

Identification of Almond (*Prunus dulcis*) Vicilin As a Food AllergenHuilian Che,^{†,‡} Yuzhu Zhang,^{*,‡,§} Shu-Chen Lyu,[§] Kari C. Nadeau,[§] and Tara McHugh[‡][†]Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Food Science and Nutritional Engineering, China Agricultural University, No. 17 Qinghua Donglu, Haidian District, Beijing 100038, P. R. China[‡]Agricultural Research Service, Pacific West Area, Western Regional Research Center, U. S. Department of Agriculture, 800 Buchanan Street, Albany, California 94710, United States[§]Division of Pediatric Immunology, Allergy, and Rheumatology, Department of Pediatrics, Stanford University School of Medicine, 269 Campus Drive, Stanford, California 94305, United States

S Supporting Information

ABSTRACT: Almond is one of the tree nuts listed by U.S. FDA as a food allergen source. A food allergen identified with patient sera has been debated to be the 2S albumin or the 7S vicilin. However, neither of these proteins has been defined as a food allergen. The purpose of this study was to clone, express, and purify almond vicilin and test whether it is a food allergen. Western blot experiment was performed with 18 individual sera from patients with double-blind, placebo-controlled clinical almond allergy. The results showed that 44% of the sera contained IgE antibodies that recognized the recombinant almond vicilin, indicating that it is an almond allergen. Identifying this and additional almond allergens will facilitate the understanding of the allergenicity of seed proteins in tree nuts and their cross-reactivity.

KEYWORDS: storage protein, Pru du vicilin, food allergen, protein expression

■ INTRODUCTION

Currently, there is no cure for food allergy. On the basis of studies in the United States, Canada, Australia, United Kingdom, and Israel, food allergies affect nearly 5% of adults and 8% of children in Westernized countries.¹ Food allergies may be life-threatening, especially peanut and tree nut allergies.^{2,3} Most food allergies are triggered by immunoglobulin E (IgE) recognition of food allergens.⁴ Most food allergens are proteins. In the United States, four of the eight major food allergen sources⁵ (peanuts, soybeans, tree nuts, and wheat) are foods of plant origin. Plant seeds are indispensable protein sources for human nutrition. There are thousands of proteins in a plant seed, but a larger percentage of the total protein weight is made up of seed-storage proteins.⁶ No biological function is known for the seed-storage proteins other than that they reserve nutrients for the growth of future young plants. Unfortunately, some of these seed-storage proteins are known to be food allergens, causing millions of people to suffer.^{7,8} Cross-reactivity among different foods poses additional risks to allergic patients,^{9,10} making identification and characterization of food allergen essential for solving the allergy problem. Moreover, development of patient-tailored risk profiles with component-resolved diagnostics (which is promising for a better understanding of food allergies¹¹) also relies on the identification and availability of the allergens.

To date, 16 peanut allergens have been given official names by the World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Subcommittee.¹² These are Ara h 1 to Ara h 17, with Ara h 3 and 4 historically named differently but now officially considered as the same allergen.^{13,14} The prevalence of tree nut allergies is as high as that of peanut allergy in the general population in the United States.¹⁵ However, the prevalence of

allergies to individual tree nut is not well-studied, and the identification of allergens in tree nuts is far from comprehensive. Many tree nut allergens remain to be uncovered. For instance, no macadamia nut allergens have been defined although severe macadamia nut allergies have been reported.^{16–20} While 10 hazelnut food allergens have been defined, the number of officially designated almond (*Prunus dulcis*) allergens is only half of that.¹² Nevertheless, four additional almond proteins were referenced when the molecular characterization and clinical relevance of almond allergens were discussed in a recent article in *Journal of Agricultural and Food Chemistry*.²¹ These included the pathogenesis-related proteins, the thaumatin-like proteins, the 2S albumin, and the γ -conglutins, which belong to the vicilins of the cupin superfamily.

The 7S globulins of the seed-storage proteins are called vicilins. They are present in monocotyledonous and dicotyledonous species. The vicilin proteins in many species have been identified as allergens, including peanut allergen Ara h 1,²² cashew allergen Ana o 1,²³ hazelnut allergen Cor a 11,²⁴ walnut allergen Jug r 2,²⁵ pine nut allergen Pin k 2,²⁶ and pecan allergen Car i 2.²⁷ The prevalence of almond allergy is relatively high but varies with the geographic regions of the patient groups.^{15,28} As a food, almond is the seed of a drupe of *Prunus dulcis*. A previously identified allergenic almond protein was reported as Pru du 2S albumin²⁹ but later suspected as Pru du vicilin.³⁰ Here we report the identification of almond vicilin as a new food allergen.

Received: September 28, 2018

Revised: November 19, 2018

Accepted: December 4, 2018

Published: December 4, 2018

MATERIALS AND METHODS

Preparation of Genomic DNA. A whole, young Nonpareil almond (collected May 26, 2016, from Travaille & Phippen, Inc.) was ground with a pestle in a mortar containing liquid nitrogen, which was added to keep the sample submerged at all times. After grinding, ~100 mg of the sample was used to isolate genomic DNA using the RNeasy Plant Mini kit and QIAprep Spin Miniprep kit (QIAGEN, Valencia, CA) with a procedure combining the manufacturer's protocols for the two kits, with minor modifications. Briefly, the sample was transferred to a dry ice-cooled RNase-free, DNase-free 2 mL microcentrifuge tube, and buffer RLT (450 μ L) from the RNA kit was immediately added to the sample. The sample was vortexed vigorously, transferred to a QIAshredder spin column, and centrifuged for 2 min at 16 200g in a microcentrifuge. The supernatant of the flow-through (the lysate) was transferred to a QIAprep spin column placed in a 2 mL collection tube and spun at 16 200g for 2 min. DNA was eluted after the column was washed following the miniprep kit protocol. Sixty microliters of Milli-Q water was used to elute the genomic DNA.

Polymerase Chain Reaction and Cloning. The Q5 high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA) was used for all polymerase chain reaction (PCR) experiments. Multiple primer pairs were designed based on the mRNA sequence of the predicted peach (*Prunus persica*) vicilin C72 (NCBI reference sequence: XM-007225564.2). A 2.5 k base-pair PCR product was obtained using the almond genomic DNA as template with primers zy342 (ccaaatgtcacactgcatg, forward) and zy343 (ccacttcaaaactccagc, reverse). The PCR product was inserted in an in-house cloning vector, pBC, prepared with *EcoR* V, and the ligation product was used to transform *dH5 α* bacteria. Positive transformants were selected on a Luria broth (LB) plate containing kanamycin. Three independent clones were cultured for plasmid DNA isolation using the Qiaprep spin miniprep kit (QIAGEN, Valencia, CA). The isolated plasmids were sequenced commercially by Elim Biopharmaceuticals, Inc. (Hayward, CA), using Elim's primers T3 (aattaacctcactaaaggg) and M13for-40 (gttttccagtcacgac). On the basis of the sequencing results, additional primers zy354 (gtcacctacttgattaa) and zy355 (gccatt-caagtagtcccat) were synthesized to cover the entire genomic DNA inserted in the vector. Introns were identified manually based on the sequence alignment between the isolated gene and the predicted vicilin sequence of peach. Intron-exon boundaries were also predicted using the GENSCAN³¹ server at MIT.³² The signal peptide was predicted based on sequence alignment of vicilins from almond (this study), peach, pine nut,³³ and pecan.²⁷ The coding sequence of the predicted mature almond vicilin was commercially synthesized as a gblock by Integrated DNA Technologies, Inc. (Skokie, IL), with codon optimization. An extra sequence (ctcgaggtgtctcaaatctctgatgt-tac) was added to the 3' end for downstream cloning purpose. The gblock was amplified with PCR using primers zy365 (aatccggat-atcgttggtgggcaagaggaagaagaa, forward) and zy340 (gtaacatcagagattt-gagacac, reverse). The PCR product was digested with *EcoR* V and *Xho* I and inserted in the pTMS1 vector³⁴ digested with *Dra* I and *Xho* I to generate pNH-PruduV. The gblock was also used in a PCR using primers zy364 (aatccggatcgttggttaacccaacccctactcttc, forward) and zy340 to amplify the structural core of the protein based on sequence alignment with pine nut allergen Pin k 2 and pecan allergen Car i 2. The PCR product was digested with *EcoR* V and *Xho* I and inserted in the pTMS1 vector³⁴ to generate pNH-PruduVc.

