JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

The *Citrus clementina* Putative Allergens: From Proteomic Analysis to Structural Features

Ilia Anna Serra,[†] Letizia Bernardo,[†] Antonia Spadafora,[†] Primetta Faccioli,[‡] Cristina Canton,[§] and Silvia Mazzuca*[,†]

[†]Dipartimento di Chimica e Tecnologie Chimiche, Università della Calabria, Rende (CS), Italy

[‡]Consiglio per la Ricerca e la Sperimentazione in Agricoltura—CRA (Centro di Ricerca per la Genomica e la postgenomica Animale e Vegetale, Fiorenzuola D'Arda (PC), Italy

[§]Ion Sources Biotechnologies, Polo Tecnologico Gerenzano (VA), Italy

Supporting Information

ABSTRACT: Several allergens have been identified and characterized in the genus *Citrus*, which belongs to the germin-like proteins (GPLs), profilins, and non-specific lipid transfer proteins (nsLTPs). In this work, in silico sequence analysis, protein purification, mass spectrometry identification, and the spectral counting method were integrated to identify new putative allergens of *Citrus clementina* and their expression level in the fruit peel. The in silico analysis revealed fifteen new sequences belonging to GLPs (Cit cl 1), and two more belonging to nsLTPs (Cit cl 3). No other new sequences were found as regards profilins (Cit cl 2). Each putative allergen from fruit peel was obtained using different protein extraction methods, and the protein sequences of the putative allergens were identified by means of LTQ-Orbitrap XL mass spectrometer. The spectral counting strategy revealed that Cit cl 1 had a higher expression level than Cit cl 2 and Cit cl 3. To predict the quaternary structure and deduced function of Cit cl 1, its primary sequence was used as a template to search a homologous protein structure in the RCSB PDB Database, getting high correspondence with the oxalate oxidase protein in barley.

KEYWORDS: *Citrus clementina, allergens, proteomic analysis, protein structure*

INTRODUCTION

Citrus fruits (genus *Citrus* in family Rutaceae) are commonly included in the population's diet in many countries, either as fresh fruits or derived beverages and foods. Citrus oils (extracted from the peel) are used as fragrance enhancers for food, perfumes, soap, household detergents, and dried citrus peels are included in tea blends and marmalades. The peel is also the major source of various bioactive compounds (phenolics, carotenoids, flavonoids, ascorbic acid), which can be used as natural antioxidants.[1,2] The consumption of citrus fruits and/or the use of byproducts containing citrus extract, however, may induce allergic reactions, with heterogeneous clinical manifestation of allergies, varying from a mild oral allergy syndrome (OAS) to severe anaphylaxis cases.[3−9] Allergens belonging to the germin-like proteins,[10−13] profilins[11,14] and non-specific lipid transfer proteins[4,6] were identified and characterized in several citrus species.

Germin-like proteins (GLPs) are part of the cupin superfamily, having different enzyme functions that include the two hydrogen peroxide generating enzymes oxalate oxidase and superoxide dismutase.[15−17] It is well-known that many plant-derived members of the cupin superfamily have allergenic properties, and globulins are the most important subgroup in this context. However, germin and other GLPs are now being increasingly recognized as significant allergens.[15]

Profilins are actin monomer binding proteins of low molecular weight, which regulate the organization of actin cytoskeleton in eukaryotes, including higher plants.[18,19] Being involved in essential cellular processes, profilins can be found in all organisms examined so far, and are therefore considered panallergens, which are responsible for many cross-reactions between pollens and food.[20,21] Finally, non-specific lipid transfer proteins (nsLTPs) are closely related to basic proteins, specific of flowering plants.[18] These proteins are involved in plant defense mechanisms against bacteria and fungi and, possibly, in the assembly of hydrophobic protective layers of surface polymers, such as cutin.[22,23] The extensive clinical evidence collected in this regard has allowed defining nsLTP an important panallergen family of both plant food and pollens.[18,21]

In the genus *Citrus* the orange GLP Cit s 1 and the profilin Cit s 2 have been well characterized and are reported to act as allergens.[11,12,14] Among the nsLTP, Cit s 3, Cit l 3, and Cit r 3 have been recognized as allergens in sweet orange, lemon, and tangerine, respectively.[4,6] The available information on the allergens of *Citrus clementina*, which is extensively cultivated in Mediterranean countries and North America, is very scarce so far. Recent molecular analysis demonstrated the expression of homologues to Cit s 1, Cit s 2, and Cit s 3 in the *clementine* pollen. The clinical data showed that *clementine* allergy is not a rare condition in pollen sensitized patients.[5] Furthermore, the current availability of the complete genome sequences of orange and clementine has allowed the in silico identification of

Received: May 29, 2013
Revised: August 5, 2013
Accepted: August 8, 2013
Published: August 8, 2013

several putative allergens in clementine, belonging to profilins and nsLTPs.[24] The main aims of this work consisted in (i) identifying new putative allergens of *C. clementina*, through in silico analysis; (ii) testing different procedures to purify each class of putative allergens from fruit peel; (iii) obtaining the primary sequences of each putative allergen by means of mass spectrometry analysis; (iv) assessing the expression of each putative allergen, by applying the spectral counting strategy;[25] and (v) predicting the structure and function of the more expressed putative allergen, searching in the RCSB Protein Data Bank.

