

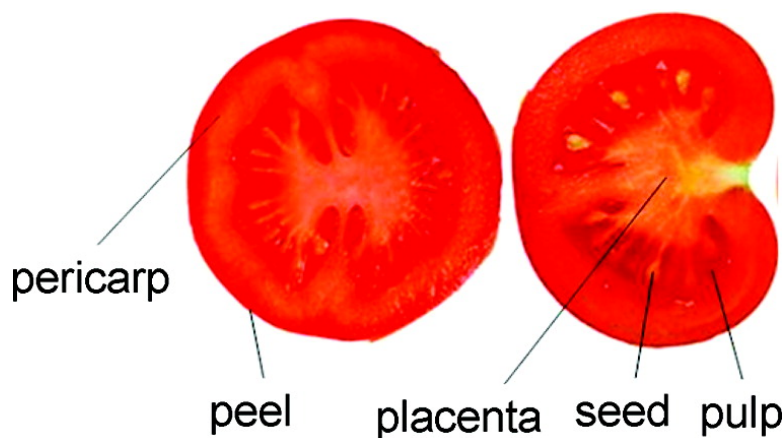
Article

Evidence for Novel Tomato Seed Allergens: IgE-Reactive Legumin and Vicilin Proteins Identified by Multidimensional Protein Fractionation#Mass Spectrometry and *in Silico* Epitope Modeling

Olivia Y. Baessler, Julia Weiss, Stefanie Wienkoop, Karola Lehmann, Christian Scheler, Sabine Döhlle, Dietmar Schwarz, Philipp Franken, Eckhard George, Margitta Worm, and Wolfram Weckwerth

J. Proteome Res., Article ASAP • DOI: 10.1021/pr800186d • Publication Date (Web): 09 February 2009

Downloaded from <http://pubs.acs.org> on February 12, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)



ACS Publications
High quality. High impact.

Journal of Proteome Research is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

Evidence for Novel Tomato Seed Allergens: IgE-Reactive Legumin and Vicilin Proteins Identified by Multidimensional Protein Fractionation—Mass Spectrometry and *in Silico* Epitope Modeling

Olivia Y. Bässler,[†] Julia Weiss,[‡] Stefanie Wienkoop,[‡] Karola Lehmann,[§] Christian Scheler,[§] Sabine Dölle,^{||} Dietmar Schwarz,[⊥] Philipp Franken,[⊥] Eckhard George,[⊥] Margitta Worm,^{||} and Wolfram Weckwerth^{*,†,‡}

Max Planck Institute of Molecular Plant Physiology, D-14467 Potsdam, Germany, Universität Potsdam GoFORSYS, D-14467 Potsdam, Germany, Proteome Factory AG, D-10117 Berlin, Germany, Allergie-Centrum-Charité, Department of Dermatology and Allergy, D-10117 Berlin, Germany, and Leibniz Institute of Vegetable and Ornamental Crops (IGZ), D-14979 Grossbeeren, Germany

Received March 12, 2008

Tomato fruit and seed allergens were detected by IgE-immunoblotting using sera from 18 adult tomato-sensitized patients selected based on a positive history skin prick test (SPT) and specific Immunglobulin (Ig) E-levels. Isolated tomato seed total protein showed high SPT activity comparable or even higher than tomato fruit protein. For the molecular characterization of tomato seed allergens, a multidimensional protein fractionation strategy and LC-MS/MS was used. Two legumin- and vicilin-proteins were purified and showed strong IgE-reactivity in immunoblots. Individual patient sera exhibited varying IgE-sensitivity against the purified proteins. *In silico* structural modeling indicates high homology between epitopes of known walnut allergens and the detected IgE-crossreactive tomato proteins.

Keywords: food allergy • tomato • legumin • vicilin • IgE • mass spectrometry • LC-MS/MS • immunoblot

Introduction

Tomato (*Lycopersicon esculentum*) is a widely produced and important crop plant. Recently, several studies demonstrated the allergenicity of tomatoes. However, molecular information about distinct tomato protein allergens is limited. Tomato is one of the most important food crops worldwide and belongs to the *Solanaceae* or nightshade family which has previously been reported to be allergenic. The prevalence of tomato allergy ranges from 1.5%^{1,2} to 20%² in different populations of food-allergic patients, indicating that tomato is indeed a relevant allergenic food in selected populations.³ Tomato allergy is an immunoglobulin E (IgE)-mediated response capable of causing serious symptoms. Allergy against tomato was shown to be linked to other allergies, for example, against grass pollen⁴ and latex⁵ by cross-reactivity with homologous protein sequences. However, the role of cross-reactivity for the development of a sensitization is still unclear. While the immune response in allergenic patients, from the initial sensitization with allergens to the release of bioactive mediators, has been widely studied, the common structure and physiological functions present in allergenic proteins have not been completely characterized.

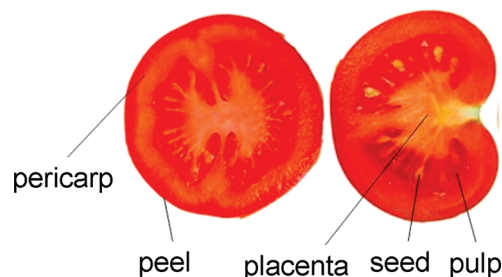


Figure 1. Parts of the tomato fruit.

The Lyc e 1-allergen (a profilin, gi2499814), Lyc e 2 (a β -fructofuranosidase, gi18542112) and the Lyc e 3-allergen (a nonspecific lipid-transfer protein, gi2497745) are known tomato allergens in fruit tissue and are listed at the International Union of Immunological Societies (I.U.I.S.) Allergen Nomenclature Sub-Committee database (<http://www.allergen.org>). The whole tomato fruit extract was analyzed by Kondo and co-workers⁶ using IgE-immunoblotting techniques to identify proteins as possible allergens. Proteins were identified by N-terminal amino acid sequencing using Edman degradation. In this study, a polygalacturonase 2A (PG2A), a superoxid dismutase and a pectinesterase (PE) were identified as putative tomato allergens exhibiting IgE cross-reactivity. The concentrations of PG2A, PE and β -fructofuranosidase were highest in the red ripening stage detected by SDS-PAGE and immunoblotting.⁶ Other proteins with capacity to bind IgE have also been described such as glucanase,⁷ peroxidase,⁸ chitinase,⁹ and patain.¹⁰

* To whom correspondence should be addressed. E-mail: Weckwerth@mpimp-golm.mpg.de. Phone: +49-331-5678109. Fax: +49-331-5678134.

[†] Max Planck Institute of Molecular Plant Physiology.

[‡] Universität Potsdam GoFORSYS.

[§] Proteome Factory AG.

^{||} Allergie-Centrum-Charité.

[⊥] Leibniz Institute of Vegetable and Ornamental Crops (IGZ).

Table 1. Characterization of the Used Patient Sera (1–18) and the Control Serum (C)

serum number	age	sex	total IgE [kU/L] ^a	grass IgE [kU/L] ^a	mugwort IgE [kU/L] ^a	birch IgE [kU/L] ^a	tomato IgE [kU/L] ^a	tomato skin prick test (SPT)
1	40	M	114	20.9	11.7	30.4	2.02	Positive
2	41	M	184	43.6	0	29.3	n.d.	Positive
3	43	M	51.1	4	1.82	24.3	n.d.	Positive
4	38	M	167	14	0	41	n.d.	Positive
5	18	F	383	69.7	0.48	37.1	0.82	Positive
6	18	F	234	n.d.	0.57	n.d.	n.d.	Positive
7	38	F	167	14	0	41	0.4	Positive
8	36	F	405	2.71	0.43	100	1.57	Positive
9	31	F	114	0	0	0	0.4	Positive
10	36	F	521	n.d.	n.d.	n.d.	17.3	n.d.
11	56	F	57.7	n.d.	n.d.	n.d.	6.87	n.d.
12	26	M	334	88	0.5	0	n.d.	Positive
13	40	M	79.4	13.3	0	0	n.d.	Positive
14	58	F	57.2	3.22	0	1.42	n.d.	Positive
15	43	F	328	52.5	0	0	n.d.	Positive
16	28	F	271	n.d.	n.d.	n.d.	2.83	n.d.
17	33	M	65.9	n.d.	n.d.	n.d.	0.35	n.d.
18	36	F	255	2.04	3.21	2.97	n.d.	Positive
C	26	F	15.5	n.d.	n.d.	n.d.	n.d.	Negative

^a Phadia CAP-FEiA System; n.d., not defined.

Until now, known allergens in tomatoes were identified exclusively in tomato fruit which comprise all tissues like seed, peel, pulp, placenta and pericarp (see Figure 1). This paper presents the analysis of different constituent parts of the tomato fruit by dot blot technique using 18 well-characterized sera from adult tomato-sensitized patients selected based on a positive history skin prick test (SPT) and specific IgE-levels. Serum from a nonallergenic donor was used as a negative control. Interestingly, the tomato seed fraction showed a strong dot blot signal. To identify corresponding tomato seed allergens, we used a proteomic approach. For protein separation two-dimensional chromatography using chromatofocusing (CF) for the first dimension and reversed phase (RP) chromatography in the second dimension was performed. The protein fractions were subjected to SDS-PAGE for further separation. Potential allergens were detected by IgE immunoblotting and analyzed using LC-MS/MS. A legumin- and a vicilin-protein were identified as IgE-cross-reactive tomato seed proteins as new putative allergens showing a significant homology to other previously reported food allergens. EST/Contig sequence alignments combined with tryptic peptide coverage analysis revealed a novel full-length vicilin protein in tomato. Sequence alignments of these 2 putative tomato allergens were performed for homology search to known allergens (<http://www.allergenonline.com/>) and structural modeling of epitopes.

Experimental Section

Material and Reagents. The protein prestain standard was purchased from Bio-Rad (Hercules, CA). Nitrocellulose membrane was obtained from Serva (Heidelberg, Germany). An Alkaline Phosphatase (AP)-conjugated polyclonal goat anti-human IgE (1 mg/mL) and a monoclonal mouse anti-human IgG1 heavy chain antibody (0.52 mg/mL) were obtained from US Biological (MA). For the detection of the anti-human IgG antibody, an AP-conjugated polyclonal goat anti-mouse IgG antibody (1 mg/mL) was purchased from Sigma (Munich, Germany). As a positive control, an IgE-standard (1 µg/mL) was acquired from Biomol (Hamburg, Germany). Nitro blue tetrazolium chloride/5-Bromo-4-chloro-3-indolyl phosphatetolui-

dine salt (NBT/BCIP) was used as the substrate for the AP and was purchased from Roche (Mannheim, Germany).

