annotations in the Pfam database. BMC Bioinform 2007;8:298.

51. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 2016;44:D279-85.

52. Vieths S, Scheurer S, Ballmer-Weber B. Current understanding of cross-reactivity of food allergens and pollen. Ann N Y Acad Sci 2002;964:47-68.

53. Geroldinger-Simic M, Zelniker T, Aberer W, Ebner C, Egger C, Greiderer A, et al. Birch pollen-related food allergy: clinical aspects and the role of allergen-specific IgE and IgG4 antibodies. J Allergy Clin Immunol 2011;127:616-22.e1.

54. Masthoff LJ, van Hoffen E, Mattsson L, Lidholm J, Andersson K, Zuidmeer-Jongejan L, et al. Peanut allergy is common among hazelnut-sensitized subjects but is not primarily the result of IgE cross-reactivity. Allergy 2015;70:265-74.

55. Midoro-Horiuti T, Mathura V, Schein CH, Braun W, Yu S, Watanabe M, et al. Major linear IgE epitopes of mountain cedar pollen allergen Jun a 1 map to the pectate lyase catalytic site. Mol Immunol 2003;40:555-62.

56. 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. J Allergy Clin Immunol 1999;104:613-7.

57. Fraczkiewicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. J Comp Chem 1998;19:319.

58. Poncet P, Senechal H, Charpin D. Update on pollen-food allergy syndrome. Expert Rev Clin Immunol 2020;16:561-78.

59. Cortegano I, Civantos E, Aceituno E, del Moral A, Lopez E, Lombardero M, et al. Cloning and expression of a major allergen from *Cupressus arizonica* pollen,

Cup a 3, a PR-5 protein expressed under polluted environment. Allergy 2004;59: 485-90.

60. Suarez-Cervera M, Castells T, Vega-Maray A, Civantos E, del Pozo V, Fernandez-Gonzalez D, et al. Effects of air pollution on Cup a 3 allergen in *Cupressus arizonica* pollen grains. Ann Allergy Asthma Immunol 2008;101:57-66.

61. Leone P, Menu-Bouaouiche L, Peumans WJ, Payan F, Barre A, Roussel A, et al. Resolution of the structure of the allergenic and antifungal banana fruit thaumatin-like protein at 1.7-A. Biochimie 2006;88:45-52.

62. Dreskin SC, Germinaro M, Reinhold D, Chen X, Vickery BP, Kulis M, et al. IgE binding to linear epitopes of Ara h 2 in peanut allergic preschool children undergoing oral immunotherapy. Pediatr Allergy Immunol 2019;30:817-23.

63. Chen X, Negi SS, Liao S, Gao V, Braun W, Dreskin SC. Conformational IgE epitopes of peanut allergens Ara h 2 and Ara h 6. Clin Exp Allergy 2016;46:1120-8.

64. Liao S, Patil SU, Shreffler WG, Dreskin SC, Chen X. Human monoclonal antibodies to Ara h 2 inhibit allergen-induced, IgE-mediated cell activation. Clin Exp Allergy 2019;49:1154-7.

65. Otsu K, Guo R, Dreskin SC. Epitope analysis of Ara h 2 and Ara h 6: characteristic patterns of IgE-binding fingerprints among individuals with similar clinical histories. Clin Exp Allergy 2015;45:471-84.

66. Mueller GA, Gosavi RA, Pomes A, Wunschmann S, Moon AF, London RE, et al. Ara h 2: crystal structure and IgE binding distinguish two subpopulations of peanut allergic patients by epitope diversity. Allergy 2011;66:878-85.

67. Bilaver LA, Chadha AS, Doshi P, O'Dwyer L, Gupta RS. Economic burden of food allergy: a systematic review. Ann Allergy Asthma Immunol 2019;122:373-80.e1.

68. Langlois A, Lavergne MH, Leroux H, Killer K, Azzano P, Paradis L, et al. Protocol for a double-blind, randomized controlled trial on the dose-related efficacy of omalizumab in multi-food oral immunotherapy. Allergy Asthma Clin Immunol 2020; 16:25.

69. Fleischer DM, Greenhawt M, Sussman G, Begin P, Nowak-Wegrzyn A, Petroni D, et al. Effect of epicutaneous immunotherapy vs placebo on reaction to peanut protein ingestion among children with peanut allergy: the PEPITES randomized clinical trial. JAMA 2019;321:946-55.

70. Goodman RE, Breiteneder H. The WHO/IUIS Allergen Nomenclature. Allergy 2019;74:429-31.

71. Schein CH. Identifying Similar Allergens and Potentially Cross-Reacting Areas Using Structural Database of Allergenic Proteins (SDAP) Tools and D-Graph Beatriz Cabanillas (ed.), Food Allergens: Methods and Protocols, Methods in Molecular Biology, vol. 2717, chapter 18.