MATERIALS AND METHODS

In Silico Sequence Analysis. The citrus genome database (http://www.citrusgenomedb.org/species/clementina) includes whole genome sequencing of *C. clementina*. Information about citrus allergens is completely lacking in this database. The official Web site of the WHO/IUIS (International Union of Immunological Society) Subcommittee on allergen nomenclature (http://www.allergen.org/) contains approved and officially recognized allergens of orange, lemon, and tangerine. Moreover, the Allergome platform for allergen knowledge (http://www.allergome.org/index.php) collected a list of allergenic sources and molecules of several citrus species, including *C. clementina*, making use of the widest selection criteria and sources. Finally, some peptides of *Citrus* allergens were described only in the literature.[12]

In this work, the amino acid sequences collected from the last two databases, combined with sequences drawn from literature, were used to identify unknown putative allergens of clementine by means of the Citrus Genome Database Basic Local Alignment Search Tool (BLAST) against *C. clementina* data sets. To detect putative signal peptide cleavage site in the sequence of putative allergens identified, we used SignalP 4.0 server program[26] available at http://www.cbs.dtu.dk/services/SignalP/.

Protein Extraction and Separation. The high quality fruits for the tests (PGI CE No. 2325/97), grown in organic cultivations, were selected in the Italian countryside (namely, in Calabria, Southern Italy). Organic growing does not make use of chemical pesticides and insecticides, which could alter protein extraction from the peel. Clementine ripe fruits, with the full orange pigmented peel, were harvested from three trees in December 2012 (180 Days After Flowering). Five fruits from each tree were collected and the peel samples of different trees and fruits were grouped. Peel portions of about 1 g each were frozen in liquid N_2 and stored at −80 °C. These samples were used as independent biological replicates for protein extraction.

We tested a different protein extraction procedure on fruit peel (both flavedo and albedo) to obtain highly purified protein extracts. The procedures include extensive organic solvent washes of powder tissue to remove water-soluble contaminants, following Pignataro et al.[13] with some modifications, and phenol extraction of proteins in presence of sodium dodecyl sulfate, according to Wang et al.[27] In short, the peel tissue was ground in liquid N_2 to obtain a fine powder that was separately processed, according to the following four procedures: (a) Procedure 1. The tissue powder was suspended in100% cold acetone; the tissue suspension was vortexed for 30 s and centrifuged at 13000 rpm for 4 min at 4 °C; the obtained pellet was extensively washed with 10% TCA in cold acetone until the supernatant became colorless, and twice with cold aqueous 10% TCA; the pellet was rinsed twice with cold 80% acetone, using both vortex and centrifugation in every step. The final tissue powder was dried at room temperature and used for protein extraction, or stored at −80 °C for further use. (b) Procedure 2. Cold 10% TCA in acetone was added to the tissue powder; the tissue suspension was vortexed for 30 s and centrifuged at 13000 rpm for 4 min at 4 °C; the pellet was extensively washed in cold aqueous 10% TCA, using both vortex and centrifugation in each step; the TCA was removed by washing it with cold 80% acetone until the supernatant was colorless and was dried at room temperature. (c) Procedure 3. Cold aqueous 20% TCA was added to the tissue powder; the tissue suspension was vortexed for 30 s and centrifuged at 13000 rpm for 4 min at 4 °C; the pellet was rinsed with cold 80% acetone, vortexed, and centrifuged until the supernatant was colorless. The final pellet was dried at room temperature. (d) Procedure 4. Cold aqueous 20% TCA was added to the tissue powder; the tissue suspension was vortexed for 30 s and centrifuged at 13000 rpm for 4 min at 4 °C; the pellet was washed twice with 20% TCA in acetone, using both vortex and centrifugation in every step; the pellet was twice rinsed with cold 80% acetone, vortexed, and centrifuged until the supernatant was colorless. The final pellet was dried at room temperature. Further, the proteins obtained from all procedures were purified by phenol extraction methods,[13,27] dissolved in a Laemmli sample buffer, and quantified by the Bradford assay.

Protein samples obtained from all procedures were processed on 1D SDS−PAGE. The Laemmli buffer system was used to cast a 6% stacking gel and 12.5% resolving gel. After denaturation at 100 °C for 3 min, proteins (20 μg for each sample) were resolved at constant 200 V in a Bio-Rad mini Protean II apparatus until bromophenol blue reached the bottom of the gel. Electrophoresis of polypeptide patterns from each procedure are shown in Figure 1. Gel slides were excised in