For the mass spectrometric analysis, sequencing grade trypsin was purchased from Roche (Penzberg, Germany) and trypsin beads (Poroszyme) were purchased from Applied Biosystem (Darmstadt, Germany).

Preparation of Tomato Protein Extract. All tomato tissues samples used for this study were obtained from a local grocery in Potsdam. The tomato peel, pericarp, pulp, seeds and the placenta were manually separated from each other (see Figure 1). Seeds were dried for 1 h at room temperature. Dried tomato seeds were homogenized with liquid nitrogen in a ball mill (Retsch, Haan, Germany) for 45 s and a frequency of 30 s⁻¹. The tomato tissues were homogenized with a blender followed by grinding with liquid nitrogen and extracted with chaotropic buffer (0.1 M imidazol, 7 M Urea, pH 8). The ratio of tissue to buffer was 1:5 (w/v). After centrifugation (2500 rpm, 30 min) of the homogenates, protein was precipitated from the supernatant overnight by freezing (–20 °C) with acetone in a dilution of 1:5 (w/v). The pellet was centrifuged (2500 rpm, 30 min, 4 °C) and dissolved in 0.5 M Tris-HCl buffer (pH 7.8). Protein amount was determined using the Bradford assay.¹¹

Dot Blot Analysis. For the dot blot analysis, the resulting tomato extract pellets were diluted in 0.5 M Tris-HCl, pH 7.8. The solutions were centrifuged for 10 min at 15 000 rpm. Aliquots of 100 µg of protein extract for each distinct tomato tissue were blotted onto the nitrocellulose membrane.

The membrane was blocked with TBS/5% nonfat dry milk/0.1% Tween 20 buffer at 4 °C overnight. The following incubation steps were performed at room temperature. First, the membrane was washed 3 times for 5 min with TBS/0.1% Tween 20 buffer (TTBS). Afterwards, the membrane was incubated with patient sera (dilution 1:20 (v/v) in TBS/2.5% nonfat dry milk/0.1% Tween 20) for 1.5 h. After washing three times for 5 min each in TTBS, the membrane was incubated with goat anti-human IgE (AP) (1:1000 v/v) for 1 h and washed three times for 5 min.

Protein Separation by CF- and RP-Chromatography. For the two-dimensional protein separation of tomato seed proteins, the protein extract was diluted in chaotropic buffer (0.1 M imidazol, 7 M Urea, pH 8). The homogenate was centrifuged at 15 000g for 20 min at 4 °C and filtered through a 0.45 µm filter (Schleicher & Schuell, Mannheim, Germany). Five milliliters of supernatant (3 mg/mL protein content) was loaded onto a two-dimensional chromatographic system (PF2D from Beckman-Coulter Proteome Laboratory). The first dimension of the protein fractionation system consists of a single piston pump (HPCF Module), manual injector for sample introduction, pH monitor, and an UV detector. First-dimension separation utilized a chromatofocusing (CF) column (250 × 2.1 mm i.d.) that generated a pH gradient from 8 (0.1 M imidazol, 7 M Urea, pH 8, with iminodiacetic acid, start buffer) to 4 (eluent buffer). The eluent buffer was performed in a 1:10 ratio of polybuffer 74 (GE Healthcare, NJ) in 7 M Urea with a pH 4 with iminodiacetic acid. Finally, a wash buffer was used comprising 1 M NaCl at a flow rate of 0.2 mL/min. Fractions were collected in 0.5 pH intervals into 96-deep well plates, resulting in 20 fractions ranging in size from 0.1 to 1 mL. The first-dimension fractions were further separated in the second dimension using a C18 reversed-phase (RP) column (33 mm × 4.6 mm i.d.; 1.5 µm nonporous silica beads derivatized with C18) utilizing an acetonitrile (ACN)/water gradient starting from 100% water to 100% ACN at a flow rate of 0.75 mL/min over

Table 2. SRMs of the Heavy Peptides Le2, Le4, Vi2 and Vi3^a

protein	peptide	peptide sequence	mass (Da)	precursor ion [<i>m/z</i>]	product ion [<i>m/z</i>]	collision energy (V)
Legumin TC156005	Le2	AGQNGFEFAVFR	1351.6	676.8	1096.5	21
					649.3	27
	Le4	GGESFLLSPQR	1196.6	599.3	487.2	33
Vicilin AM932874	Vi2	SNNPYLFESQR	1360.6	681.3	600.3	17
					1046.5	24
	Vi3	NSFNLQKGDVIR	1396.7	699.3	666.3	23
					559.1	19
					387.2	34

^a Heavy amino acids are labeled in bold type.

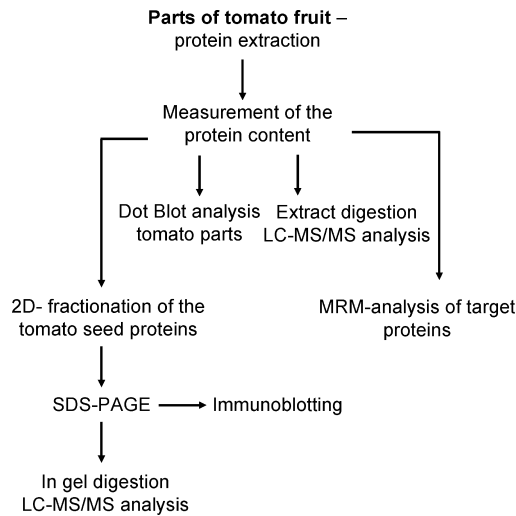


Figure 2. Sample preparation scheme for the identification of new tomato seed allergens.

110 min. A total of 500 μ L of each first-dimension fractions was loaded onto the second dimension. Protein separation was monitored by UV absorbance at 280 nm. The hardware is controlled by 32 Karat software.

Serum Samples and Skin Prick Testing. Serum samples were obtained from patients with a positive history of tomato allergy after written informed consent at the Allergy-Centre-Charité, Charité-Universitätsmedizin Berlin, Germany. For negative control, a serum from an individual lacking a positive history and having negative SPT and negative CAP-FeIA-test reactivity was used (see Table 1, serum C). CAP-FeIA is a blood test for detecting specific IgE antibodies against allergens. Measurement of allergen-specific IgE was performed with the CAP-FeIA system (Phadia, Uppsala, Sweden), according to the manufacturer's instruction. The sera were kept at -20°C until use. The SPT was performed with native tomatoes. A wheal >3 mm was considered as a positive reaction,¹² sodium chloride as negative and histamine (1 mg/mL) as positive control. The study was approved by the local ethical committee, Charité-Universitätsmedizin Berlin, Germany.

IgE Western Immunoblotting. Tomato protein samples were prepared for 1-D SDS-PAGE by boiling the samples in a Laemmli sample buffer¹³ for 5 min. Proteins were separated on 15% SDS gels and stained with Coomassie blue as control and for further LC-MS/MS analysis and blotted on nitrocellulose membrane for immuno analysis. Nitrocellulose membranes containing blotted proteins and prestained protein standard were blocked overnight in TBS/5% nonfat dry milk/0.1% Tween 20 at 4°C . The following incubation steps were performed at room temperature. The membranes were washed

three times for 5 min each in TBS/0.1% Tween 20 (TTBS) and incubated with patient sera and control sera from a nonallergic individual, respectively (dilution 1:20 (v/v) in TBS/2.5% nonfat dry milk/0.1% Tween 20) for 1.5 h on a shaker. After washing three times for 5 min each with TTBS, the membranes were incubated with goat anti-human IgE (AP) for 1 h, washed three times for 5 min, and exposed with 10 mL of NBT/BCIP (Bio-Rad, CA) for 30 min.

Nano-LC-MS/MS Analysis. For tryptic digestion, the complex tomato protein mixture of each tissue was solubilized in 0.5 M Tris-HCl, pH 7.8, and digested with 2 μ L of trypsin beads (Porozyme) per 10 μ g of protein for 16 h at 30°C . After centrifugation, the beads and insoluble material were removed and the tryptic peptides were separated through a Chromolith CapRod RP18e, 0.1 mm i.d. capillary column (Merck, Darmstadt, Germany) with a nano-HPLC (Agilent 1100 Series LC, Böblingen, Germany) and analyzed by an LTQ linear ion trap mass spectrometer (Thermo Finnigan, Waltham, MA).

Protein gel bands corresponding to detected bands in immunoblotting were excised and digested for 16 h at 30°C with trypsin as described previously.¹⁴ In brief, peptides were extracted from the gel by passive diffusion in three consecutive steps using increasing percentages of acetonitrile (5, 50, and 90%, each containing 1% formic acid). A Chromolith CapRod RP18e, 0.1 mm i.d. capillary column with a liquid chromatography-MS system comprising a nano-HPLC instrument and an LTQ linear ion trap mass spectrometer was used for analysis. Eluting peptides of complex tomato mixture and protein gel bands were continuously analyzed by selecting the three most abundant ion signals of a survey scan (mass-to-charge ratio range of 500–2000) for sequential MS/MS fragmentation. The MS/MS spectra were searched against a tomato nucleotide database (<http://www.tigr.org>) using Turboquest¹⁵ implemented in Bioworks 3.1 (Thermo-Finnigan). Matches with minimum Xcorr of 1.9, 1.5, and 3.3 for singly, doubly, and triply charged tryptic peptides, a minimum correlation of 0.1, and a minimum of 2 different peptides per locus were retained.¹⁶ Using DTA Select,¹⁷ all redundancy of scans, peptides and proteins were removed. Additionally, spectra were evaluated manually using the DTA Select graphical user interface (GUI). All the spectra can be visualized for quality assessment using the mass spectral reference database PROMEX (<http://promex.mpimp-golm.mpg.de/home.shtml>).¹⁸

Protein Extraction from Commercial Tomato Sauce. Protein from commercial tomato sauce was extracted using a 80% ammonium sulfate precipitation at 4°C for 1 h. After centrifugation (3400 rpm for 30 min), the pellet was desalted with small NAP-columns (GE Helathcare, Buckinghamshire, U.K.) in 0.5