The inserted sequence in pNH-PruduV was confirmed by DNA sequencing with the primers above, and the plasmid was transformed into the *E. coli* strain BL21(DE3)* (Invitrogen). The bacteria carrying pNH-PruduV were grown in LB with kanamycin (50 mg/L, Gold Biotechnology, St. Louis, MO) in 1 L batches at 37 °C. When the OD₆₀₀ of the culture reached 1.0, the temperature of the incubator was reduced to 18 °C. The culture was incubated for an additional 45 min before isopropyl β -D-1-thiogalactopyranoside (Gold Biotechnology) was added to a final concentration of 1 mM. The expression of the proteins was allowed for 16 h, and the bacteria were collected by

centrifugation at 3 500g. The pellets were stored at -80 °C for further use.

Frozen cell pellets were thawed at the temperature of tap water and resuspended in 50 mL of buffer HB (20 mM imidazole, pH 8.0, 500 mM NaCl) containing 1 mM benzamidine. The bacteria were sonicated for 6 min in a beaker surrounded by ice and water. The lysate was subject to centrifugation at 20 000g for 40 min at 18 °C. The supernatant was loaded onto a 5 mL HisTrap FF Crude column (GE Healthcare, Piscataway, NJ). The column was washed with 25 mL of buffer HW (45 mM imidazole, pH 8.0, 500 mM NaCl) before the protein was eluted with buffer HE (300 mM imidazole, pH 8.0, 500 mM NaCl).

SDS-PAGE Analysis. Samples were incubated at 96 °C in a sodium dodecyl sulfate (SDS) sample buffer (50 mM Tris-HCl, pH6.8, 2% SDS, 0.1% bromophenol blue, and 10% glycerol) containing 100 mM β -mercaptoethanol for 10 min before loading. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using 4–20% polyacrylamide gels with a Tris-HEPES-SDS running buffer (100 mM Tris, 100 mM HEPES, 3 mM SDS, pH 8.0). Prestained protein molecular weight standards of 10, 15, 20, 25, 37, 50, 75, 100, 150, and 250 kDa (Bio-Rad, Hercules, CA) were used as references. Gels were stained as previously described³⁵ and documented with an ImageQuant LAS 400 Imager (GE Healthcare).

N-terminal Protein Sequencing. Purified recombinant was subjected to SDS-PAGE as described above and transferred to a polyvinylidene difluoride (PVDF) membrane (GE Healthcare) using a Tris-glycine buffer (48 mM Tris, 39 mM glycine, pH 9.2) and a Tans-Blot SD Semidry Transfer Cell (Bio-Rad, Hercules, CA) following the manufacturer's protocols, and the blot was stained with the conventional Coomassie Brilliant Blue method. The peptide bands of interest were excised and sent to the analytical core facility at Tufts Medical School for N-terminal amino acid sequencing.

Preparation of Almond Protein Extract. Twenty grams of ripe Nonpareil almond kernel was ground in 150 mL of a Tris buffer (10 mM, pH 8.0) for 5 min in a KitchenAid blender precooled at -20 °C. The sample was centrifuged at 4 000g for 15 min. The supernatant was collected and incubated at 0 °C for 30 min in an ice-water bucket. The sample was then centrifuged at 24 000g for 40 min. The supernatant was collected as almond protein extract-S, and the pellet was collected as almond protein extract-P. Extract-S and extract-P were mixed at 10:1 and used as almond protein extract.

Patient Sera. Eighteen sera (nos. 1–18 corresponding to deidentified patient codes S00210, MC12-2, S00309, S00066, MC04-4, M32-5, M37-2, 07-T07, 07-T04, 07-T14, 07-T12, M22-5, M02-4, MC08-5, M29-5, S00281, M40-4, and M04-4, respectively) were collected from patients with a positive oral food allergy challenge to almond. The patients were enrolled in the food allergy study at Stanford University under institutional review board approval with informed consent (IRB approval certificate no. 8629).

IgE Recognition of Almond Vicilin. Almond protein extract and recombinant vicilin were heated to 96 °C for 10 min in 1 $\times$ SDS sample buffer containing 100 mM β -mercaptoethanol and separated by electrophoresis with 4–20% gels. Multiple SDS gels were used to assess IgE binding by Western blot and for protein detection by Coomassie Brilliant Blue (CBB) staining. For Western blot, protein bands in the SDS gels were transferred to PVDF membranes as described above. The membranes were blocked for 1 h at room temperature in TBST (25 mM Tris, pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.1% Tween 20) containing 5% nonfat dry milk. In the meantime, individual serum (150 μ L) was incubated with 10 μ L of protein A immobilized with agarose (Pierce, Rockford, IL, U.S.A.) at 4 °C in 3 mL of TBST with 1% nonfat milk. After incubation, the serum samples were subjected to centrifugation at 220g for 2 min. Each supernatant was used to incubate with a blocked membrane for 45 min at room temperature. The membranes were washed with TBST 3 $\times$ 5 min followed by incubating with a peroxidase-conjugated secondary antibody against human IgE (Sigma) for 45 min at room temperature. The antibody was diluted 5 000 times with TBST containing 1% nonfat milk before use. The membranes were then

washed 3×5 min with TBST, and a Pierce ECL Western Blotting substrate (Rockford, IL) was used for detection using the Image-Quant LAS4000 imaging system.

RESULTS AND DISCUSSION

Isolation of Almond Vicilin Gene. The cupin superfamily is among the few protein families with numerous known food allergens. Both the 11S and 7S seed-storage proteins belong to the cupin family, and they have been identified as allergens in peanut and a number of tree nuts, including cashew, hazelnut, pecan, pistachio, and walnut. Additionally, the 11S proteins are also food allergens in almond and Brazil nut; the 7S protein in pine nut is also known to be a food allergen. However, the almond vicilin has not been investigated although an almond allergen previously believed to be the 2S albumin²⁹ was speculated as almond vicilin.³⁰ The sequence of vicilin from peach, a close relative of almond, is available in the NCBI database (XM-007225564.2). PCR primers designed based on the sequence of peach vicilin were used to amplify the vicilin gene from almond genomic DNA. As shown in Figure S1 of the Supporting Information, a 2517 base pair cDNA was isolated. The vicilin gene sequence was predicted to contain four introns, and the protein deduced from the predicted coding sequence has 547 amino acids. Searching the nonredundant protein database at NCBI with BLAST using the predicted protein as a query reported vicilins from many species as hits, with the highest sequence identity (98%) to peach vicilin C72. Pine nut allergen Pin k 2 was the first vicilin to be shown as a copper protein.³³ Multiple sequence alignment with vicilins of peanut and other tree nuts that are known food allergens showed that one of the copper coordinating residues is not conserved in almond vicilin (Figure S2), indicating that, unlike the other tree nut allergens, the almond vicilin and Ara h 1 do not have copper ligands.

Expression and Purification Almond Vicilin. Many food allergens were defined by the following experiment sequence: When sera of subjects were provided by patients allergic to a given food, proteins were extracted from the food. The extraction was separated into protein bands with SDS polyacrylamide gels, and bands that could be recognized by IgE antibodies in the sera were identified. The bands were isolated and sequenced to define the proteins. In the case of identifying almond allergens, a 12 kDa IgE reacting protein band was considered to be the 2S albumin²⁹ but was suspected to be almond vicilin later.³⁰ To this end, however, neither the 2S albumin nor the 7S vicilin of almond has been unambiguously identified as a food allergen. Isolating almond vicilin based on the property of known vicilins from other species was not successful (data not shown). Thus, we decided to produce recombinant almond vicilin, with a His tag, based on its coding sequence determined above. In parallel, the structural core of almond vicilin was predicted by aligning its sequence with those of Ara h 1 and Car i 2. A plasmid for expressing the almond vicilin structural core (rPruduVc) was constructed. The yields of the expression for both rPruduV and rPruduVc, however, were quite low, especially compared with those for the peanut allergen Ara h 1^{36–38} and the pecan allergen Car i 2.²⁷ Nevertheless, a portion of these recombinant proteins was soluble and remained in the supernatant after sonication and centrifugation. The rPruduV and rPruduVc purified with a Ni^{2+} column, along with the almond protein extract, was analyzed with SDS-PAGE as shown in Figure 1. Because the expression level of the soluble rPruduV and