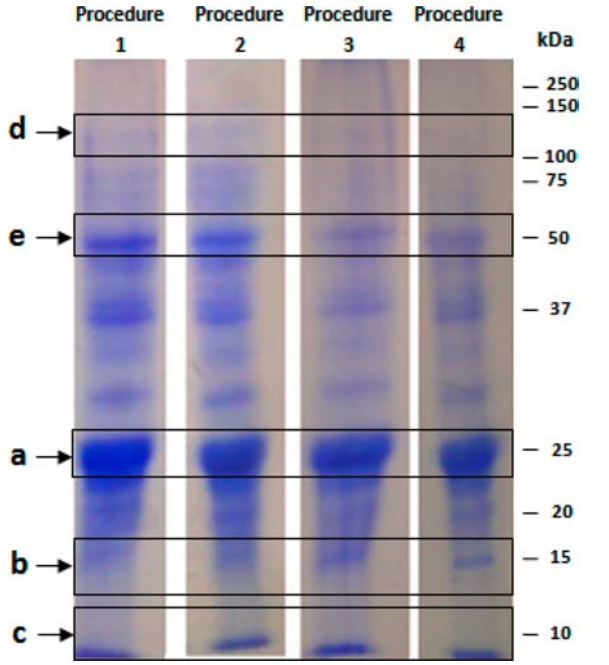

Figure 1. CBB-stained 1DE gels of proteins extracted from *Citrus clementina* peel by means of four different procedures (see the text for details). Arrows indicate gel slices that have been digested for the mass spectrometry analysis. Each lane was loaded with 20 μg of purified protein extract. Gel slices: a (27−23 kDa); b (18−12 kDa); c (12−8 kDa); d (130−110 kDa); e (55−40 kDa).

the ranges of molecular weights corresponding to those of the putative allergens (see Table 1) as follows: (a) range 27−23 kDa for the putative Cit cl 1 GLPs; (b) range 18−12 kDa for putative Cit cl 2 (profilins); (c) range 12−8 kDa for putative Cit cl 3 (nsLTPs) (see boxes in Figure 1); additional slices were excised at the range 130−110 kDa (d) and at the range 55−40 kDa (e) corresponding respectively to the homoexamers and to the dimers of the germin-like proteins detected in the lemon peel.[13] After protein reduction and alkylation, slices were digested overnight with trypsin at 37 °C.[28] The tryptic fragments were immediately processed for mass spectrometry analysis though a LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) for protein identification.

Mass Spectrometry Analysis. Tryptic peptides were analyzed by liquid chromatography−tandem mass spectrometry (LC−MS/MS) using a high resolution LTQ-Orbitrap XL mass spectrometer.

Table 1. Putative Allergens of *Citrus clementina*, Obtained through in Silico Analysis

allergen	transcript name[a]	length (aa)	pI/MW[b] (kDa)
GLPs (putative Cit cl 1)	**clementine0.9_021092m**	**224**	**5.8/24.2**
	clementine0.9_021067m	224	5.9/24.1
	clementine0.9_021078m	224	5.7/24.2
	clementine0.9_031237m	224	5.7/24.1
	clementine0.9_021090m	224	6.3/24.0
	clementine0.9_021094m	224	5.4/24.1
	clementine0.9_021076m	224	5.5/24.1
	clementine0.9_021233m	221	7.8/23.7
	clementine0.9_021308m	220	7.8/23.6
	clementine0.9_032110m	220	7.7/23.5
	clementine0.9_032828m	220	7.8/23.5
	clementine0.9_035688m	220	7.8/23.5
	clementine0.9_032180m	220	7.6/23.6
	clementine0.9_032269m	220	7.8/23.5
	clementine0.9_033886m	225	8.6/24.4
profilin (putative Cit cl 2)	**clementine0.9_025303m**	**131**	**5.0/14.1**
	clementine0.9_025205m	133	5.4/14.2
	clementine0.9_025288m	131	4.6/14.1
nsLTP (putative Cit cl 3)	clementine0.9_023997m	164	8.7/17.5
	clementine0.9_025859m	115	9.3/11.7
	clementine0.9_025486m	**126**	**9.7/13.2**
	clementine0.9_025843m	115	9.8/12.0
	clementine0.9_025808m	117	8.9/11.9
	clementine0.9_025527m	125	9.0/13.2
	clementine0.9_032770m	117	8.8/11.9

[a]The underlined transcripts are those identified in this work through in silico analysis. The transcripts in boltd are those confirmed by means of mass spectrometry analysis (see also Table 2 and Figure 2). [b]pI/M_w: Theoretical molecular mass and isoelectric point of the protein sequence calculated by Expasy Compute pI/Mw tool.

Chromatography separations were conducted on a Waters XBridge C18 column (300 μm ID × 100 mm length and 3.5 μm particle size); using a linear gradient from 5 to 90% ACN, containing 0.1% formic acid with a flow of 4 μL/min, including the regeneration step, one run lasted 70 min. Acquisitions were performed in the data-dependent MS/MS scanning mode (full MS scan range of 250−1800 m/z followed by full MS/MS scan for the most intense ion from the MS scan). Peptide sequences generated by mass spectrometry were searched using GPM software (Global Proteome Machine) against plant databases, including some sequences of *C. sinensis* (http://plant.thegpm.org/tandem/thegpm_tandem.html/). Peptide sequences that were not identified with the above method were further searched on the GPM Web site using X!Tandem algorithm against a local database built with all putative allergen sequences of *C. clementina* identified by the in silico analysis (see the paragraphs referring to In Silico Sequence Analysis). Quantization was carried out by dividing the number of tandem mass spectra of all peptides assigned to each allergen by the total number of spectra identified and assigned to proteins in each sample.[25]