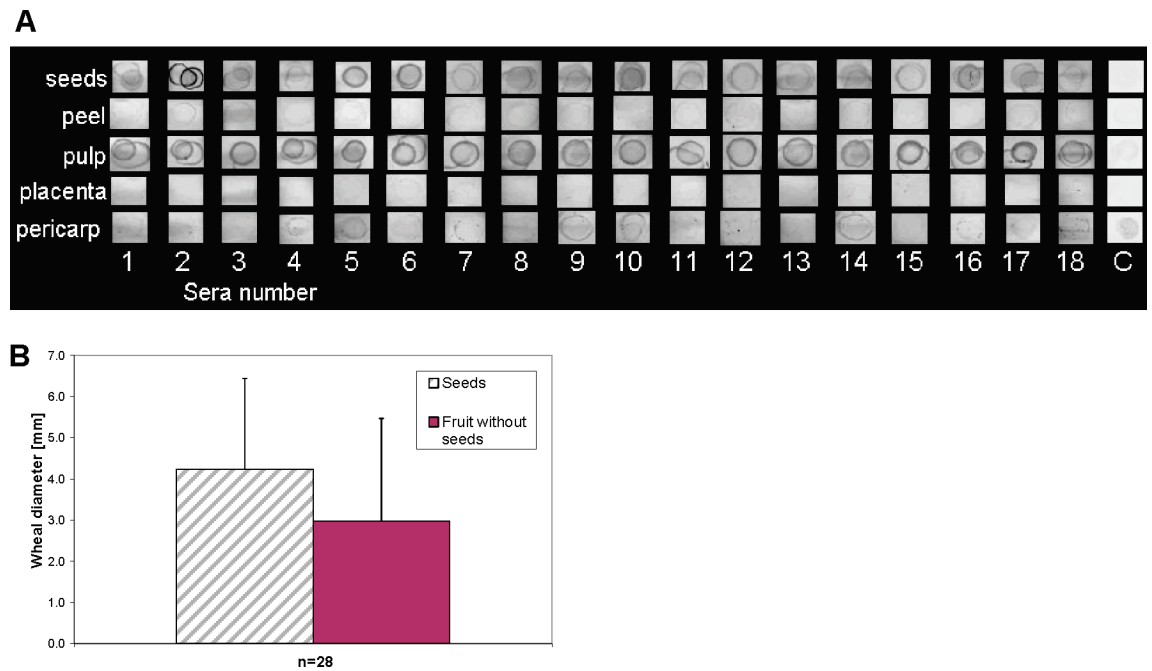


Figure 3. (A) IgE-dot blot of different constituent parts of the tomato fruit (seed, peel, pulp, placenta and pericarp) using sera from 18 tomato allergenic patients and a healthy individual (C). (B) Skin prick test (SPT) with seed protein or fruit-without-seed- protein ($n = 28$).

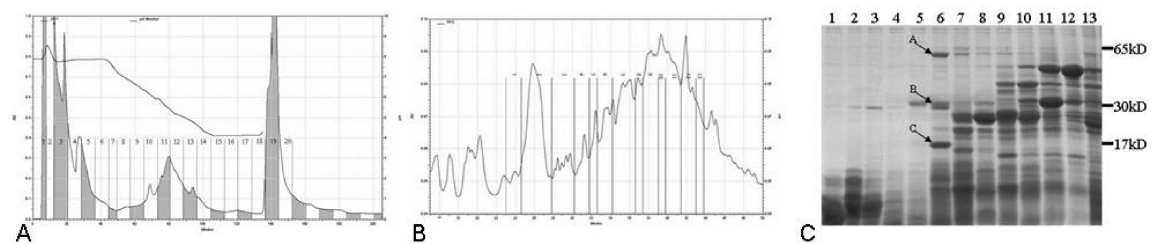


Figure 4. CF-chromatogram (A) and RP- chromatogram of the CF-fraction 3 (B) and the corresponding SDS PAGE (C) of the RP-fractions of the tomato seed extract.

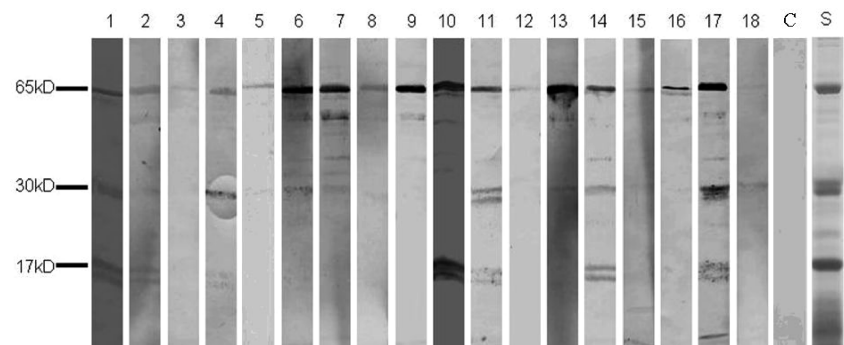


Figure 5. IgE immunoblot of the RP fraction number 6 with 18 patient sera (1–18), a control serum (C) and die Coomassie stained SDS lane (S). An individual pattern against the 65, 30 and 17 kDa protein bands is observed for each patient.

M Tris-HCl buffer, pH 7.8, and digested with 2 μ L of trypsin beads (Porozyme) per 10 μ g of protein for 16 h at 30 $^{\circ}$ C.

Single Reaction Monitoring (SRM) of Putative Tomato Allergens in Commercial Tomato Sauce. To identify and quantify the proteins legumin and vicilin in commercial tomato products, stable isotope-labeled synthetic peptides of the corresponding tryptic peptides of the protein amino acid sequences were used as internal reference peptides in mass spectrometric analysis. The method follows recent protocols for targeted identification and quantification of proteins/

α -unit (~30kDa)
SLAMAVSTRVLLALLFSAFLLSAANAYRQYPPQGGHGORSIPLTQAQCCRLQRLTG
SQPSQRIESEGgyTEVVDENEQFQCVGVAPMRNVIRQNSLSLPNFHPMPRLVYIEQG
RGLISVINPGCAETFSQSTFQAGRMAGEERGERRNDQHKVHRIRQGDVVAIPAGA
AHWCYNDGEELVAVSINDLNHRSNQLDQNFRAFYLAGGVQETSGRQIGVGKQSTQ
KFQNIIFRAFNTLMAEAFNPVEIVRRMQEEQNERGLIVNVREGMSMIRADEEEEEFE
GRSQRGRWWAEVTG

β -unit (~17kDa)
GLEENICTMKIRTNMDTQRRADIYSRQACKIKHVNQNFILRDMDSASKGTLYPN
ALMSPHWSVNGHSVTVYVQRGDAQVQVVDHTGOQVMNDRVNQGEFVVPQYYAST
VKAGQNGFEFAVFTSSQPMNSQLAGYTSVIRAMPLEVLITNAYQISPNEAQRLEKMR
GGESFLLSPQRESF

Figure 6. Putative cleavage site in the tomato legumin protein resulting in a α - and a β -unit (gray: identified peptides).

Table 3. Identified Proteins via Nanoflow Liquid Chromatography Coupled with Mass Spectrometry after RP Fractionation

	accession number	proteine name	peptide sequence	charge	Xcorr	peptide no.
65 kDa band	AM932874	Vicilin	GRTQGPFNLM*K	2	2.1926	10
			SNNPYLFESQR	2	2.7999	
			GQGVINIAEQDNK	2	3.4311	
			IVGFGVDAENNK	2	3.6676	
			QDQSYFVAGPEHR	2	2.8716	
			QDQSYFVAGPEHR	3	3.5279	
			LFVYVLAK	2	2.2196	
			LVMVIEGNAR	2	2.711	
			EVEEIFQR	2	2.6038	
			NSFNLQKGDVIR	2	2.814	
			AFYLAGGVQETSGR	2	2.6927	
			LVYIEQGR	1	1.9012	
			SNQLDQNFRR	2	2.9054	
30 kDa band	TC156005	Legumin-like protein (α -chain)	LTGSQPSQR	2	1.8133	10
			NVIRQNSLSLPNFHPM*PR	3	4.4289	
			QNSLSLPNFHPM*PR	2	2.7236	
			QNSLSLPNFHPM*PR	3	3.0206	
			QSTQKFQNFRR	3	2.6102	
			FQNFRR	2	2.0641	
			QEEQNERGLIVNVR	3	3.4864	
			GGESFLLSPQR	2	3.4553	
			HVNSQNFPIRR	2	3.6679	
			IKHVNSQNFPIRR	3	3.5734	
17 kDa band	TC156005	Legumin-like protein (β -chain)	AMPIEVLTNAYQISPNEAQR	2	2.2502	5
			AGQNGFEFAVFR	2	2.9657	

peptides in complex samples.^{19,20} For identification and quantification, peptide-multiple-reaction-monitoring (MRM) on a triple stage quadrupole mass spectrometer (TSQ Quantum, Thermo Fisher Scientific, San José, CA) was used. Stable isotope labeled internal peptide standards (¹³C-leucin) (Thermo Fisher Scientific, Dreieich, Germany) were synthesized. Peptide sequences were derived from legumin and vicilin sequences identified in previous mass spectrometric analyses. Peptide sequences, precursor and product ions and the corresponding collision energies for collision-induced-dissociation (CID) are listed in Table 2. All the spectra are stored in ProMEX, a reference library for mass spectra which can be used to design such internal stable-isotope labeled peptides.¹⁸ Mass spectrometry was performed on a TSQ Quantum Discovery MAX mass spectrometer (Thermo Fisher Scientific, San José, CA) operated in the positive ion mode. The scan width for all MRMs was 0.7 mass units. The resolution for Q1 was 0.2 mass units; the resolution for Q3 was set to 0.7 mass units. The mass spectrometer was tuned to its optimum sensitivity for each standard peptide as described in detail previously.²¹ The dwell time per transition was 50 ms. For chromatography, a 1D nanoflow LC system with precolumn (Agilent, Germany) was used. A monolithic column (Merck, Darmstadt, Germany) of 15 cm length and an i.d. of 0.1 mm was coupled with the TSQ. Peptides were eluted during a 30 min gradient ranging from 5% MeOH to 100% MeOH in 0.1% FA with a flow rate of 0.7 μ L/min. A total of 100 fmol per standard peptide was added to each sample digest prior to analysis.²²

Homology Alignment Analysis of the Identified Legumin and Vicilin Proteins with Known Allergens. The identification of homologues of the identified putative protein allergens was performed using genebee (<http://www.genebee.msu.su/index.html>) and searches against well-known allergens from the FARRP-database (<http://www.allergenonline.com/>). The FARRP-database contains a comprehensive list of unique proteins of known and putative allergens and allows a direct protein

alignment. The results were confirmed using the allergome homepage (<http://www.allergome.org/>).