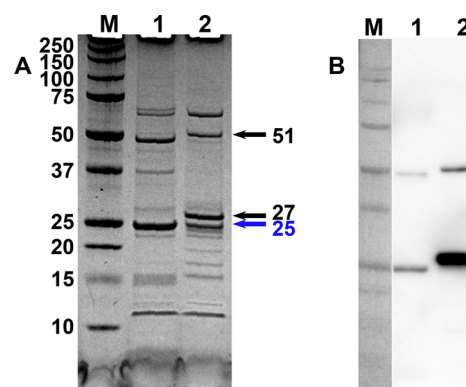


Figure 1. SDS-PAGE analysis of recombinant almond vicilin. Samples of Ni^{2+} column purified recombinant almond vicilin (lane 2) and its structure core (lane 1) were analyzed with homemade 4–20% SDS polyacrylamide gels. The gel was stained with Coomassie Brilliant Blue (A) or transferred to a PVDF membrane for Western blot analysis using an antibody against His-tag (B). Protein standards (see text) were loaded in lane M. Western blot results were documented with the imaging system in the chemiluminescence mode. White light images of the membranes were taken with otherwise identical settings. The positions of the bands of the prestained standards in the second image where overlaid with the chemiluminescence image. The molecular mass of each of the protein standards and the major components of the vicilin preparation (in kDa) are labeled next to the corresponding band in panel (A).

rPruduVc was low, their purity after Ni^{2+} column purification was also low because of the high relative abundance of the impurities. Several bands with similar strength could be seen for each of the proteins as shown in Figure 1A. To ensure the recombinant expression of these proteins was successful and to identify which band(s) was the expressed target, an antibody produced against poly-His (Invitrogen) was used in a Western blot to identify His-tag proteins. As shown in Figure 1B, two bands with apparent molecular masses of ~ 50 and ~ 25 kDa, respectively, were recognized by the antibody in the rPruduVc preparation. For rPruduV, two bands with slightly higher molecular mass than those from rPruduVc were detected. The band ~ 50 kDa was consistent with the theoretical molecular mass of rPruduV (51.1 kDa) and rPruduVc (49.9 kDa) as calculated by SPHERE.^{39,40} In addition, the insoluble portion of the expressed rPruduVc in the inclusion body also showed two bands at 50 and 25 kDa, respectively, while the insoluble portion of rPruduV showed strong bands at 51, 27, and 25 kDa, respectively. These results suggested that the recombinant proteins were processed by an unknown peptidase in the bacteria post-translationally. It was previously known that, in many species,^{25,41,42} vicilins can become more than one fragment of different molecular weight as a result of post-translational peptidase processes. Peanut vicilin (Ara h 1), which is a major allergen,^{43,44} was also truncated when expressed in bacteria, which was attributed to a rare arginine codon.⁴⁵ Walnut vicilin consisted of multiple bands when it was first extracted from walnut kernels and identified as food allergen Jug r 2.²⁵ Recombinant Jug r 2 was also prone to truncation.²⁵ If the recombinant almond vicilin was processed by a peptidase, it would be likely that the same peptide bond was cleaved in rPruduV and rPruduVc. This was exactly what the data in Figure 1 suggested because rPruduV and its N-terminal fragment had slightly higher molecular masses than those of the corresponding bands in rPruduVc. It seemed that

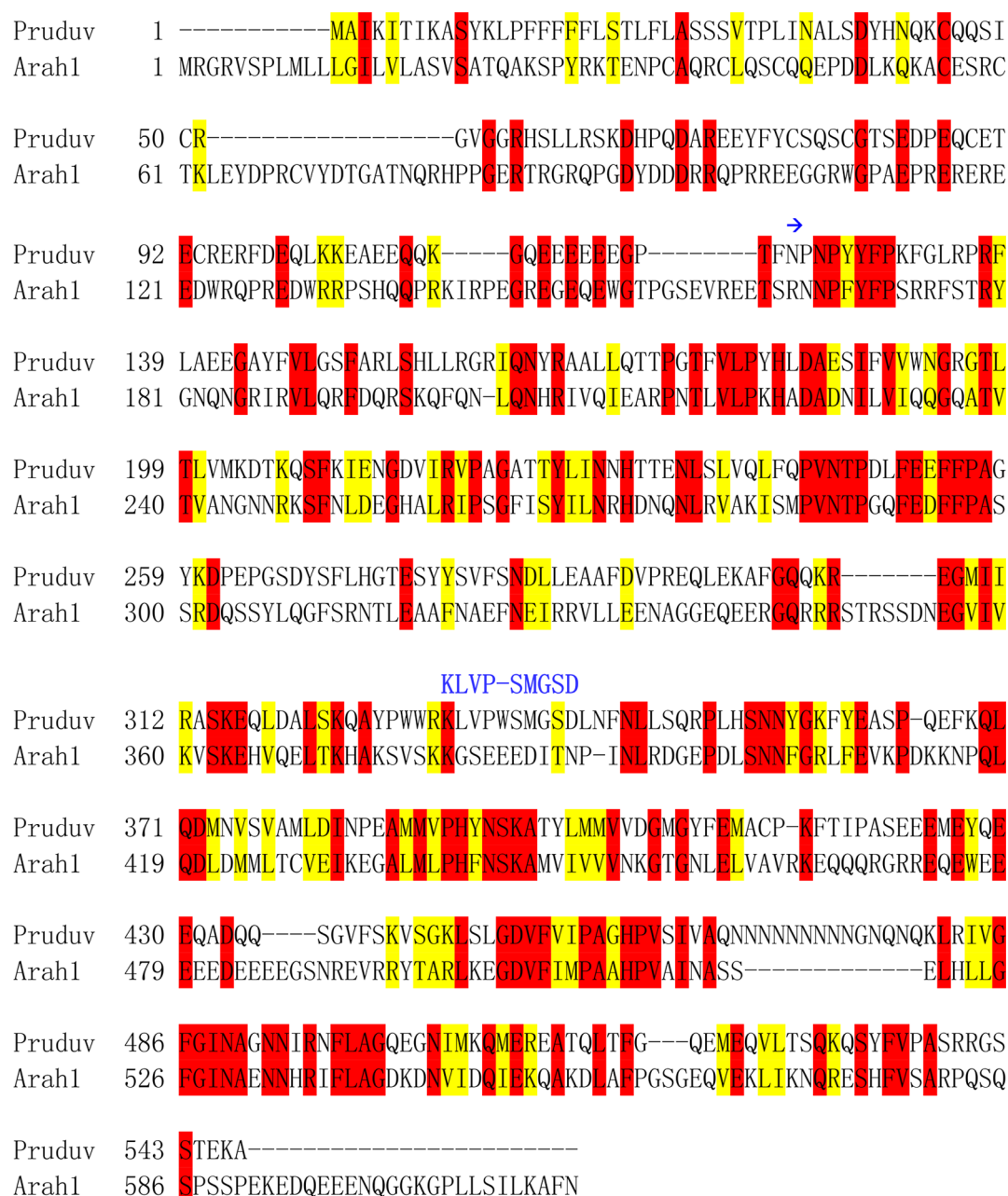


Figure 2. Sequence alignment of almond vicilin and Ara h 1. Red and yellow background indicates identical and similar residues. The N-terminal sequencing result of the C-terminal fragment is shown in blue font above the PruduV sequence. The start of the structural core is marked by a blue arrow above the PruduV sequence.

the cleavage site was close to the middle of rPruduVc, making the two cleaved fragments overlapped on the SDS polyacrylamide gels. The two cleavage products could be resolved for rPruduV because it had a longer N-terminal fragment. To confirm this prediction, the smaller fragment of rPruduV (as indicated by the blue arrow in Figure 1A) was isolated and subjected to N-terminal sequencing. The first 10 residues for this peptide was KLVP-SMGSD, matching a sequence in the almond vicilin as shown in Figure 2, considering W often give little or no signal in N-terminal sequencing. The theoretical molecular masses for the N- and C-terminal fragments are 26.95 and 24.14 kDa, respectively. Together, these data in Figures 1 and 2 indicated that the three relatively strong bands resolved from the rPruduV preparation at ~25, 27, and ~51 kDa at the point of the arrows in Figure 1A are the C-terminal

fragment, the N-terminal fragment, and the full-length rPruduV, respectively. This preparation was then used in testing whether any of the sera contained IgE specific to almond vicilin.

Identification of Almond Vicilin As an Allergen.