Protein Structure. The amino acid sequence of GLP identified in clementine by mass spectrometry (Cit cl 1) was used to refine a model for a monomer structure using the 1FI2 PDB structure deposited in the RCSB Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) as a template by homology modeling. At first, the monomer model was used to create a possible dimer structure of the putative allergen by ZDOCK server 3.0.2 (http://zdock.umassmed.edu/) freely available.[29,30] The dimeric structures given by ZDOCK software, using a fast Fourier transform algorithm and a large supercomputing cluster, were finally used in the ClusPro software for protein−protein docking.[31,32] The model with highest confidence score was selected to predict the 3D structure assembly. The Cit cl 1 protein sequence of homoexameric structure was calculated by the pI/Mw tool in Expasy Web site http://web.expasy.org/compute_pi/.

RESULTS AND DISCUSSION

In Silico Sequence Analysis. To date, the official allergen list of IUIS Allergen Nomenclature Subcommittee comprises only a single germin-like protein (GLPs), a single profilin, and four lipid transfer proteins (nsLTPs) from different species of *Citrus*, excluding clementine. A larger number of allergens were listed in the Allergome database, which includes one profilin and five nsLTP sequences of *C. clementina*. The latter were derived from an in silico analysis of the genome information[24] but not validated by mass spectrometry. In this work the in silico analysis has been improved, identifying ten new putative allergens belonging to GLPs and two new nsLTPs, whereas three putative profilin allergens were confirmed (Table 1). According to the allergen nomenclature adopted in the Allergome platform, we named the *C. clementina* GLPs as Cit cl 1, the profilin as Cit cl 2, and the nsLTP as Cit cl 3. GLPs are encoded by a family of genes found in all plants. They are part of the cupin superfamily, which comprises proteins with conserved tertiary structure, but has limited similarity in the primary sequences and in the different enzymatic and nonenzymatic activities.[15] Newly identified putative genes of *Cit cl 1* have an open reading frame (ORF) length ranging from 663 to 678 bp, encoding a polypeptide of 220−225 amino acid residues. Full-length Cit cl 1 amino acid sequences have a calculated mass ranging from 23.5 to 24.1 kDa and a theoretical p*I* from 5.4 to 8.6 (Table 1), consistent with usual citrus GLPs features.[10,11,13] Profilins are a family of small (12 to 15 kDa) and highly conserved molecules with sequence identity exceeding 75%, even between members of distantly related organisms.[21] The Cit cl 2 proteins identified have a calculated molecular mass about of 14 kDa, corresponding to a polypeptide of 131−133 amino acid residues, and a theoretical p*I* from 4.6 to 5.4 as expected from profilin allergens (Table 1).

Non-specific LTPs are small basic (p*I* 8−10) proteins and have been classified into two families according to the observed molecular masses: LTP1 (approximately 10 kDa) and LTP2 (approximately 7 kDa).[23,33] Although individual members of this protein family show little sequence similarity, they share nonetheless common structural features based on a conserved scaffold of eight cysteine residues.[34] Moreover, both families present a signal peptide at the amino terminal region. This peptide is then excised, targeting the LTPs to cell secretory pathway, where they are exported to the apoplast.[23] The nsLTPs of *C. clementina* (putative Cit cl 3 proteins) have a calculated molecular mass ranging from 11.7 to 17.5 and theoretical p*I* from 8.7 to 9.8 (Table 1). These calculated molecular masses were larger than those reported for this kind of protein, suggesting that they correspond to the pro-protein, still containing the signal peptide. Using SignalP 4.0 Server, we predicted the signal peptide cleavage site of the seven identified nsLTPs, thus obtaining putative signal peptide ranging from 22 to 31 amino acids. In this way, the calculated mass of mature proteins resulted as ranging from 9.3 to 10.8 kDa. As an exception, the mature protein translated from clementine 0.9_023997m nucleotide sequence showed a larger molecular size (15.1 kDa) than the typical nsLTP1 allergens. On the other hand, no significant similarity was found through BLASTp searches at the NCBI BLAST Web site for the last 49 amino acid residues at C-termini of this sequence, suggesting that this peptide could contain sequence errors. Regarding the protein

dx.doi.org/10.1021/jf4023367 | J. Agric. Food Chem. 2013, 61, 8949−8958

sequence homology, the presence of two consensus pentapeptides (T/SXXDR/K and PYXIS) in six putative Cit cl 3 identified sequences strongly suggested that these proteins belong to the nsLTP1 family.

In conclusion, we identified several new putative allergens in *C. clementina* fruits. First, this result allowed us to restrict the 1DE gel-based proteomics to the range of molecular weights assigned to putative allergens. Moreover, the identified sequences were used to create a local database, thereby optimizing the protein identification by mass spectrometry analysis and bioinformatics (see Mass Spectrometry and Bioinformatics).