Structural Epitope Analysis of the Identified Legumin and Vicilin Proteins in Comparison to Known Allergens. For analysis of the three-dimensional structural properties of the putative tomato allergens (TC156005, AM932874) and the known homologous allergens Anao2 and Arah1, HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>) was used for finding proteins with similar structures. HHpred is a fast server for remote protein homology detection and implements pairwise comparison of profile hidden Markov models (HMMs).²³ On the basis of the amino acid sequences of the proteins TC156005, Anao2 (gi 25991543), AM932874 and Arah1 (gi 1168390), HHpred found the following structures which we used as templates for the homology modeling: 1fxz, 2evx, 2d5f, 2cav and 2ea7 for TC156005; 1fxz, 2evx, 2d5f, 2cav and 1uij for Anao2; 2ea7, 1uij, 2cav, 2phl and 2evx for AM932874; 2ea7, 1uij, 2cav, 2phl and 1fxz for Arah1 (pdb-codes). Calculation of tertiary structural models based on multiple templates of the proteins was performed using the program MODELER which returns a file containing the atomic coordinates in the Protein Data Bank (PDB) format.^{24,25} The structure of the proteins TC156005 and AM932874 were superposed with the structures of the proteins Anao2 and Arah1 which contain known epitopes. The regions which structurally correspond to the known epitopes were labeled in the tomato proteins TC156005 and AM932874.

Results and Discussion

Analysis of the Tomato Fruit Using IgE-Dot Blot Technique and Skin Prick Testing (SPT). In Figure 2, the work flow of the present study is shown. The IgE-dot blot technique using different patient sera against different parts of the tomato fruit allows a fast prescreening of patient-specific allergen profiles. For that purpose, 100 μ g of tomato protein of each tissue extract were blotted onto a nitrocellulose membrane. The

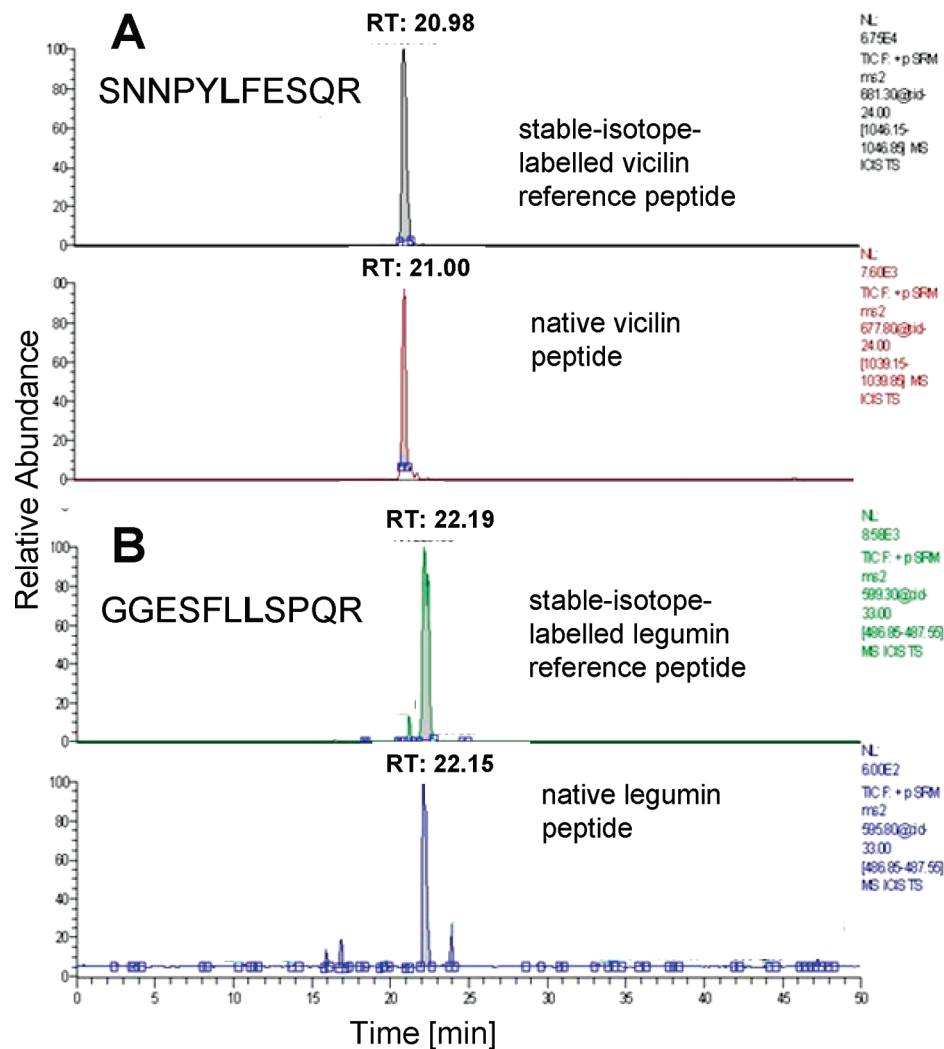


Figure 7. LC-Triple-Quadrupole-MRM-MS analysis of putative tomato allergens in commercial tomato sauce. (A) MRM-extracted ion trace chromatogram for the stable-isotope-labeled vicilin reference peptide (upper part) and the native tryptic vicilin peptide (lower part). (B) MRM-extracted ion trace chromatogram for the stable-isotope-labeled legumin reference peptide (upper part) and the native tryptic legumin peptide (lower part).

Table 4. Identity and Similarity between the Identified Putative Allergenic Proteins and Known Allergens

protein	allergen	organism	identity [%]	similarity [%]
Legumin protein (TC156005)	Ber e 2 11S globulin (gi 30313867)	brazil nut	45.5	76.5
	Ana o 2 legumin (gin)	cashew nut	43	77
	Fag e 1 legumin (gi 29839419)	buckwheat	49	74
Vicilin-like protein (AM932874)	Jug n 2 vicilin (gi 31321944)	black walnut	47	81
	Ara h 1 vicilin (gi 1168390)	peanut	30.5	66.5

dot blots were measured in triplicates. In this study, 18 characterized patient sera were used (see Table 1). Bound IgE was detected using an anti-IgE second antibody.

In the IgE dot blot experiment (see Figure 3A), about 4 patients showed a weak IgE recognition signal against the tomato pericarp. No signal against the tomato pericarp was found with the other serum samples. No patient sera exhibited a signal against the tomato placenta tissue and the peel of the tomato, whereas tomato pulp extract was recognized by all patient sera. Interestingly, the tomato seed extract showed significant IgE reactivity for about 10 of 18 allergic patient sera (see Figure 3A). As a control, the tomato tissue protein blot was incubated with serum from a nonallergic individual in the same dilution (1:20) (see lane C in Figure 3A). Bound serum

IgE antibodies were detected using a second anti-IgE (AP) antibody. The negative control shows no IgE-reactivity in the dot blot.

In Figure 3B, skin prick test (SPT) results are shown using isolated seed protein or only tomato fruit protein after separation from seeds. Intriguingly, the significance of the allergic reaction against the seed protein only is high or even higher compared to the tomato fruit protein measured by the wheal diameter after the reaction. This result indicates the relevance of seed proteins as a source for potential allergens. Therefore, in this study, we focused on the isolation of the IgE-reactive parts from tomato seed proteins.

Analysis of Tomato Fruit Proteins Using LC-MS/MS. We performed LC-MS/MS analysis of complex protein samples of

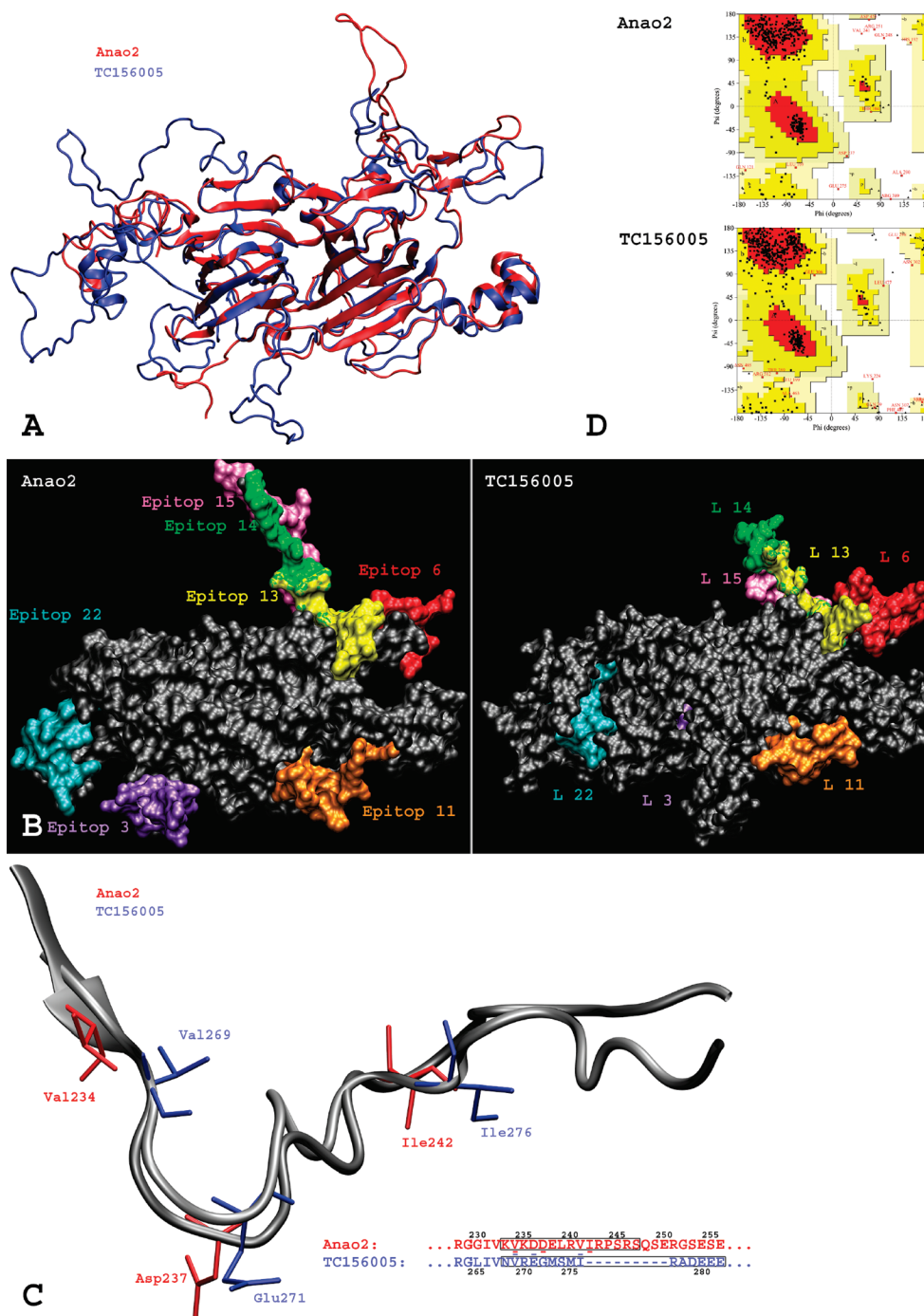


Figure 8. Ribbon diagrams of the walnut Anao2 allergen and tomato legumin (TC156005) (A), with putative epitopes (B), epitope 13 of Anao2 (red) and the homologous sequence of TC156005 protein (blue) shown in detail (C), a plot of the Φ and Ψ angles for the amino acids in Anao2 and TC156005 (D).

the different tomato tissues. In the pericarp protein extract, a profilin (TC162738) and a pathogenesis related protein (TC162694) were detected. However, no sequence homology between the known profilin Lyc e 1 allergen and the identified profilins is found. The identified pectinesterase protein (TC162057) within the tomato pericarp is assumed to be an allergen.⁶ The pectinesterase was identified by peptide alignment found in the Kondo study. In the placenta protein extract, no protein with homology to known tomato allergens was identified, which correlates with the absence of an IgE signal in the dot blot. In contrast, the pulp protein extract shows a high IgE cross-reactivity to all patient sera in the dot blot experiment

correlating with the presence of Lyc e 2 β -fructofuranosidase (TC161907), a known tomato allergen. In the tomato peel, no protein sequence homologies to other tomato allergens were detected. In the tomato seed extract, no known tomato allergen protein sequence was identified, although a clear signal in the dot blot was observed. We performed further protein fractionation and immunoblot analysis to identify putative allergens in the tomato seed extract (see next section).