Eighteen sera were used individually to test whether they recognize almond vicilin. All patients were diagnosed by double-blind placebo-controlled food challenge (DBPCFC, which is considered the gold standard method for food allergy diagnosis) to be allergic to almond. The patients were enrolled in the food allergy study at the Stanford University. Seventeen of the patients were children with their ages between 3 and 13 (median 7). A 28-year-old adult was also included. Some of them also have an allergy to walnut. The clinical information on the patients is shown in Table 1.

Table 1. Clinical Findings of Almond-Allergic Patients

patient no.	PPID	age	symptoms ^a	IgEb (kU/L) ^b	allergy to food ^c
1	S00210	10	A,U	33.60	almond
2	MC12-2	13	R,U	54.9	almond
3	S00309	28	no data	no data	almond
4	S00066	6	U	8.4	almond
5	MC04-4	3	A,R,U	2.42	almond
6	M32-5	7	A,R,U	0.8	almond
7	M37-2	7	R,U	1.26	almond
8	07-T07	10	A,R,U	1.95	almond
9	07-T04	8	U	1.02	almond
10	07-T14	7	A,R,U	no data	almond
11	07-T12	7	A,R,U	12.8	almond
12	M22-5	8	A,R,U	36.8	almond, walnut
13	M02-4	5	A,R,U	2.98	almond, walnut
14	MC08-5	11	A,U	100	almond, walnut
15	M29-5	5	A,R,U	7.44	almond, walnut
16	S00281	5	A,R,U	2.03	almond, walnut
17	M40-4	9	A,R,U	0.35	almond, walnut
18	M04-4	11	A,R,U	55.6	almond, walnut

^aAbbreviations for symptoms: A, asthma; R, rhinitis; U, urticaria; dbpcfc = double-blind, placebo-controlled food challenge. ^bImmunoCAP specific IgE. ^cFood allergies diagnosed by dbpcfc.

Whether these sera contained IgE antibodies against almond vicilin was tested by Western blot experiments. Almond protein extract and rPruduV were separated by SDS polyacrylamide gels and analyzed. Among the 18 sera, 8 contained IgE antibodies that recognized recombinant vicilin as shown in Figure 3, indicating that almond vicilin is a food allergen. Six of the sera (nos. 2, 12, and 15–18) recognized only one of the cleaved products of rPruduV besides the full-length protein. Serum 11 and 14 recognized both the N- and C-terminal fragments.

All sera that recognized rPruduV also recognized two bands in the lane of the almond protein extract just above the position of the 37 kDa protein standard in the marker. All other sera, except number 10, recognized more than one other protein bands, albeit weak, but not rPruduV. The patterns of peptide bands that were recognized by a given serum differ from one another. Thus, the result from one serum could be used as a control for other sera because each serum was tested individually. Hence, no sera from normal subjects were included as negative controls.

Comparison with Peanut Vicilin. Cross-reactivity between food allergens also has a negative impact on agricultural products. Although peanut and tree nuts are botanically far separated, allergens from these foods can cross-react. For example, incubating sera from peanut allergy patients with Brazil nut extract can inhibit the binding of specific IgE to peanut extract.^{10,46,47} The sequence of this newly identified almond allergen was compared with that of peanut allergen Ara h 1 by pairwise sequence alignment. The longest stretch of identical amino acids consists of six residues as shown in Figure 2, indicating that cross-reaction between these two food allergens due to linear IgE epitope is not very likely. The patient group who donated samples in this study did not show an allergy to peanut. However, the stretch of six identical residues was separated from two additional identical residues in the aligned sequences by one amino acid. The 9-residue peptide stretch may potentially contribute to cross-reactivity.

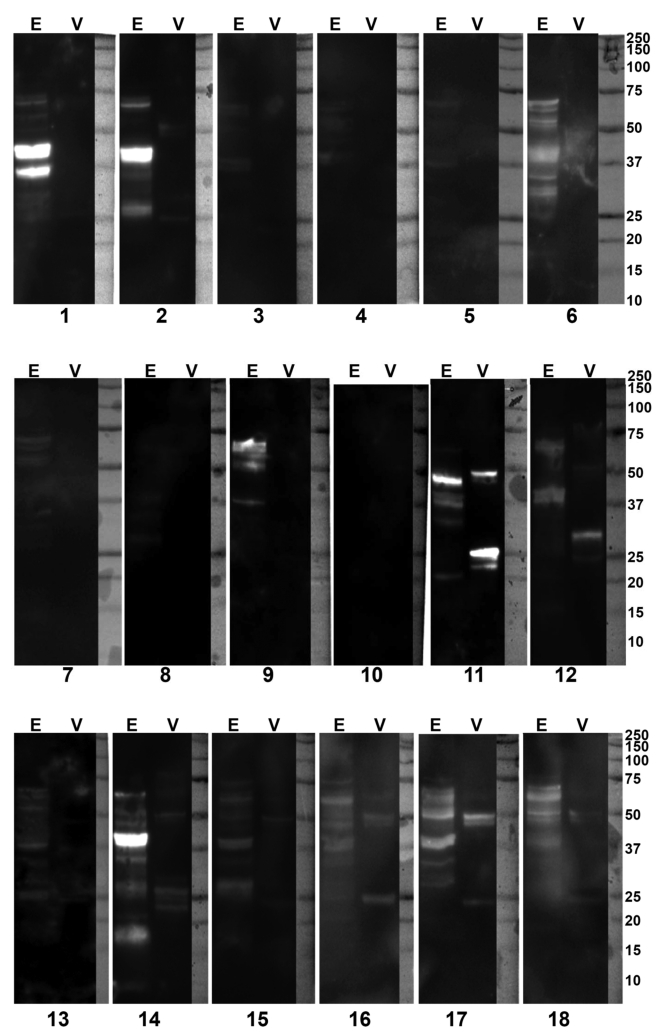


Figure 3. Western blot with 18 patient sera. Proteins in almond protein extract (lane E, see text) and a sample of Ni²⁺ column purified recombinant almond vicilin (lane V) were denatured, separated with 4–20% SDS-gels, and transferred to PVDF membranes. The membranes were blotted with individual serum from 18 almond allergic patients. The serum number is labeled at the bottom of the blot. The images were obtained, and the molecular masses of the protein standards are labeled as described in Figure 1.

A BLAST search of the PDB database returned a number of hits. These are vicilin orthologs from other species, and some of them are known food allergens. Among the BLAST hits, Car i 2²⁷ had the highest sequence identity with almond vicilin (49%), followed by the eggplant (*Solanum melongena*) vicilin⁴⁸ (40%). A structural model of almond vicilin was built with SWISS-MODEL⁴⁹ using the structure of Car i 2 as a starting structure. The model structure of almond vicilin can largely be superimposed with the structure of peanut allergen Ara h 1 as shown in Figure 4A. The molecular surfaces of almond vicilin and peanut allergen Ara h 1 were calculated using the program PyMol.⁵⁰ The charge distribution on the surfaces of the two molecules did not reveal a high potential of cross-reactivity between these two food allergens (Figure 4B and data not shown).

At present, 19 nuts are included in the FDA list of tree nut allergen sources.⁵¹ Almond is one of the most commonly produced and consumed nuts. In a study of 5149 subjects with self-reported food allergy, almond was reported to be the third