Protein Extraction and Separation. Protein extracts obtained from peel and analyzed by 1DE PAGE apparently showed a good contaminant removal and protein yield, independently of the procedures used for extraction. However, the results from mass spectrometry analyses showed that peptides belonging to the Cit cl 1 protein were found in protein samples obtained from all extraction procedures, whereas those belonging to the Cit cl 2 protein were identified exclusively in the sample of the extraction procedure 3. Finally, the higher number of peptides assigned to Cit cl 3 protein were obtained using the extraction procedure 4 (Table 2).

The detection of peptides belonging to Cit cl 1 (clementine0.9_021092m) in all procedures used, and the very high percentage of spectra found for this protein (Tables 2 and 3, see also Mass Spectrometry and Bioinformatics), validate the hypothesis that Cit cl 1 has high expression level in the peel tissue of clementine fruits, in compliance with high amount of GLPs found in the proteomic analysis of *C. limon*[13] and in gene expression study of *C. sinensis*.[35]

Two peptides, with quite a low percentage of spectra, belonging to the putative Cit cl 2 clementine0.9_025303m were obtained following only the procedure 3 (Tables 2 and 3). This finding strongly suggests that the Cit cl 2 profilin has low expression in peel tissue, according to a previous study on sweet orange allergens that showed the accumulation of profilin Cit s 2 mainly in the fruit pulp.[14] Finally, the largest number of peptides for the nsLTP (putative Cit cl 3 clementine0.9_025486m) was obtained following procedure 4 (Tables 2 and 3).

Hence, procedures 3 and 4 seem more appropriate for the extraction of proteins of low molecular weight (profilins and nsLTPs), less abundant than GLPs in the peel tissue of *clementine* fruits. Both extraction methods involve, as a first step, the addition of cold aqueous 20% TCA to the tissue powder, instead of 100% acetone (procedure 1) or 10% TCA in acetone (procedure 2). TCA, by interacting with peptides, induces an increase in the hydrophobicity of proteins, which may lead to aggregation through hydrophobic interactions, and the efficiency of aggregation can depend on TCA concentration.[36] Moreover, in an aqueous solution, the TCA is almost completely dissociated and modifies the ionic strength of the solvent.

Mass Spectrometry and Bioinformatics. The expression of putative allergens identified through in silico analysis was verified by mass spectrometry using Orbitrap-LC-MS/MS. Spectra were used to identify peptide sequences, by searching both in public online plant databases and in a local database containing all previously identified putative allergen sequences. This procedure was used to optimize the protein identification, because available searching engines analyzing mass spectra are not capable of interrogating the citrus genome database. Spectral counting analyses for each putative allergen in every extraction procedure are shown in Table 3. Mass spectrometry statistical parameters for each protein identified are reported in Tables S1−5, as Supporting Information.

Six peptide fragments detected from mass spectrometry gave high score matching with the clementine0.9_021092m protein sequence (Table 2). It is the only Cit cl 1 isoform confirmed among those found through in silico analysis (Table 1). Four of these peptides were also found in *Citrus limon* by Pignataro et al.[13] and annotated on Allergome as partial sequence of Cit l 1. Moreover, alignment between putative full-length Cit cl 1 protein and the translated open-reading frame sequence for homologous orange Cit s 1 showed that these two sequences have very high amino acid identities (99%, Figure 2a). N-terminal sequence of Cit cl 1 (22 residues) is probably a cleavable signal sequence, such as assumed for identical peptide of Cit s 1.[12] As a matter of fact, by using SignalP 4.0 Server, we succeeded in predicting a signal peptide cleavage site positioned between Ala22 and Thr23 of both Cit s 1 and putative Cit cl 1 sequences (Figure 2a). The sequence of mature protein was then validated by the mass spectrometry analysis with a coverage of 67.3% (Table 2).

A high number of Cit cl 1 spectra waere found in all used extraction procedures (Table 3), with a higher number assigned to peptides in the molecular range of 27−23 kDa (Table 3, gel slices a), corresponding approximately to the molecular weight of the GLP monomer, as pointed out by the in silico analysis in this work and well reported in the literature.[10−12,37,38] Some peptides of Cit cl 1 were found at lower molecular weights (Table 3, b and c slices), suggesting that the degradation of this protein occurred during extraction. Two peptides were detected at 50 kDa and 120 kDa (Table 3, d and e slices), probably due to the rearrangement of the monomer in oligomer intermediate forms, in spite of denaturing electrophoresis conditions. It is well-known that the typical quaternary structure of GLP is characterized by a homoexameric assembly of six monomers, and that these proteins can assume an anomalous migration on SDS−PAGE, due to the rearrangement of monomers producing an apparently high molecular mass.[13,17,38,39]