Analysis of the Tomato Seed Protein Fraction Using Two-Dimensional Protein Chromatography. Tomato seeds are a rich source of proteins (22–33% on a dry weight basis),²⁶ where storage proteins like globulins constitute about 70% (on

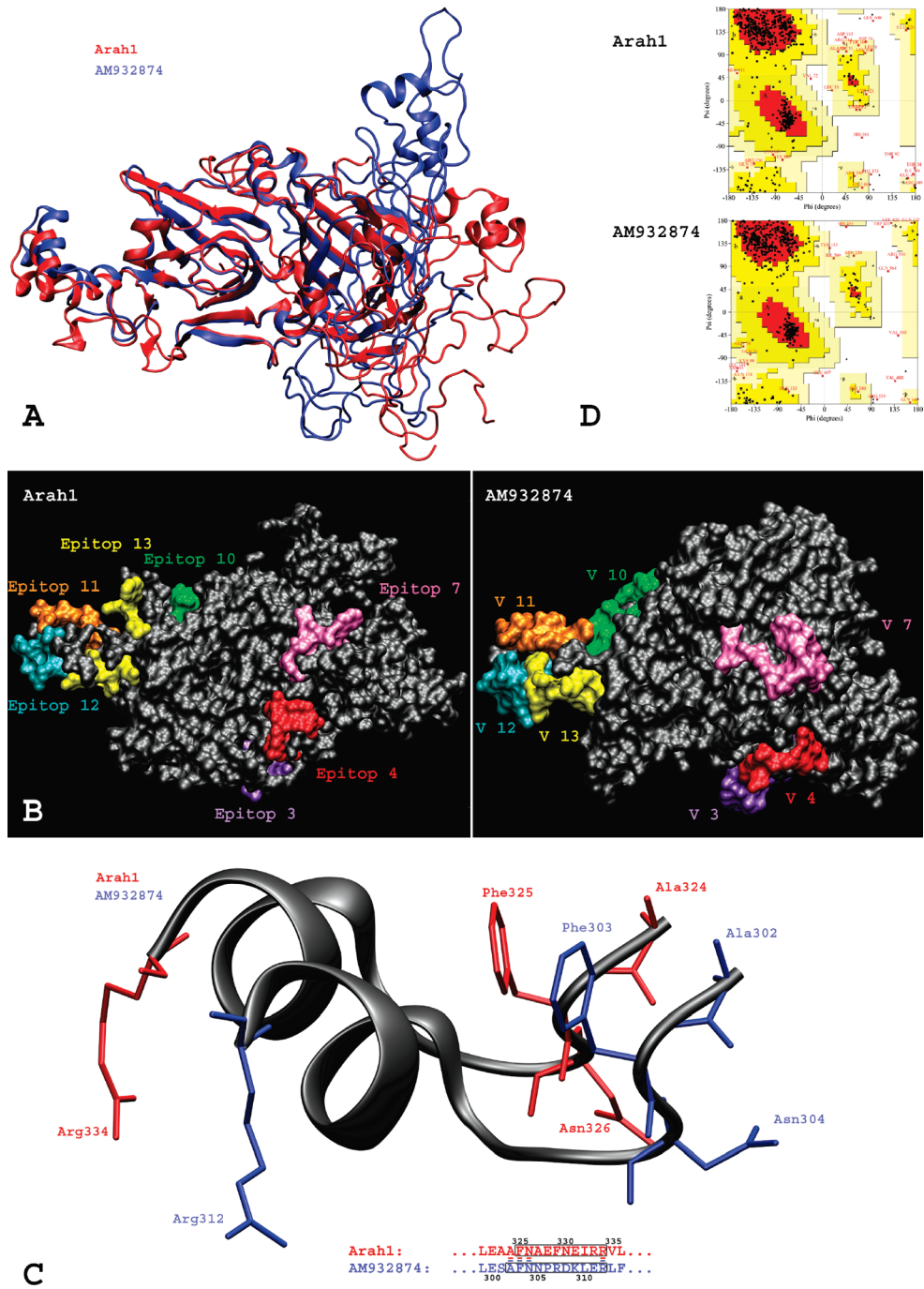


Figure 9. Ribbon diagrams of the walnut Arah1 allergen and tomato vicilin (AM932874) (A), with putative epitopes (B), epitope 12 of Arah1 (red) and the homologous sequence of AM932874 protein (blue) shown in detail (C), a plot of the Φ and Ψ angles for the amino acids in Arah1 and AM932874 (D).

a dry weight basis).²⁷ To cope with the high dynamic range and complexity in the seed proteome, we performed multidimensional protein fractionation in combination with immunoblotting techniques. A two-dimensional chromatographic system consisting of CF- and RP-chromatography (PF2D from Beckman-Coulter Proteome Laboratory) was used. For separation of the tomato seed proteins, 5 mL of filtered extract was loaded onto the column with a protein content of 3 mg/mL. The tomato seed extract was diluted in a chaotropic buffer containing 7 M Urea. Fractions were collected in 0.5 pH intervals into 96-deep well plates, resulting in 20 fractions ranging in size from 0.1 to 1 mL (see Figure 4A). IgE-Western

blot revealed three dominant protein signals in the fractions 3–6 with a molecular weight of 65, 30 and 17 kDa in a pI-range from 7.9 to 7.7 (data not shown). The bands were excised from the SDS-gel and analyzed by LC-MS/MS (see supplemental 1 in Supporting Information). In the 65 kDa band, several proteins ranging from an elongation factor to a vicilin precursor were identified. In the 30 kDa band, legumin and LEA proteins were identified. Within the major immuno-detected 17 kDa protein band, three isoforms of the legumin-group (TC156005, TC169589 and TC168461) and a pathogenesis-related protein (TC153878) were identified by several significant peptides. To clarify which specific proteins are responsible for the IgE-

Table 5. Linear Ana o 2 Epitopes and Homologous Amino Acid (aa) Range of the Tomato Legumin Allergen (TC156005)^a

epitope Ana o 2	aa position Ana o 2	aa sequence Ana o 2	aa position (TC156005)	aa sequence (TC156005)	label color
3	15–29	SRQEWQQQDECQIDR	40–54	RSIPLTQAQQCRLQR	Violet
6	105–119	YQAPQQGRQQGQSGR	130–150	FQSQTFFQAGRMAGEERGERRG	Red
11	185–199	VFQQQQQHQSRRNL	215–233	TSGRQIGVGKQSTQKFQNI	Orange
13	233–247	KYKDDWELRVIRPSRS	268–282	VNVREGMSMIRADEEE	Yellow
14	241–255	VIRPSRSQSERGSES	275–287	MIRADEEEEEFE	Green
15	257–271	EESEDEKRRWGQRDN	289–302	RSQRQRWWAEVTGN	Pink
22	425–439	FQISREDARKIKFNN	505–522	FAKMKQERLVFRSKCNE	Cyan

^a Shared amino acids are underlined.

Table 6. Linear Ara h 1 Epitopes and Homologous Amino Acid (aa) Range of the Tomato Vicilin Allergen (AM932874)

epitope Ara h 1	aa position Ara h 1	aa sequence Ara h 1	aa position (AM932874)	aa sequence (AM932874)	label color
3	65–74	LEYDPRLUYD	38–47	YQDPQEKLRE	Violet
4	89–98	GERTGRQPG	54–62	RQQPGQQKQ	Red
7	123–132	REREEDWRQP	78–93	QQQHGGGETGEDDLGNNR	Pink
10	294–303	TPGQFEDFFP	271–281	NAPGNLQEYFS	Green
11	311–320	SYLQEFSSNT	288–312	ESFYRAFSSDI	Orange
12	325–334	FNAEFNEIRR	302–312	AFNNPRDKLER	Cyan
13	344–353	EQEERGQRRW	314–323	FGQHKEGIII	Yellow

reactivity, further separation was necessary. Immunoreactive CF-fractions were further separated in the second dimension using a C18 reversed-phase (RP) column. A total of 500 μ L of CF-fraction 3 with a protein content of 1 mg/mL was loaded onto the column. The resulting chromatogram is shown in Figure 4B. Resulting fractions were applied on a 15% SDS-gel (Figure 4C). IgE-reactive protein bands were detected in fraction 6 by IgE-dot blot. Fraction 6 was tested with 18 individual sera from allergic patients and a control serum from a nonallergenic individual (see Figure 5). Three enriched and purified immunoreactive bands (17, 30, 65 kDa) were detected in the IgE-immunoblots. This experiment was performed in duplicates. Within the 65 kDa band, a single vicilin (AM932874) tomato protein, and in the 30 and 17 kDa band, a single legumin (TC165005) protein were identified. According to the definitions of the FAO/WHO,²⁸ the legumin is designated as a minor allergen because approximately 39% of the patient sera react in the immunoblot with the legumin bands (TC165005, 30 and 17 kDa). In comparison, 67% of the patient sera showed IgE-reactivity against the 65 kDa vicilin protein band.