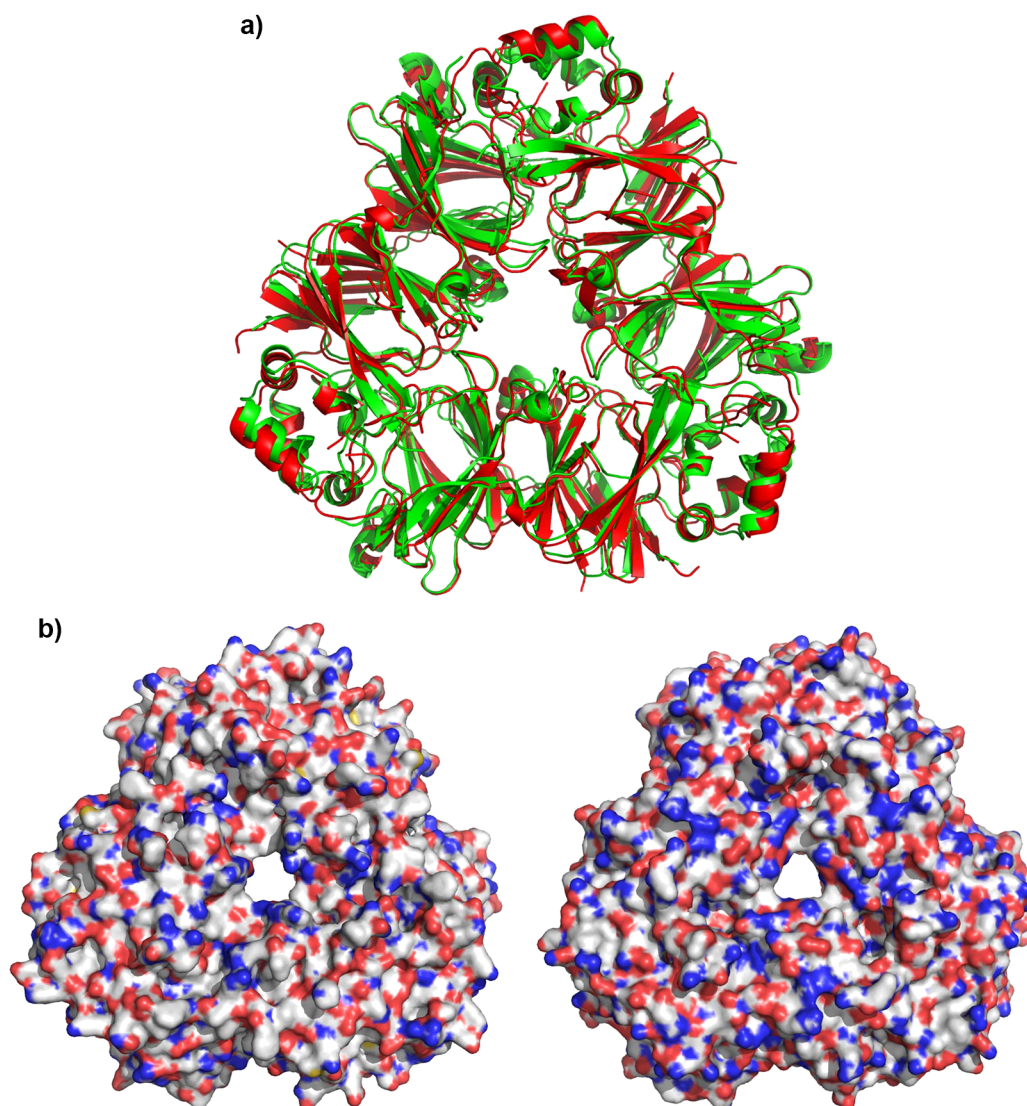


Figure 4. Comparison of a model structure of almond vicilin and the structure of peanut allergen Ara h 1. (a) Superimposition of the two structures. The structure of Ara h 1 is shown in green, and that of almond vicilin is shown in red. (b) Molecular surface of almond vicilin (left) and Ara h 1 (right) viewed from the same angle of the imposed proteins. Positive charges are shown in blue, and negative charges are shown in red.

leading cause of tree nut allergies behind walnut and cashew,⁵² while in another study of 122 doctor-diagnosed tree nut allergies, almond was the second leading cause following only walnut.⁵³ Identification of almond allergens will facilitate a better understanding of the allergenicity of plant proteins from almond and other tree nuts. This study showed that the almond 7S vicilin reacted with IgE antibodies in 8 of the 18 sera, making it the fifth almond food allergen being characterized. However, the almond allergen revealed one and a half decades ago remains to be identified.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.8b05290.

Gene sequence of almond vicilin (PDF)

Multiple sequence alignment of the newly identified and previously known peanut and tree nut vicilin allergens (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: 510 559 5981. Fax: 510 559 5818. E-mail: Yuzhu.Zhang@ars.usda.gov.

ORCID

Yuzhu Zhang: 0000-0001-7882-5692

Funding

This work was supported by U. S. Department of Agriculture, Agricultural Research Services.

Notes

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

The authors declare no competing financial interest.

■ REFERENCES

- (1) Sicherer, S. H.; Sampson, H. A. Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *J. Allergy Clin. Immunol.* **2014**, 133, 291–307. and quiz 308.