Concerning profilin, two peptide fragments gave high score matching with the Cit cl 2 clementine0.9_025303m sequence (Table 2). The peptides were identified in the samples obtained by following procedure 3 (Table 3, slice b), with a 24% protein coverage. A generic BLAST search in NCBI of the two peptides resulted in the identification of several profilin proteins. In particular, it was pointed out that these peptides match with a large number of plant allergens, such as Hev b 8 (*Hevea brasiliensis*), Ole e 2 (*Olea europaea*), Lit c 1 (*Litchi chinensis*), Pru du 4 (*Prunus dulcis x Prunus domestica*), Amb a 8 (*Ambrosia artemisiifolia*), Pyr c 3 (*Pirus communis*), Fra a 4 (*Fragaria x ananassa*), and Mal d 4 (*Malus x domestica*), with a high level of amino acid identity, ranging from 97% to 91%. This result was expected, considering that profilins are panallergens showing great similarity in the primary sequences.[21] Moreover, alignment between identified protein Cit cl 2 and Cit s 2 allergen of sweet orange (UniProt Accession No. P84177) showed 82% amino acid identity (Figure 2b).

Concerning nsLTPs, the peptide fragments detected by mass spectrometry gave high score matching with the Cit cl 3 clementine0.9_025486m protein sequence (Table 2). Bioinformatic analysis with SignalP server program pointed out that the N-terminal of this sequence contains a eukaryotic signal sequence with a cleavage site positioned between Ala23

Table 2. Peptide Sequences of Putative Allergens Detected by Mass Spectrometry According to the Protein Extraction Procedures 1−4[a]

allergen	procedure	log(e) protein[b]	attribution/residue alignment	detected peptide sequence	m/z[c]	z[d]	log(e) peptide[b]	% protein coverage
GLP (putative Cit cl 1)	1	−512.3	clementine0.9_021092m/aa 23−39; 57−74; 75−99; 100−116; 145−168; 169−180	TDPGHLQDVCVAINDPK	939.9; 626.9	2; 3	−9.0	67.3
				AEDFFFSGLGKPGNTANR	964.4; 643.3	2; 3	−6.2	
				LGVDETDANVEQIPGLNTLGISAFR	877.1	3	−8.4	
				IDYAPYGQRPPHIHPR	639.6	3	−1.6	
				VLNKGDVFVFPQGLIHFQFNIGK	873.1	3	−3.3	
				TNAAAYSALNSQFPGEVTIADTVFGANPSINPDFLGK	1266.6	3	−7.5	
	2	−508.0	clementine0.9_021092m aa 23−39; 57−74; 75−99; 100−116; 145−168; 169−180	TDPGHLQDVCVAINDPK	939.9; 626.9	2; 3	−2.7	67.3
				AEDFFFSGLGKPGNTANR	964.4; 643.3	2; 3	−6.5	
				LGVDETDANVEQIPGLNTLGISAFR	877.1	3	−8.6	
				IDYAPYGQRPPHIHPR	639.6	3	−3.1	
				VLNKGDVFVFPQGLIHFQFNIGK	873.1	3	−3.2	
				TNAAAYSALNSQFPGEVTIADTVFGANPSINPDFLGK	1266.6	3	−8.1	
	3	−541.1	clementine0.9_021092m aa 23−39; 57−74; 75−99; 100−116; 145−168; 169−180	TDPGHLQDVCVAINDPK	939.9; 626.9	2; 3	−6.0	67.3
				AEDFFFSGLGKPGNTANR	964.4; 643.3	2; 3	−5.7	
				LGVDETDANVEQIPGLNTLGISAFR	877.1	3	−7.5	
				IDYAPYGQRPPHIHPR	639.6	3	−2.7	
				VLNKGDVFVFPQGLIHFQFNIGK	873.1	3	−6.0	
				TNAAAYSALNSQFPGEVTIADTVFGANPSINPDFLGK	1266.6	3	−8.4	
	4	−507.9	clementine0.9_021092m aa 23−39; 57−74; 75−99; 100−116; 145−168; 169−180	TDPGHLQDVCVAINDPK	939.9; 626.9	2; 3	−4.5	67.3
				AEDFFFSGLGKPGNTANR	964.4; 643.3	2; 3	−5.9	
				LGVDETDANVEQIPGLNTLGISAFR	877.1	3	−10.2	
				IDYAPYGQRPPHIHPR	639.6	3	−3.7	
				VLNKGDVFVFPQGLIHFQFNIGK	873.1	3	−7.5	
				TNAAAYSALNSQFPGEVTIADTVFGANPSINPDFLGK	1266.6	3	−9.2	
profilin (putative Cit cl 2)	3	−9.6	clementine0.9_025303m aa 53−71; 72−84	DFEEPGSLAPTGLHLGGTK	963.5; 642.6	2	−2.6	24.4
				YMVIQGEPGAVIR	716.9	2	−4.3	
nsLTP (putative Cit cl 3)	1	−10.6	clementine0.9_025486m aa 24−41;84−100	ITCGQVTASLAPCIPFLR	1002.5	2	−2.5	34.0
				GIKPNVAAGLPSQCGVR	862.4; 575.3	2	−5.2	
	3	−8.5	clementine0.9_025486m aa 24−41;84−100	ITCGQVTASLAPCIPFLR	1002.5	2	−2.0	34.0
				GIKPNVAAGLPSQCGVR	862.4; 575.3	2	−3.7	
	4	−22.7	clementine0.9_025486m aa 24−41;46−56; 84−100	ITCGQVTASLAPCIPFLR	1002.5	2	−1.6	44.7
				FPPPPCCSGVR	637.3	2	−2.5	
				PPPPCCSGVR	563.7	2	−3.0	

dx.doi.org/10.1021/jf4023367 | J. Agric. Food Chem. 2013, 61, 8949−8958

Table 2. continued

allergen	procedure	log(e) protein[b]	attribution/residue alignment	detected peptide sequence	m/z[c]	z[d]	log(e) peptide[b]	% protein coverage
				GIKPNVAAGL PSQCGVR	862.4; 575.3	2	−3.1	