The legumin protein TC165005 with a molecular weight of 17 kDa in the IgE immunoblot is most probably a legumin β -chain after proteolytic degradation (see Discussion below). In an epitope study of Yoshioka and co-workers, the β -chain of the major allergenic protein Fag e 1 from buckwheat with a molecular weight of 22 kDa was analyzed.²⁹ The β -chain of the buckwheat legumin shows a 49% identity to the putative β -chain legumin in this study (see supplemental 3 in Supporting Information). The 60 kDa buckwheat legumin contains an asparagin–glycin (N–G) cleavage site which is cleaved by an specific endopeptidase into a 22 and 38 kDa unit. The N–G cleavage site is conserved among prolegumins from diverse plant species and also found in the prolegumin sequence TC165005 identified in our study (see Figure 6).³⁰ A cleavage at the N–G peptide bond results in a product of 17 kDa (β -subunit) and a product of 30 kDa (α -subunit) which corresponds well with the observed molecular weights in our study (see Figure 5). The protein coverage of the β -chain of the legumin protein using LC-MS/MS is 41%. It is possible that due to degradation processes of the β -unit the legumin protein was detected as a double band in the immunoblot. The 30 kDa double band could be identified as the N-terminal α -unit of

the legumin protein (see Figure 6). The protein coverage of the α -chain of the legumin protein is 17%. However, in the IgE immunoblot, a double protein band is visible (see Figure 5). MS-analysis of both protein bands revealed a degradation product of the same protein. The immuno-detected band with a molecular weight of 65 kDa matched to a vicilin protein with two accessions (TC164469 and BE458586, Supporting Information Table 1) (see Table 3). The protein coverage is 17%. All resulting spectra corresponding to the identified proteins in Table 3 are listed in the public ProMEX database (<http://promex.mpimp-golm.mpg.de/home.shtml>).¹⁸

The heat-stable legumin and vicilin proteins are the main storage proteins in angiosperms and gymnosperms representing also major food allergens in several other plant seeds. One of the most common allergen group is the cupin superfamily comprising the major globulin storage proteins from legumes and nuts. Globulins have been found to be highly relevant allergens in peanut, soybean, buckwheat, walnut, hazelnut and sesame.^{29,31–33} A comparable protein pattern of globulins (7S and 11S globulin) was also demonstrated via immunoblot characterization of cashew nut allergens.³⁴ Via immunoblot, Wang and co-workers identified the Ana o 1 (vicilin) and the Ana o 2 (legumin) allergen.^{35,36}

Prediction and Mass Spectrometric Support of a Putative Full-Length Tomato Vicilin-like Protein. Two distinct sequences for the vicilin protein (BE458586 and TC164469) can be assigned to the expressed sequence tags (EST) 413878,³⁷ 278357,³⁷ and gi124222276.³⁸ Those three EST's were aligned against a tomato-contig_TC164469 derived from the Gene Index Databases (<http://compbio.dfci.harvard.edu/tgi/plant.html>). The alignment revealed identities between the full-length contig and the EST's in different parts of the nucleotide sequence (see Supplemental 2 in Supporting Information). The detection of peptides along the contig and the inner part of the sequence by LC-MS/MS confirmed the predicted protein sequence (see Table 3, sequence coverage ~20%). On the basis of these results, a predicted full-length sequence of the tomato vicilin protein was submitted to the EMBL Nucleotide Sequence

Database (<http://www.ebi.ac.uk/>) and annotated with the accession number AM932874.

Identification of the Tomato Seed Proteins Legumin and Vicilin in Commercial Tomato Sauce. Tomato sauce was analyzed for the presence of seed proteins, especially the putative allergens detected in this study, to investigate whether there is a relevance of those putative allergens for tomato-allergic patients. Because of the lack of antibodies specific for the detected legumin and vicilin protein isoforms, we used a mass spectrometric multiple-reaction-monitoring approach (MRM) with internal stable-isotope-labeled reference peptides.^{19,20} To quantify legumin and vicilin, corresponding tryptic peptides of the detected polypeptide were synthesized using stable-isotope-labeled amino acids and spiked as internal standards into the digested protein extract of tomato sauce. This complex protein extract was further analyzed by HPLC coupled with a triple quadrupole mass spectrometer. The MRM experiment during the chromatography was performed by specifying the parent mass of the tryptic peptide for MS/MS fragmentation and then specifically monitoring for a single fragment ion (see Figure 7). With the use of this method, both proteins were detected with approximately 40–50 pg per 50 mL of tomato sauce. Accordingly, the presence of these putative allergens in tomato sauce was demonstrated. Their presence is most probably caused by a harsh preparation procedure of the tomato sauce during the production process. However, due to the relevance for allergic patients, these results are very preliminary and need more systematic investigations in the future. Furthermore, for peanut allergens, several orders of magnitude higher concentrations are normally considered as relevant.

Sequence Homology of the Identified Legumin and Vicilin Proteins with Known Allergic Proteins. According to the FAO/WHO guidelines,^{28,39} cross-reactivity between the expressed protein and a known allergen has to be considered in case there is more than 35% identity in the amino acid sequence of a segment of at least 80 residues. The AllergenOnline database entries were compiled primarily from entries identified by searching publicly available protein databases using the Entrez (NCBI) search and retrieval system based on keyword and protein sequence searches (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>).

Searching the FARRP-database with the immuno-detected legumin (TC156005) allergen candidate resulted in three significant hits. The highest identity (45.5%) was detected to the Ber e 2 legumin allergen (gi30313867, supplemental 3 in Supporting Information) from brazil nut (*Bertholletia excelsa*). Furthermore, a 43% identity was found to the Ana o 2 allergen (gi25991543) from cashew nut (*Anacardium occidentale*) and a 49% identity to the Fag e 1 (gi29839419) allergen (see Table 4). In the study of Yoshioka et al., the major allergenic protein Fag e 1 from buckwheat was analyzed.²⁹ This 22 kDa allergen is the β -subunit of the 13S globulin and reveals a 49% identity (see Table 4, Supplemental 3A in Supporting Information) to the putative tomato legumin β -unit with a molecular weight of 17 kDa identified in our study. The vicilin (AM932874) with a molecular mass of approximately 65 kDa shows high identity to the vicilin seed storage allergen Jug n 2 (gi31321944) from black walnut (*Juglans nigra*) (see Table 4, Supplemental 3B in Supporting Information). In Supplemental 4 (Supporting Information), the alignment of the sequence of the allergic black walnut protein with a 47% identity to the vicilin (AM932874)

protein is shown. Furthermore the allergen Ara h 1 (gi 1168390) from peanut (*Arachis hypogaea*) shows a 30.5% identity and 66.5% homology to the tomato vicilin protein (see Table 4).

Structural Epitope Alignment with the Identified Legumin and Vicilin Proteins. Tertiary structure patterns of the putative tomato allergens legumin (TC156005) and vicilin (AM932874) and the corresponding homologous allergens Ana o 2 (gi25991543) and Ara h 1 (gi 1168390) were generated as pdb-files using the program MODELER (see Figures 8A and 9A and Tables 5 and 6). In an epitope study of Wang et al., seven major linear epitopes were detected with two out of three serum pools using synthetic peptides immobilized on a membrane (SPOT-technique).⁴⁰ A space-filling model depicting the surface accessibility of the seven epitopes of the Ana o 2 allergen and the appropriate putative epitopes is shown in Figure 8B. The putative epitopes were derived by comparing the structures of TC156005 and Ana o 2. The linear epitope 13 of the Ana o 2 allergen and the appropriate structure of the tomato legumin loop is shown in detail in Figure 8C. The picture shows the high structural similarity between the Ana o 2 epitope and the tomato legumin. All amino acid epitope sequences of the Ana o 2 allergen and the appropriate amino acids analyzed by computational alignment of the putative tomato legumin (TC156005) are listed in Table 5. Epitopes 6, 11, 13, 14 and 15 of the Ana o 2 allergen exhibit sequence and structure homology to the corresponding tomato legumin putative epitopes, whereas for the putative epitopes 3 and 22 corresponding to L3 and L22, no structural similarity was evident (see Figure 8B).

Furthermore, a study of Shin et al. demonstrates 23 linear IgE binding epitopes of a Ara h 1 vicilin allergen using synthetic peptides probed with pooled serum from 15 patients (see Table 6).^{41–43} These Ara h 1 epitope sequences were used for putative epitope search on the surface of the putative tomato vicilin allergen as well. A space-filling model depicting the surface accessibility of seven epitopes and the appropriate putative epitopes of the tomato allergen vicilin is shown in Figure 9B. We choose seven out of the 23 linear epitopes of the Ara h 1 allergen to show the structural homology between Ara h 1 and the putative vicilin allergen. In our study, not all epitopes described in the Shin study align well with the tomato structure. The linear epitope sequences of the Ara h 1 allergen and the appropriate amino acid sequences of the putative tomato vicilin (AM932874) are listed in Table 6. The structures of the Ara h 1 linear epitope 12 and the corresponding fragment of the tomato vicilin are shown in detail in Figure 9C. After the structural epitope analysis of the putative tomato allergens legumin and vicilin it is evident that the epitope domains are rather located in variable loop regions than in α -helices or β -sheets. These results agree well with the observation that epitopes are usually associated with turns or loops which are primarily located on the surface of proteins.⁴⁴ With these investigations, we could show that the identified tomato proteins show high sequence, structural and epitope homology to known allergens, thus, indicating novel allergens according to the FAO/WHO guidelines.²⁸ Furthermore, it has been recognized that immunologic cross-reactivity of allergens is a common fact, especially in food allergy.^{45,46} Our results suggest that the proteins legumin and vicilin are strong candidates for tomato allergens. Future studies will focus further on experimental and structural epitope analysis of the identified proteins.

Acknowledgment. This study was part of the German BioProfile Nutrigenomik project “Allergen-Profiling in crop plants and plant-derived food using novel proteomic technologies”. The financial support provided by a grant from the German Federal Ministry of Education and Research (BMBF) is greatly acknowledged. The authors thank Patrick May, Dirk Walther and Stefan Kempa for helpful comments on the manuscript. We thank Eva Marie Fiedler for the serum characterization.