- (2) Sampson, H. A.; Mendelson, L.; Rosen, J. P. Fatal and near-fatal anaphylactic reactions to food in children and adolescents. *N. Engl. J. Med.* **1992**, *327*, 380–384.
- (3) Yunginger, J. W.; Sweeney, K. G.; Sturner, W. Q.; Giannandrea, L. A.; Teigland, J. D.; Bray, M.; Benson, P. A.; York, J. A.; Biedrzycki, L.; Squillace, D. L.; et al. Fatal food-induced anaphylaxis. *Jama* **1988**, *260*, 1450–1452.
- (4) Asero, R.; Ballmer-Weber, B. K.; Beyer, K.; Conti, A.; Dubakiene, R.; Fernandez-Rivas, M.; Hoffmann-Sommergruber, K.; Lidholm, J.; Mustakov, T.; Oude Elberink, J. N.; Pumphrey, R. S.; Stahl Skov, P.; van Ree, R.; Vlieg-Boerstra, B. J.; Hiller, R.; Hourihane, J. O.; Kowalski, M.; Papadopoulos, N. G.; Wal, J. M.; Mills, E. N.; Vieths, S. IgE-mediated food allergy diagnosis: Current status and new perspectives. *Mol. Nutr. Food Res.* **2007**, *51*, 135–147.
- (5) USFDA. Food Allergies: What You Need to Know. <http://www.fda.gov/Food/ResourcesForYou/Consumers/ucm079311.htm> (accessed Sep 19, 2018).
- (6) Shewry, P. R. Plant storage proteins. *Biol. Rev. Camb. Philos. Soc.* **1995**, *70*, 375–426.
- (7) Añón, M. A.; Maselli, J. P.; Sanz, M. L.; Fernandez-Benitez, M. Allergy to pine nut. *Allergol. Immunopathol.* **2002**, *30*, 104–108.
- (8) Garcia-Menaya, J. M.; Gonzalo-Garijo, M. A.; Moneo, I.; Fernandez, B.; Garcia-Gonzalez, F.; Moreno, F. A 17-kDa allergen detected in pine nuts. *Allergy* **2000**, *55*, 291–293.
- (9) Beardslee, T. A.; Zeece, M. G.; Sarath, G.; Markwell, J. P. Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *Int. Arch. Allergy Immunol.* **2000**, *123*, 299–307.
- (10) de Leon, M. P.; Drew, A. C.; Glaspole, I. N.; Suphioglu, C.; O'Hehir, R. E.; Rolland, J. M. IgE cross-reactivity between the major peanut allergen Ara h 2 and tree nut allergens. *Mol. Immunol.* **2007**, *44*, 463–471.
- (11) Flores Kim, J.; McCleary, N.; Nwaru, B. I.; Stoddart, A.; Sheikh, A. Diagnostic accuracy, risk assessment, and cost-effectiveness of component-resolved diagnostics for food allergy: A systematic review. *Allergy* **2018**, *73*, 1609–1621.
- (12) WHO/IUIS-Allergen-Nomenclature-Sub-committee. Allergen Nomenclature. <http://www.allergen.org/> (accessed Sep 19, 2018).
- (13) Pomes, A.; Davies, J. M.; Gadermaier, G.; Hilger, C.; Holzhauser, T.; Lidholm, J.; Lopata, A. L.; Mueller, G. A.; Nandy, A.; Radauer, C.; et al. WHO/IUIS Allergen Nomenclature: Providing a common language. *Mol. Immunol.* **2018**, *100*, 3–13.
- (14) Radauer, C.; Nandy, A.; Ferreira, F.; Goodman, R. E.; Larsen, J. N.; Lidholm, J.; Pomes, A.; Raulf-Heimsoth, M.; Rozynek, P.; Thomas, W. R.; Breiteneder, H. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. *Allergy* **2014**, *69*, 413–419.
- (15) Cox, A.; Sicherer, S. H. Peanut and tree nut allergy. *Chem. Immunol. Allergy* **2015**, *101*, 131–144.
- (16) Sutherland, M. F.; O'Hehir, R. E.; Czarny, D.; Suphioglu, C. Macadamia nut anaphylaxis: demonstration of specific IgE reactivity and partial cross-reactivity with hazelnut. *J. Allergy Clin. Immunol.* **1999**, *104*, 889–980.
- (17) De Knop, K. J.; Hagendorens, M. M.; Bridts, C. H.; Stevens, W. J.; Ebo, D. G. Macadamia nut allergy: 2 case reports and a review of the literature. *Acta Clin. Belg.* **2010**, *65*, 129–132.
- (18) Lerch, M.; Egger, C.; Bircher, A. J. Allergic reactions to macadamia nut. *Allergy* **2005**, *60*, 130–131.
- (19) Ekbote, A.; Hayman, G.; Bansal, A. Macadamia nut allergy: potentially misleading specific IgE results. *Allergy* **2010**, *65*, 1345.
- (20) Knott, E.; Gurer, C. K.; Ellwanger, J.; Ring, J.; Darsow, U. Macadamia nut allergy. *J. Eur. Acad. Dermatol. Venereol.* **2008**, *22*, 1394–1395.
- (21) Costa, J.; Mafra, I.; Carrapatoso, I.; Oliveira, M. B. Almond allergens: molecular characterization, detection, and clinical relevance. *J. Agric. Food Chem.* **2012**, *60*, 1337–1349.
- (22) Viquez, O. M.; Konan, K. N.; Dodo, H. W. Structure and organization of the genomic clone of a major peanut allergen gene, Ara h 1. *Mol. Immunol.* **2003**, *40*, 565–571.
- (23) Wang, F.; Robotham, J. M.; Teuber, S. S.; Tawde, P.; Sathe, S. K.; Roux, K. H. Ana o 1, a cashew (*Anacardium occidentale*) allergen of the vicilin seed storage protein family. *J. Allergy Clin. Immunol.* **2002**, *110*, 160–166.
- (24) Lauer, I.; Foetisch, K.; Kolarich, D.; Ballmer-Weber, B. K.; Conti, A.; Altmann, F.; Vieths, S.; Scheurer, S. Hazelnut (*Corylus avellana*) vicilin Cor a 11: molecular characterization of a glycoprotein and its allergenic activity. *Biochem. J.* **2004**, *383*, 327–334.
- (25) Teuber, S. S.; Jarvis, K. C.; Dandekar, A. M.; Peterson, W. R.; Ansari, A. A. Identification and cloning of a complementary DNA encoding a vicilin-like proprotein, jug r 2, from english walnut kernel (*Juglans regia*), a major food allergen. *J. Allergy Clin. Immunol.* **1999**, *104*, 1311–1320.
- (26) Zhang, Y. Z.; Du, W. X.; Fan, Y. T.; Yi, J.; Lyu, S. C.; Nadeau, K. C.; McHugh, T. H. Identification, characterization, and initial epitope mapping of pine nut allergen Pin k 2. *Food Res. Int.* **2016**, *90*, 268–274.
- (27) Zhang, Y.; Lee, B.; Du, W. X.; Lyu, S. C.; Nadeau, K. C.; Grauke, L. J.; Zhang, Y.; Wang, S.; Fan, Y.; Yi, J.; McHugh, T. H. Identification and Characterization of a New Pecan [*Carya illinoensis* (Wangenh.) K. Koch] Allergen, Car i 2. *J. Agric. Food Chem.* **2016**, *64*, 4146–4151.
- (28) McWilliam, V.; Koplin, J.; Lodge, C.; Tang, M.; Dharmage, S.; Allen, K. The Prevalence of Tree Nut Allergy: A Systematic Review. *Curr. Allergy Asthma Rep.* **2015**, *15*, 54.
- (29) Poltronieri, P.; Cappello, M. S.; Dohmae, N.; Conti, A.; Fortunato, D.; Pastorello, E. A.; Ortolani, C.; Zacheo, G. Identification and characterisation of the IgE-binding proteins 2S albumin and conglutin gamma in almond (*Prunus dulcis*) seeds. *Int. Arch. Allergy Immunol.* **2002**, *128*, 97–104.
- (30) Garino, C.; De Paolis, A.; Coisson, J. D.; Arlorio, M. Pru du 2S albumin or Pru du vicilin? *Comput. Biol. Chem.* **2015**, *56*, 30–32.
- (31) Burge, C.; Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **1997**, *268*, 78–94.
- (32) Burge, C. The GENSCAN Web Server at MIT. <http://genes.mit.edu/GENSCAN.html> (accessed Sep 19, 2018).
- (33) Jin, T.; Wang, Y.; Chen, Y. W.; Fu, T. J.; Kothary, M. H.; McHugh, T. H.; Zhang, Y. Crystal structure of the Korean pine (*Pinus koraiensis*) 7S seed storage protein with copper ligands. *J. Agric. Food Chem.* **2014**, *62*, 222–228.
- (34) Zhang, Y.; Fan, Y. A Mutant Sumo Facilitates Quick Plasmid Construction for Expressing Proteins with Native N-termini After Tag Removal. *Mol. Biotechnol.* **2017**, *59*, 159–167.
- (35) Zhang, Y.; Du, W. X.; Fregevu, C.; Kothary, M. H.; Harden, L.; McHugh, T. H. Expression, Purification, and Characterization of Almond (*Prunus dulcis*) Allergen Pru du 4. *J. Agric. Food Chem.* **2014**, *62*, 12695–12700.
- (36) Cabanos, C.; Urabe, H.; Masuda, T.; Tandang-Silvas, M. R.; Utsumi, S.; Mikami, B.; Maruyama, N. Crystallization and preliminary X-ray analysis of the major peanut allergen Ara h 1 core region. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2010**, *66*, 1071–1073.
- (37) Chruszcz, M.; Maleki, S. J.; Majorek, K. A.; Demas, M.; Bublin, M.; Solberg, R.; Hurlburt, B. K.; Ruan, S.; Mattisohn, C. P.; Breiteneder, H.; Minor, W. Structural and immunologic characterization of Ara h 1, a major peanut allergen. *J. Biol. Chem.* **2011**, *286*, 39318–39327.
- (38) Cabanos, C.; Urabe, H.; Tandang-Silvas, M. R.; Utsumi, S.; Mikami, B.; Maruyama, N. Crystal structure of the major peanut allergen Ara h 1. *Mol. Immunol.* **2011**, *49*, 115–123.
- (39) Zhang, Y.-Z. Protein and peptide structure and interactions studied by hydrogen exchange and NMR. Dissertation, University of Pennsylvania, 1995; Dissertations available from ProQuest, AA19532311.
- (40) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75–86.
- (41) Casey, R. Distribution and Some Properties of Seed Globulins. In *Seed Proteins*; Shewry, P. R., Casey, R., Eds.; Springer: Dordrecht, 1999; pp 159–169.

- (42) Jin, T.; Albillos, S. M.; Chen, Y. W.; Kothary, M. H.; Fu, T. J.; Zhang, Y. Z. Purification and Characterization of the 7S Vicilin from Korean Pine (*Pinus koraiensis*). *J. Agric. Food Chem.* **2008**, *56*, 8159–8165.
- (43) Burks, A. W.; Shin, D.; Cockrell, G.; Stanley, J. S.; Helm, R. M.; Bannon, G. A. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur. J. Biochem.* **1997**, *245*, 334–339.
- (44) Barre, A.; Sordet, C.; Culerrier, R.; Rance, F.; Didier, A.; Rouge, P. Vicilin allergens of peanut and tree nuts (walnut, hazelnut and cashew nut) share structurally related IgE-binding epitopes. *Mol. Immunol.* **2008**, *45*, 1231–1240.
- (45) Becker, W. M. Characterization of Ara h 1 by two-dimensional electrophoresis immunoblot and recombinant techniques: new digestion experiments with peanuts imitating the gastrointestinal tract. *Int. Arch. Allergy Immunol.* **2004**, *113*, 118–121.
- (46) de Leon, M. P.; Drew, A. C.; Glaspole, I. N.; Suphioglu, C.; Rolland, J. M.; O'Hehir, R. E. Functional analysis of cross-reactive immunoglobulin E antibodies: peanut-specific immunoglobulin E sensitizes basophils to tree nut allergens. *Clin. Exp. Allergy* **2005**, *35*, 1056–1064.
- (47) de Leon, M. P.; Glaspole, I. N.; Drew, A. C.; Rolland, J. M.; O'Hehir, R. E.; Suphioglu, C. Immunological analysis of allergenic cross-reactivity between peanut and tree nuts. *Clin. Exp. Allergy* **2003**, *33*, 1273–1280.
- (48) Jain, A.; Kumar, A.; Salunke, D. M. Crystal structure of the vicilin from *Solanum melongena* reveals existence of different anionic ligands in structurally similar pockets. *Sci. Rep.* **2016**, *6*, 23600.
- (49) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.
- (50) PyMol homepage. <http://pymol.org/> (accessed Sep 19, 2018).
- (51) USDA. What nuts are considered "tree nuts?". <http://www.fda.gov/ForIndustry/FDABasicsforIndustry/ucm238807.htm> (accessed Sep 19, 2018).
- (52) Sicherer, S. H.; Furlong, T. J.; Munoz-Furlong, A.; Burks, A. W.; Sampson, H. A. A voluntary registry for peanut and tree nut allergy: characteristics of the first 5149 registrants. *J. Allergy Clin. Immunol.* **2001**, *108*, 128–132.
- (53) Sicherer, S. H.; Burks, A. W.; Sampson, H. A. Clinical features of acute allergic reactions to peanut and tree nuts in children. *Pediatrics* **1998**, *102*, No. e6.