[a]Attribution of tryptic fragments, aa residue alignment, and protein coverage are shown. [b]log(e): the base-10 log of the expectation that any particular protein assignment was made at random (*E*-value). [c]Mass/charge. [d]Charge of precursor ion.

Table 3. Quantitative Analysis of the Putative Allergens Cit cl 1, Cit cl 2, and Cit cl 3 by Means of Spectral Counting Approach, According to the Protein Extraction Procedures

	% tandem mass spectra[a]			
procedure	gel slice[b]	Cit c 1	Cit c 2	Cit c 3
1	a	**95.5**	0	0
	b	54.4	0	0
	c	0	0	5.8
	d	5.3	0	0
	e	31.1	0	0
2	a	**100**	0	0
	b	74.7	0	0
	c	0	0	0
	d	0	0	0
	e	23	0	0
3	a	**96**	0	0
	b	43.8	0	0
	c	5.7	**8.6**	7.7
	d	16.3	0	0
	e	13	0	0
4	a	93	0	0
	b	52.5	0	0
	c	0	0	**35**
	d	14.3	0	0
	e	30	0	0

[a]% tandem mass spectra = the total number of tandem mass spectra that match peptides to the particular allergen divided by the total number of spectra assigned to all proteins found in each slice. Values in bold are the best found for each allergen with every extraction procedure. [b]Gel slice: a (27−23 kDa); b (18−12 kDa); c (12−8 kDa); d (130−110 kDa); e (55−40 kDa).

and Thr24. The identified protein coverage, excluding the signal peptide, totaled 44.7%, whereas the theoretical isoelectric point and the molecular weight of mature protein were 9.85 and 10.84 kDa, respectively. Initial peptide of mature protein identified by mass spectrometry showed high amino acid sequence identity (77.8%) with N-terminal of both Mal d 3 (*Malus x domestica*) and Pru p 3 (*Prunus dulcis x Prunus persica*) allergens. Also the other two identified matched with the same allergens, showing sequence identity by 68.7% and 62.5%, respectively. Finally, sequence alignments of putative Cit cl 3 protein with available full or partial sequences of the nsLTPs in the genus *Citrus* showed identity scores ranging from 72 to 90% (Figure 2c). Three peptides with a high percentage of mass spectra were identified in the samples obtained following procedure 4. Only two peptides with low percentages of mass spectra were instead identified under extraction procedures 1 and 3 (Table 3, slices c).

All together, these data pointed out that the Cit cl 1, Cit cl 2, and Cit cl 3 proteins, belonging to the GLPs, profilins, and nsLTPs respectively, are expressed in *clementine* fruit peel. These proteins showed considerable sequence similarity with allergens from taxonomically distant (e.g., peach and apple) and close (e.g., orange and lemon) species. Quantitative analyses performed through the spectral counting strategy revealed that Cit cl 1 had a higher expression level than Cit cl 2 and Cit cl 3. A deeper proteomic approach of *clementine* fruit pulp will define whether these proteins are also expressed in this tissue.

Though further functional analysis of the identified proteins would be required, our study provides the molecular basis for