Supporting Information Available: Supplemental 1, identified proteins via nanoflow liquid chromatography coupled with mass spectrometry from CF dimension fractions; Supplemental 2, nucleotide sequence alignment of AM932874 (vicilin contig TC164469) against EST's (EST413878, EST278357 and gil24222276); Supplemental 3A, protein sequence alignment between legumin (TC156005) and the Brazil nut legumin Ber e 2 (*B. excelsa*, gi 30313867) allergen; Supplemental 3B, protein sequence alignment between the putative legumin β -unit (TC156005) and the Fag e 1 allergen; Supplemental 4, 47% homology between the tomato vicilin (AM932874, contig TC164469) and the walnut vicilin Jug n 2 (*J. nigra*, gi 31321944), red, identical amino acids. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Westphal, S.; Kempf, W.; Foetisch, K.; Retzek, M.; Vieths, S.; Scheurer, S. Tomato profilin Lyc e 1: IgE cross-reactivity and allergenic potency. *Allergy* **2004**, *59* (5), 526–532.
- Ortolani, C.; Ispano, M.; Pastorello, E.; Bigi, A.; Ansaloni, R. The oral allergy syndrome. *Ann. Allergy* **1988**, *61* (6 Pt 2), 47–52.
- Zuberbier, T.; Edenharter, G.; Worm, M.; Ehlers, I.; Reimann, S.; Hantke, T.; Roehr, C. C.; Bergmann, K. E.; Niggemann, B. Prevalence of adverse reactions to food in Germany—a population study. *Allergy* **2004**, *59* (3), 338–345.
- de Martino, M.; Novembre, E.; Cozza, G.; de Marco, A.; Bonazza, P.; Vierucci, A. Sensitivity to tomato and peanut allergens in children monosensitized to grass pollen. *Allergy* **1988**, *43* (3), 206–213.
- Beezhold, D. H.; Sussman, G. L.; Liss, G. M.; Chang, N. S. Latex allergy can induce clinical reactions to specific foods. *Clin. Exp. Allergy* **1996**, *26* (4), 416–422.
- Kondo, Y.; Urisu, A.; Tokuda, R. Identification and characterization of the allergens in the tomato fruit by immunoblotting. *Int. Arch. Allergy Immunol.* **2001**, *126* (4), 294–299.
- Palomares, O.; Villalba, M.; Quiralte, J.; Polo, F.; Rodriguez, R. 1,3-beta-glucanases as candidates in latex-pollen-vegetable food cross-reactivity. *Clin. Exp. Allergy* **2005**, *35* (3), 345–351.
- Weangsripanaval, T.; Nomura, N.; Moriyama, T.; Ohta, N.; Ogawa, T. Identification of suberization-associated anionic peroxidase as a possible allergenic protein from tomato. *Biosci. Biotechnol. Biochem.* **2003**, *67* (6), 1299–1304.
- Diaz-Perales, A.; Collada, C.; Blanco, C.; Sanchez-Monge, R.; Carrillo, T.; Aragoncillo, C.; Salcedo, G. Cross-reactions in the latex-fruit syndrome: A relevant role of chitinases but not of complex asparagine-linked glycans. *J. Allergy Clin. Immunol.* **1999**, *104* (3 Pt 1), 681–687.
- Reche, M.; Pascual, C. Y.; Vicente, J.; Caballero, T.; Martin-Munoz, F.; Sanchez, S.; Martin-Esteban, M. Tomato allergy in children and young adults: cross-reactivity with latex and potato. *Allergy* **2001**, *56* (12), 1197–1201.
- Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.
- Zuberbier, T.; Worm, M. Allergies and the skin, an interdisciplinary approach in GA(2)LEN and EAACI activities. *Allergy* **2006**, *61* (12), 1373–1376.
- Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227* (5259), 680–685.
- Otto, A.; Thiede, B.; Muller, E. C.; Scheler, C.; Wittmann-Liebold, B.; Jungblut, P. Identification of human myocardial proteins separated by two-dimensional electrophoresis using an effective sample preparation for mass spectrometry. *Electrophoresis* **1996**, *17* (10), 1643–1650.
- Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
- Peng, J. M.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J. Proteome Res.* **2003**, *2* (1), 43–50.
- Tabb, D. L.; McDonald, W. H.; Yates, J. R. DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **2002**, *1* (1), 21–26.
- Hummel, J.; Niemann, M.; Wienkoop, S.; Schulze, W.; Steinhäuser, D.; Selbig, J.; Walther, D.; Weckwerth, W. ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinf.* **2007**, *8*, 216.
- Lehmann, U.; Wienkoop, S.; Weckwerth, W. If the antibody fails—a mass Western approach. *Plant J.* **2008**, *55* (6), 1039–1046.
- Wienkoop, S.; Weckwerth, W. Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in Arabidopsis thaliana. *J. Exp. Bot.* **2006**, *57* (7), 1529–1535.
- Glinski, M.; Weckwerth, W. Differential multisite phosphorylation of the trehalose-6-phosphate synthase gene family in Arabidopsis thaliana—A mass spectrometry-based process for multiparallel peptide library phosphorylation analysis. *Mol. Cell. Proteomics* **2005**, *4* (10), 1614–1625.
- Wienkoop, S.; Larrainzar, E.; Niemann, M.; Gonzalez, E.; Lehmann, U.; Weckwerth, W. Stable isotope-free quantitative shotgun proteomics combined with sample pattern recognition for rapid diagnostics - a case study in Medicago truncatula nodules. *J. Sep. Sci.* **2006**, *29*, 2793–2801.
- Soding, J.; Biegert, A.; Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33* (Web Server issue), W244–248.
- Sanchez, R.; Sali, A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins: Struct., Funct., Genetics* **1997**, *50*–58.
- Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.
- Tsatsaronis GC, B. D. Amino acid and mineral salt content of tomato seed and skin waste. *J. Sci. Food Agric.* **1975**, *26* (4), 421–423.
- Sogi, D. S.; Arora, M. S.; Garg, S. K.; Bawa, A. S. Fractionation and electrophoresis of tomato waste seed proteins. *Food Chem.* **2002**, *76*, 449–454.
- Food and Agriculture Organization of the United Nations, Report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology. *Food and Agriculture Organization of the United Nations*, **2001**.
- Yoshioka, H.; Ohmoto, T.; Urisu, A.; Mine, Y.; Adachi, T. Expression and epitope analysis of the major allergenic protein Fag e 1 from buckwheat. *J. Plant Physiol.* **2004**, *161* (7), 761–767.
- Nielsen, N. C.; Bassuner, R.; Beaman, T. The biochemistry and cell biology of embryo storage proteins. In *Cellular and Molecular Biology of Plant Seed Development*; Larkins, B. A., Vasil, I. K., Eds.; Kluwer Academic Publisher: Dordrecht, 1997; Chapter 4, pp 151–220.
- Mills, E. N.; Jenkins, J. A.; Alcocer, M. J.; Shewry, P. R. Structural, biological, and evolutionary relationships of plant food allergens sensitizing via the gastrointestinal tract. *Crit. Rev. Food Sci. Nutr.* **2004**, *44* (5), 379–407.
- Breiteneder, H.; Radauer, C. A classification of plant food allergens. *J. Allergy Clin. Immunol.* **2004**, *113* (5), 821–830; quiz 831.
- Breitling, R.; Herzyk, P. Biological master games: using biologists' reasoning to guide algorithm development for integrated functional genomics. *OMICS* **2005**, *9* (3), 225–232.
- Teuber, S. S.; Sathe, S. K.; Peterson, W. R.; Roux, K. H. Characterization of the soluble allergenic proteins of cashew nut (*Anacardium occidentale* L.). *J. Agric. Food Chem.* **2002**, *50* (22), 6543–6549.
- Wang, F.; Robotham, J. M.; Teuber, S. S.; Sathe, S. K.; Roux, K. H. Ana o 2, a major cashew (*Anacardium occidentale* L.) nut allergen of the legumin family. *Int. Arch. Allergy Immunol.* **2003**, *132* (1), 27–39.
- 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* (1), 160–166.

- (37) Alcala, J. V. J.; White, R.; van der Hoeven, R. S.; Holt, I. E.; Liang, F.; Hansen, T. S.; Craven, M. B.; Bowman, C. L.; Ronning, C. M.; Niernan, W.; Fraser, C. M.; Martin, G. B.; Giovannoni, J. J.; Tanksley, S. D. Generation of ESTs from tomato fruit tissue, immature green. Unpublished, 2000.
- (38) Yamamoto, N.; Tsugane, T.; Watanabe, M.; Yano, K.; Maeda, F.; Kuwata, C.; Torki, M.; Ban, Y.; Nishimura, S.; Shibata, D. Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars. *Gene* **2005**, 356, 127–134.
- (39) WHO, Evaluation of the allergenicity of genetically modified foods. *Report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology, 22–25 January, 2001*, **2001**.
- (40) Wang, F.; Robotham, J. M.; Teuber, S. S.; Sathe, S. K.; Roux, K. H. Ana o 2, a major cashew (*Anacardium occidentale* L.) nut allergen of the legumin family. *Int. Arch. Allergy Immunol.* **2003**, 132 (1), 27–39.
- (41) 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 (2), 334–339.
- (42) Shin, D.; Sampson, H. A.; Huang, S. K.; Compadre, C.; Burks, A. W.; Bannon, G. A. Characterization of a major peanut allergen: Mutational analysis of the Ara h 1 IgE binding epitopes. *J. Allergy Clin. Immunol.* **1997**, 99 (1), 570–570.
- (43) Shin, D. S.; Compadre, C. M.; Maleki, S. J.; Kopper, R. A.; Sampson, H.; Huang, S. K.; Burks, A. W.; Bannon, G. A. Biochemical and structural analysis of the IgE binding sites on ara h1, an abundant and highly allergenic peanut protein. *J. Biol. Chem.* **1998**, 273 (22), 13753–13759.
- (44) Pacios, L. F.; Tordesillas, L.; Cuesta-Herranz, J.; Compes, E.; Sanchez-Monge, R.; Palacin, A.; Salcedo, G.; Diaz-Perales, A. Mimotope mapping as a complementary strategy to define allergen IgE-epitopes: Peach Pru p 3 allergen as a model. *Mol. Immunol.* **2008**, 45 (8), 2269–2276.
- (45) Ortolani, C.; Ispano, M.; Ansaloni, R.; Rotondo, F.; Incorvaia, C.; Pastorello, E. A. Diagnostic problems due to cross-reactions in food allergy. *Allergy* **1998**, 53 (46 Suppl), 58–61.
- (46) Asero, R.; Mistrello, G.; Roncarolo, D.; Amato, S.; Zanoni, D.; Barocci, F.; Caldironi, G. Detection of clinical markers of sensitization to profilin in patients allergic to plant-derived foods. *J. Allergy Clin. Immunol.* **2003**, 112 (2), 427–432.