1 CCAAATGTACACACTGCATGCTCAGCCATGCCTTATTAAAAGCCTCTCACTATAAACAAACACCTGATAT


71 AACAACTATAGCCCTAAACATGGCGATCAAGATCACCATCAAAGCTAGCTATAAGCTCCCTTTCTTCTT
141 CTTCTTCTCTCCACTCTCTTCCTAGCCTCCTCCTCTGTAACACCACTGATCAACGCCCTCAGTGATTAT
211 CATAATCAGAAATGCCAGCAGTCGATTTGCAGAGGCGTTGGCGGCCGCCATAGCTTGTTAAGATCCAAAG
281 ATCATCCTCAAGATGCTCGAGAGGAATACTTTTATTGCTCGCAGAGCTGCGGAACGAGTGAGGACCCCGA
351 ACAGTGCAGACAGAGTGTCTGAGAGCGATTTGACGAGCAATTGAAGAAAGAAGCTGAAGAGCAGCAGAAG
421 GGACAAGAGGAGGAGGAGGAGGAGGGGCCGACTTTTAATCCCAACCCTTACTATTTCCCCAAGTTCGGGT
491 TGAGGCCTAGGTTTTTGGCTGAAGAAGGTGCCTATTTTGTCTTGGGAGCTTTGCCAGGCTCTCGCATTT
561 GCTCCGGGGAAGAATTCAAACTACCGCGCTGCCTTGCTTCAGACTACGCCTGGCACTTTTGTCTCCCT
631 TACCACCTCGACGCCGAATCCATTTTTGTGTGTTGGAACG GTGAGACATGGAACTAGTTTCAAAATTGT
701 AGTTATTATATTGAGTCGATTTTGTGTTTGTCTTAATTTAATTAAATTGGAATTACAATTTACAGGAAGAGG
771 CACTCTCACTCTTGTGATGAAGGACACGAAACAGTCCCTCAAGATTGAAAATGGAGATGTAATTAGAGTT
841 CCTGCCGGGGCTACCACCTACTTGATTAACAATCACACCACAGAAAATCTTAGCCTCGTACAGCTCTTCC
911 AGCCAGTCAACACTCCTGACCTATTTGAGGTAATTAAATTTTTGTGTTTTGTTTTCTTTTTTCGAATTTCTA
981 TGCTAGCTATTAGAAAACTACTTAAAATTTTCACATAGAAACGTGAAAATGAGTTACTTCTTTTTCTTT
1051 TAGGAAAAAAAAAACTTTTATAGAATTTTTTATAGGAATTGCTTTGATCATTGGCTGCAGGAATTCTTCC
1121 CTGCTGGTTATAAAGATCCCGAACCCGGGTCAGACTATAGTTTTTTTGCACGGTACCGAGTCATACTACAG
1191 CGTTTTTCAGCAATGACCTTCTTGAAGCTGCCTTTGATGTAAGTTTGTGTTTTTCTTCCTTAAATATTTCC
1261 CTTTTATTTCTTTTTAATTTGTTTCAATTTACACTAATTGCTAATTTGATTATCAGGTCCCAAGAGAGCAGC
1331 TAGAAAAGGCCTTTGGACAGCAGAAGCGGGAGGGCATGATCATAAGAGCCTCAAAGAGCAACTCGACGC
1401 ATTGAGTAAACAAGCTTATCCATGGTGGCGCAAACCTAGTCCCTTGGTCTATGGGTAGTGATTTAAATTTT
1471 AATCTGCTAAGCCAAAGACCTCTGCACTCCAACAACCTATGGCAAGTTCTATGAGGCTTCCCCTCAGGAGT
1541 TCAAACAGCTCCAGGACATGAATGTCTCAGTCGCTATGCTTGACATCAATCCGGTAGGGCAAACCTTAA
1611 ACATTAATTTCCAGAAGAAATTTAATTCCATATACAAATGTTTTAGTAAGGGAAATTTAAAACTAGCTA
1681 TTGCAAGGGGATTGTACTTCAAGCAGTTAAGAACAGTTCCCCTTGCAACCCTCAGGTCCTGTGTTCGAATC
1751 CCTTCACTCCTCCCTCTTCCAATATTACTTGTATAAAAGATTAAAAACAGTGAGACAATGCATGTGTACT
1821 TGATGAATTAAAATCTGTGAAATTTTTGCAGGAAGCTATGATGGTGCCACACTACAATTCTAAAGCAACA
1891 TATTTGATGATGGTGGTAGATGGAATGGGATACTTTGAAATGGCCTGCCCAAAGTTCACAATACCAGCGA
1961 GCGAAGAAGAGATGGAGTATCAAGAAGAGCAAGCAGATCAGCAGAGCGGAGTGTTTCAAGCAAGGTAAGTGG
2031 CAAGCTTTCCCTTGGTGATGTCTTTGTGATCCAGCAGGCCATCCTGTGTCTATTGTTGCCCAAACAAC
2101 AACACAACAACAACAACAACGGCAACCAAAACCAGAAATTGAGAATTGTGGGATTCGGAATCAATGCCG
2171 GAAACAACATAAGGAACCTTCTTAGCAGGTAAATTGTTAATCAAGTCAGACTTAATATGCTTAATAATTAC
2241 ATTTTGTGTTTAAATTAAGTGGGTGTGTATATAAATGATATGTTCTCTGCTTTGTTAAATATATGCAGGGCA
2311 AGAGGGCAACATAATGAAGCAGATGGAGAGAGAGGCTACGCAGCTGACATTTGGGCAAGAGATGGAACAA
2381 GTATTGACCAGTCAGAAGCAATCTTATTTTGTGCCTGCCTCGAGAAGAGGGAGCAGCACAGAGAAGGCAT
2451 AAGCAATCCCAACTGGCTTCAAGTACGTAGAGCAGAAGTAGTATGCTTTGCTGGAGTTTTGAAGTGG


Figure S1. Gene sequence of almond vicilin. There 5' and 3' untranslated sequences and the predicted introns are shown in red. The coding sequence are shown in blue.

Pink2	1	-----
Arah1	1	-----
Cari2	1	MVTKAKIPLFLFLSALFLALVCSSLALETEDLSNELNPHHPESHREWFQQCQERCQHEERGQRQAQQCQRRCEEQLRER
Jugr2	1	-----
Pisv3	1	-----
Cora11	1	-----
Pruduv	1	-----
Anao1	1	-----
jugn2	1	-----

Pink2	1	-----
Arah1	1	-----
Cari2	81	EREREREEIVDPREPRKQYEQCRETCEKQDPRQQPCERRCERQFQEQQERERRRRRGRDDDDKENPRDPREQYRQCEE
Jugr2	1	-----
Pisv3	1	-----
Cora11	1	-----
Pruduv	1	-----
Anao1	1	-----
jugn2	1	-----

Pink2	1	-----
Arah1	1	-----MRGRVSPLMLLLGILVLASVSATQAKS
Cari2	161	HCRRQGQGQRQQQQCQSRCEERFEEEQRRQEEERERRRRGRDNDDEENPRDPREQYRQCQEHCCRQGGQRQQQQCQSRCEE
Jugr2	1	-----RGRDDDDDEENPRDPREQYRQCQEYCCRQGGQRQQQQCQIRCEE
Pisv3	1	-----
Cora11	1	-----
Pruduv	1	-----MAIKITIKASY
Anao1	1	-----MGPTTKFSFSLFLSVLV
jugn2	1	-----

Pink2	1	---MAFVSLLTILAISSSS-----V
Arah1	28	PYRK TENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYDTGATNQRHPPGERTRGRQPGDYDDDRQPRREEGG
Cari2	241	RLEEEQRKQEERERRRRGRDDEDDQNPDPPEQRYEQCQQQCERQRRGQEQLCRRRCEQRRQEEERERQGRDRQDPQQQYH
Jugr2	45	RLEEDQRSQEERERRRRGRDVDDQNPDPPEQRYEQCQQQCERQRRGQEQLCRRRCEQRRQEEERERQGRDRQDPQQQYH
Pisv3	1	-----KTDPELKQCKHQCKVQRQYDEEQKEQCAKGCEKYKKEKGREQE-----EWWWGSGRGRGDEFSTHEGKRL
Cora11	1	---MLPKEDPELKCKHKCRDERQFDEQQ
Pruduv	12	KLPFFFFLSTLFLASSSVTPLINALS-DYHNQKQQSICRGVGGRHSLLSK-----DHPQDAREEYF
Anao1	19	LCLGFALAKIDPELKQCKHQCKVQRQYDEEQKEQCVKECEKYKKEKGREREHEEEEEWGTGG--VDEPSTHEAEKHL
jugn2	1	-----GRDRQDPQQQYH

Pink2	19	ALTEPVASMAQQGVFPEQHGRGHHGVFPEEHGRGHHRG-----GREEREENPVVFHSDRFIRASSHAGEI
Arah1	108	RWGPAEPREREEDWRQPEDWRRPSHQPRKIRPEGREGEQEWGTPGSEVREETSRNNPFYFPPSRRESTRYGNQNGRI
Cari2	321	RCQRRQCQTQEESPERQRCQQRCERQYKEQQGREWGPDQ---ASPRRESRGREEEQQRHNPYYFHSQGLSRHESCEGEV
Jugr2	125	RCQRRCQIQEQSPERQRCQQRCERQYKEQQGRERGPE---ASPRRESRGREEEQQRHNPYYFHSQSIISRHESCEGEV
Pisv3	72	SQCMKQCERQDGGQQKQLRFRCQEKYKERRHSYSR---DEEEEGDEHQEEDENPVVFEDHFTRVKTEQCKV
Cora11	27	-----RRDG---KQICEEKARERQQEE-----GNSSEE---SYGKEEENPVVFQDEHFESRVKTEEGRV
Pruduv	75	YCSQSCGTSED-PE---QCETEERERFDEQLKKEAEQ---QKGQEEEEEGPTFNPNPYPFKFGLAPFLAEEGAY
Anao1	97	SQCMKQCERQEGGQQKQLRFRCQERYNKERQHNKYR---EDDEDED-EDAEEDENPVVFEDHFTTKVKTEQCKV
jugn2	13	RCQRRCQIQEQSPERQRCQQRCERQYKEQQGRERGPE---ASPRRESKGREEEQQRHNPYYFHSQSIISRHESCEGEV

Pink2	86	RALPNFGEVSELLEQ--ISRYRVTCIEMKPNTVMIPHYIDAKWTLVYTGGRGYIAYQQNELVKRKLLEGDVFGVPSGHIIF
Arah1	188	RVLQRFDQRSKQFQN-LQNHRIVQIEARPNTLVLPKHADADNLLVIQQGQATVTVANGNNKSFNLDEGHALRIISGFIS
Cari2	398	KYLERFTFTELLRG-IENYRVVILEANPNTFVLPYHKDAESVIVVTRGRATLTLFSQERRESFNLEYGDIVRVPAGATE
Jugr2	201	KYLERFTFTELLRG-IENYRVVILDANPNTSLPHHKDAESVAVVTRGRATLTLFSQETRESFNLECGDIVRVPAGATV
Pisv3	148	VVLPKFTKRSKLLRG-LEKYLRAFLVANPQAFVVPNIMDADSIFFVSWGRGITITKIRENKRESMVKQGDIIIRIRAGTPF
Cora11	81	QVLENFTKRSRLLSG-IENFRLAILEANPHIFISPAHFDAELVLFVAKGRATITMREEKRESFNVEHGDIIIRIPAGTPV
Pruduv	146	FVLGSEARLSHLRGRIQNYRAALLQTTPTGTVLPYHLDAESIFVVWNGRGITLIMKDKTSFKLENGDIVRVPAGATT

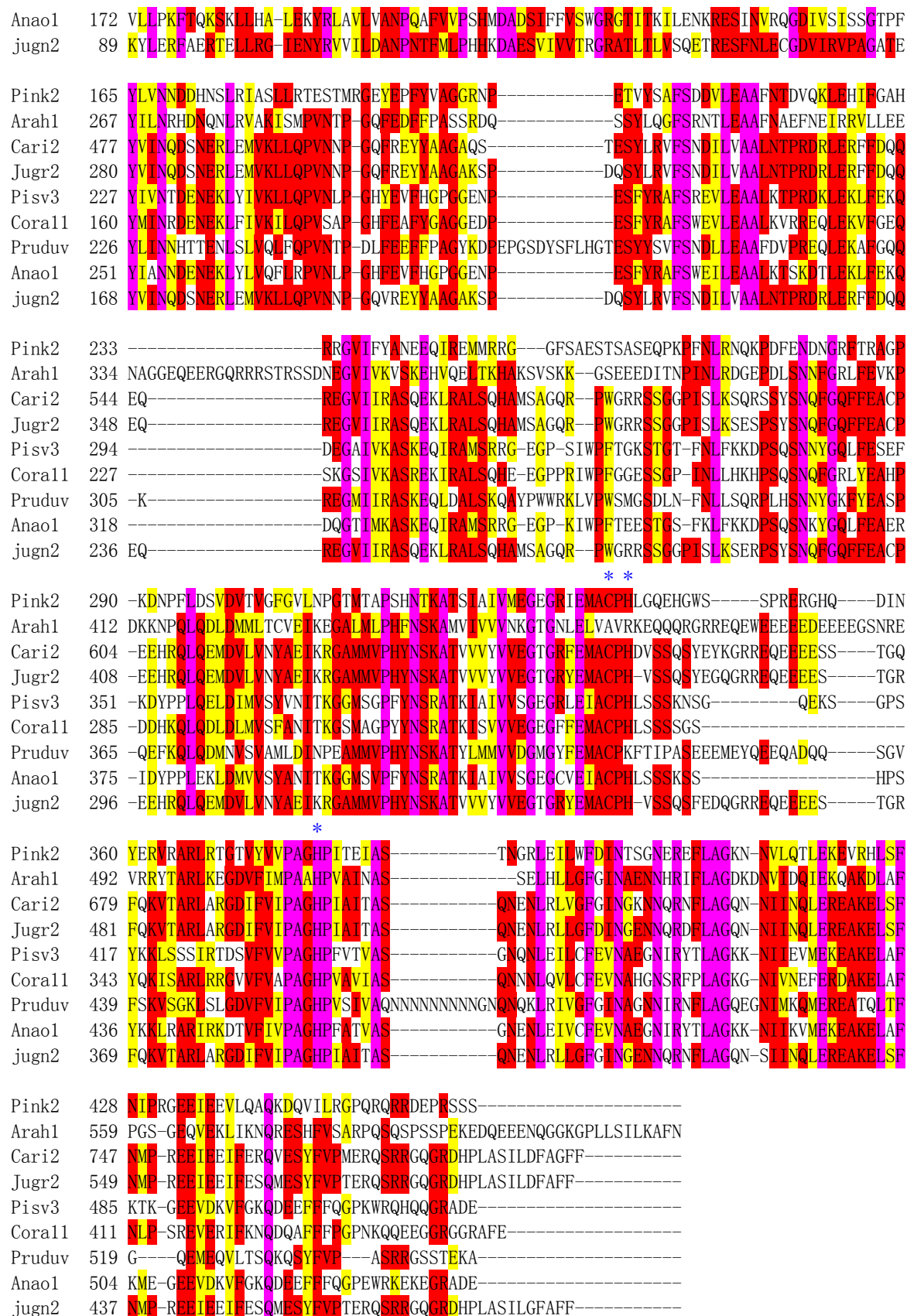


Figure S2. Multiple sequence alignment of the newly identified and previously known peanut and tree nut vicilin allergens. Magenta background indicates conserved residues. Red and yellow background indicates identical and similar residues in more than 60% of the sequences. Blue asterisks on top of the Pin k 2 sequence indicate the copper-coordinating residues in the pine nut allergen.