PR800186D

SUPPLEMENTAL

Supplemental 1: Identified proteins via nanoflow liquid chromatography coupled to mass spectrometry from CF dimension fractions

accession number	proteine name	peptide sequence	charge	Xcorr	peptide no.
65kD band					
TC164469	weakly similar to UP Q9ZWI3 (Q9ZWI3) PV100	IVGFGVDAENKKK	2	2.7573	2
		QDQSYFVAGPEHR	2	2.7343	
		GNLNVGDVLVIPAGHPITFVATGNSNLR	2	3.3943	
BE458586	Vicilin GC72-A precursor (Alpha-globulin A)	SNNPYLFESQR	2	2.7648	3
		LFAGSNVYLLNK	2	2.928	
		SVNAPGNLQEYFSAGGQNPESFYR	2	3.8037	
TC161869	Elongation factor 1-alpha (EF-1-alpha)	LPLQDVYK	2	2.4024	3
		IGGIGTVPVGR	2	3.5341	
		VETGVIKPGMVVTFGPTGLTTEVK	3	4.9154	
TC162433	heat shock protein (Type I)	SSSMFDFPSIDVDFPFR	2	4.8091	2
		ASMENGVLTVTPK	2	2.2462	
TC164140	expressed protein	DYFICTGPASMLVPLVINPGEWR	2	2.9859	2
		DYFICTGPASMLVPLVINPGEWR	3	3.9934	
17kD band					
TC156005	Legumin-like protein	HVNSQNFPILR	2	3.119	7
		AGQNGFEFAVFR	2	3.2199	
		AGQNGFEFAVFR	2	3.4129	
		VNQGEMFVVPQYYASTVK	2	2.8054	
		GDAQVQVVDHTGQQVMNDR	2	3.4333	
		AMPIEVLTNAYQISPNEAQR	2	3.731	
		AMPIEVLTNAYQISPNEAQR	3	4.3222	
		ISTVNSNSLPVLNWLQLSAER	3	4.785	
TC169589	11S globulin, partial	ISTVNSNSLPVLNWLQLSAER	2	3.8364	8
		TNDNAMISPLAGR	2	4.2642	
		SVFDDQVR	2	2.1251	
		KAGEQGLDYIAFK	2	3.5704	
TC168461	11S globulin	FNRDELSVFGPGAR	2	3.6754	8
		FNRDELSVFGPGAR	3	3.4429	
		GTGQIQVVGDTGNSVFDEVR	2	5.3313	
		TNDEAITSALAGR	2	4.3359	
		QMIVVPQNFAIMK	2	2.0694	
		AMPEEVLMSYQISR	1	1.9815	
		KAGDQGLEYIAFK	2	4.1223	
		KAGDQGLEYIAFK	3	3.5156	
		AGDQGLEYIAFK.	1	2.4887	
		AGDQGLEYIAFK.T	2	2.1674	
TC153878	Pathogenesis-related protein STH-2	KINFVEGGPIK	2	1.9378	3
		ALVIDGDNLIPK	2	3.3208	
		INFVEGGPIKYLK	2	1.9483	
30kD band					
TC156005	Legumin-like protein	SNQLDQNFR	2	2.0199	2
		ADEEEEEFEGR	2	2.3523	
TC158893	Seed storage protein Lec2SA1 large chain	VVQQUEEQTGKVQGR	2	1.93	2
		VVQQUEEQTGKVQGR	3	2.6185	
TC164188	ypersensitive response assisting protein	FKTPIQYFNR	3	1.7004	2
		VQNIIVPR	2	2.9069	
AW930142	Late embryogenesis abundant protein D-7 (LEA D-7)	ITDTASGTGQYVAGNLK	2	2.6132	2
		DKITDTASGTGQYVAGNLK	3	3.449	

10

2

Supplemental 3A: Protein sequence alignment between legumin (TC156005) and the brazil nut legumin

Ber e 2 (*Bertholletia excelsa*, gi 30313867) allergen.

```

TC156005  SLAMAVSTRVLLALLFSAPLLSAANAYRQYPPFQQCQHGQRSTPLTQAQQCLQLRLTCSQP
Ber e 2      MAKLFLLSLGIFLLFHCCLAIETFEQEELYECRIQLRLTAQEP
TC156005  SQRIESRGCYTEVMDENEEQFCQCVGVAFMRNVIRQNSLSLPNFHMPRLVYIEQGRGLIS
Ber e 2  QYRLAEAGVSEVMDYTDQFRCAGVAALRNTIRPQCLLPVYTNAFKLYVVTQGRGILG
TC156005  VIMPCCARTFQSQT-FQAGFMACHEERGERRCNDQHQKVHRIHQCDVVAIPAGAAHWCYND
Ber e 2  VLMPCGPETFQMSQFQGSREQEEREG--RFQDQHQKVHHLKKGDIATPAGVALWCYND
TC156005  GEHELVAVSINDLNHRSNQLDQNFRAFYLAGGVQETSGRQIG--VGKQSTQKF--QNIFFR
Ber e 2  GDEDLVTVLVQHTASDLNQLDQNPFRHFLAGNIQR-SQKQRGERYCLGGQQLADNVFK
TC156005  AFNTELMAEA--FNIPVEIVRMQREERGLIVNVRECMIMIRADEEEEREFEGRSQRGQ
Ber e 2  GFNMEALADVLGFCMDTETARKVVEGDDQRGHIVRVEQCLKVIIRPPRIEEL--QDEGC
TC156005  EWMVAEVTGNCLEENICTMKIRTNMDTQRRADIYSRQAGKIKHVNSQNFPILRDMDSASK
Ber e 2  GY-----NGLKETICSATFIQNIDNPAAEDFYNPRACRLTTVNSLKVPIILTLQLSAMK
TC156005  GTLYPNALMSFHWNVNCHSVIYVQRGDAQVQVVDHTCQQVMNDRVNQGEFVVPQYYAST
Ber e 2  GVLYENAMMAPLWRLNANSVVYAVRGEARVQIVDHRGETVFDNLRQGMVVPQNFVVV
TC156005  VKAGQNGCFEFAVFTSSQPMNSQLAGYTSVIRAMPIEVLTNAYQISPNEAQRLKMNREGGE
Ber e 2  KQAGSRGFEWVVFNTNDNALFSTAAGRTSPLRGIFVGVLANAYRLSQEARRKLNMD-E
TC156005  SFLLSPQRRESFNGFNVSHLGYRLVVQVSSLCFAKMKQERLVFRSKCNSSSSCLFVPRSHP
Ber e 2  AVLFQPCSRSEGRASA
TC156005  RCSFSLCKISESTGRIFYFHINQIINSQC

```

Supplemental 3B: Protein sequence alignment between the putative legumin β -unit (TC156005) and the Fag e 1 allergen.

20

```

Fag e 1      ****+ .*.+.* *.. ****+. .****. *.+.*.***. .+***. .** ****+
TC156005  GLEQAFCNLKFQNVNRPSSRADVFNPFRAGINTVDSNNLPILFEIQLSAQHVVLKYNAIL
GLEENICTMKIRTNMDTQRRADIYSRQAGKIKHVNSQNFPILRDMDSASKGTLYPNALM
Fag e 1      *.+.*.***+*** **+.***** *.+.*.***+.*.***** +* .****+.*
TC156005  GPRWNLNHSALYVTRGEGRVQVVGDEGRSVFDDNVQRGQILVVPQGFVAVLKAGREGLE
SPHWNVNCHSVIYVQRGDAQVQVVDHTGQQVMNDRVNQGEFVVPQYYASTVKAGQNGFE
Fag e 1      .. .+.... .** +** *****+*****.+***.*** ** *** .* * .. *
TC156005  WVELKNDDNAITSPIAGKTSVLRAIPVEVLANSYDISTKEAFRLKNGR-QEVEVFRPFQS
FAVFTSSQPMNSQLAGYTSVIRAMPIEVLTNAYQISPNEAQRLKMNREGESFLLSLP-QR
Fag e 1      *. .+.. +**
TC156005  RD----EKERERFSIV
RSfngfNVSHLGYRLV

```

Supplemental 4: 47% homology between the tomato vicilin (AM932874, contig TC164469) and the walnut vicilin Jug n 2 (*Juglans nigra*, gi 31321944), red: identical amino acids.

```

contigTC164469 CRQRCQYRKEQQQHQHGGCTGDDLCNRGPDKSYKRLQECQRRCCQ-SRQQCQRLQECQQ
                .....
Jug n 2         -----GRDRQDPQQQYHR---CQRRCCQIQEQSPERQRCQQ

contigTC164469 RCQQEYQREKQHQHC-ETNPQWE----QQEKSNNPYLFESQRFESFFRASHGDFRILER
                .....
Jug n 2         RCERQYKQCCREKGPASPPRESHGKREERQQHNPYYFHSQSIRSPHESECEVKYLER

contigTC164469 FNQPSQLLKCIKRYVAILELEPQSFLVPHHCDCGATYVVVKGQGVINIAEQDNKNSFNL
                .....
Jug n 2         FAERTELLRCIENYRVVILDANPTFMLPHHKDAESVIVVTRGRATLTLVSQETRESFNL

contigTC164469 QKGDVIRLFAGSNVYLLNKKDNEKLFVYVLAKSVNAPCNLQHYFSAGGQNPES-SFYTRAFS
                .....
Jug n 2         ECGDVIRVPACATEYVINQDSNERLEMVKLLQPVNPGQVRHYTAACAKSPDQSYLRVFS

contigTC164469 SDILESFAFNNDKLEELFGQH--KEGIIKASEEQIRAISEHASSTQQTRGR--TQGP
                .....
Jug n 2         NDIIVAALNTPEDRLERFFDQEQREGVITRASQEKRLALSQHAMSAGQRPWGRSSCGP

contigTC164469 FNLMEKRPVFESRFGQFFEAPRYEQQLDLDAAVGFMNINQGMVLPYYNTRKTLVMV
                .....
Jug n 2         ISLKSERPSTYSNFGQFFEACPEEHRQLQEMDVLVNYAEIKRGANDMPPHYNSKATVVVVYV

contigTC164469 IECMARFEMACPHLGRQSQSPWRCQGREREQEQQEQEGDVHYQKIRGNLNVGDVLVI
                .....
Jug n 2         VEGTGRYEMACPHVSSQS----FEDQGR---REQEEREESTG--RFQKVTARLARGDIFVI

contigTC164469 PAGHPITFVATGNSNLEIVGFCVDAENNHKGNFLAGKQNIWRNIDREAKELSFNMPGREVE
                .....
Jug n 2         PAGHPITAITASQENLELLGFCINGENNQNFLAGQNSIINQLEREAKELSFNMPDEEIE

contigTC164469 EIFQRQDSYFVACPEHRQQRERGEGRGQDOYLSILDVVF-
                .....
Jug n 2         EIFESQMSYFV--PTERQSR-RGQ---GRDHPLASILGFAFF

```