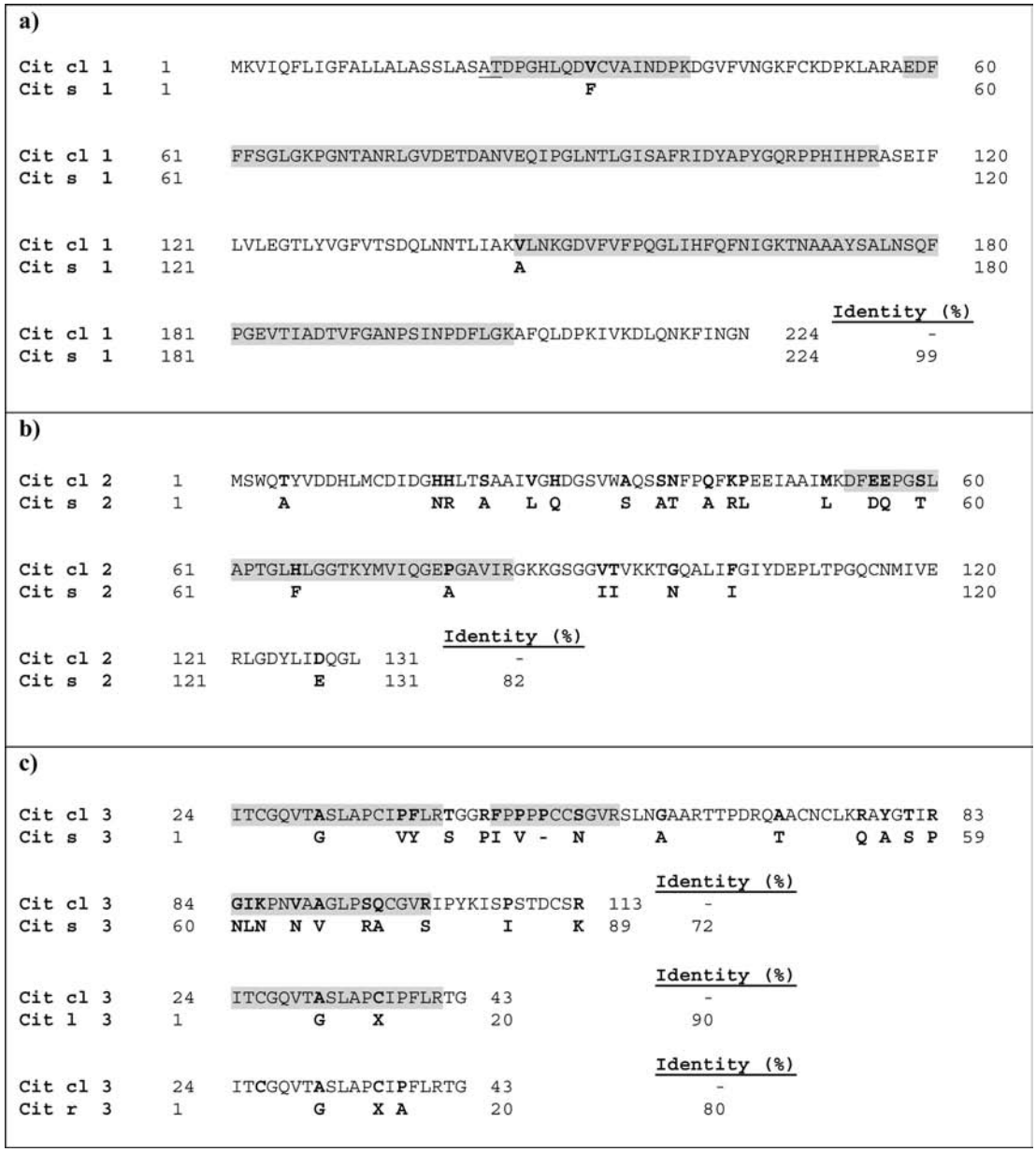

Figure 2. Alignments of amino acid sequences for Cit cl 1, Cit cl 2, and Cit cl 3 with homologous proteins for known allergens of citrus fruits. Only differential residues were shown. The experimentally determined amino acid sequences in this study are marked in gray. (a) Alignment of Cit cl 1 (clementine0.9_021092m) and sweet orange GLP Cit s 1.[12] The signal peptide cleavage site positioned between Ala22 and Thr23 was underlined. (b) Alignment of Cit cl 2 (clementine0.9_025303m) and sweet orange profilin Cit s 2 (Accession No. P84177). (c) Alignment of Cit cl 3 (clementine0.9_025486m) with sweet orange nsLTP Cit s 3 (Accession No. Q6EV47), lemon nsLTP Cit l 3 (Accession No. P84160), and tangerine nsLTP Cit r 3 (www.allergen.org).

the in vitro and in vivo assays to confirm the clinical relevance of *C. clementina* allergens.

Protein Structure. Since the primary structure of Cit cl 1 was almost completely characterized by mass spectrometry, it was used as a template to search a homologous protein structure in the RCSB Protein Data Bank. A correspondence was found with a primary sequence of barley germin with oxalate oxidase activity.[40−42] A raw model of Cit cl 1 was used for preliminary docking modeling to build a possible putative tridimensional structure of this GLP putative allergen. The secondary structure of the monomer resulted in an α-helix folding and two β-sheets (Figure 3A). The quaternary structure was established by assembling the monomers in a dimer (Figure 3B) and by arranging them into a homohexameric form (Figure 3C). As the structural similarity between proteins is a very good predictor of functional similarity,[43] this model confirmed the structural relationship with the oxalate oxidase of barley and assigned this putative function to Cit cl 1. The computed p*I* and monoisotopic mass of predicted native structure (5.5 and 131.3 kDa respectively) are in compliance with the experimental results on SDS−PAGE.

The conserved N-glycosylation site of GLPs[13] is expected at the Asn138 (Figure 3D), as reported in the homologue protein Cit s 1 of *C. sinensis*, in which the complex N-linked glycans play a prominent role in the IgE-binding.[10,12] The N-glycosylation site in the tridimensional model of Cit cl 1 is located in α-helix proximity (Figure 3C, see legend). In the predicted 3D Cit cl 1 homoexamer, the human IgE-binding

dx.doi.org/10.1021/jf4023367 | J. Agric. Food Chem. 2013, 61, 8949−8958

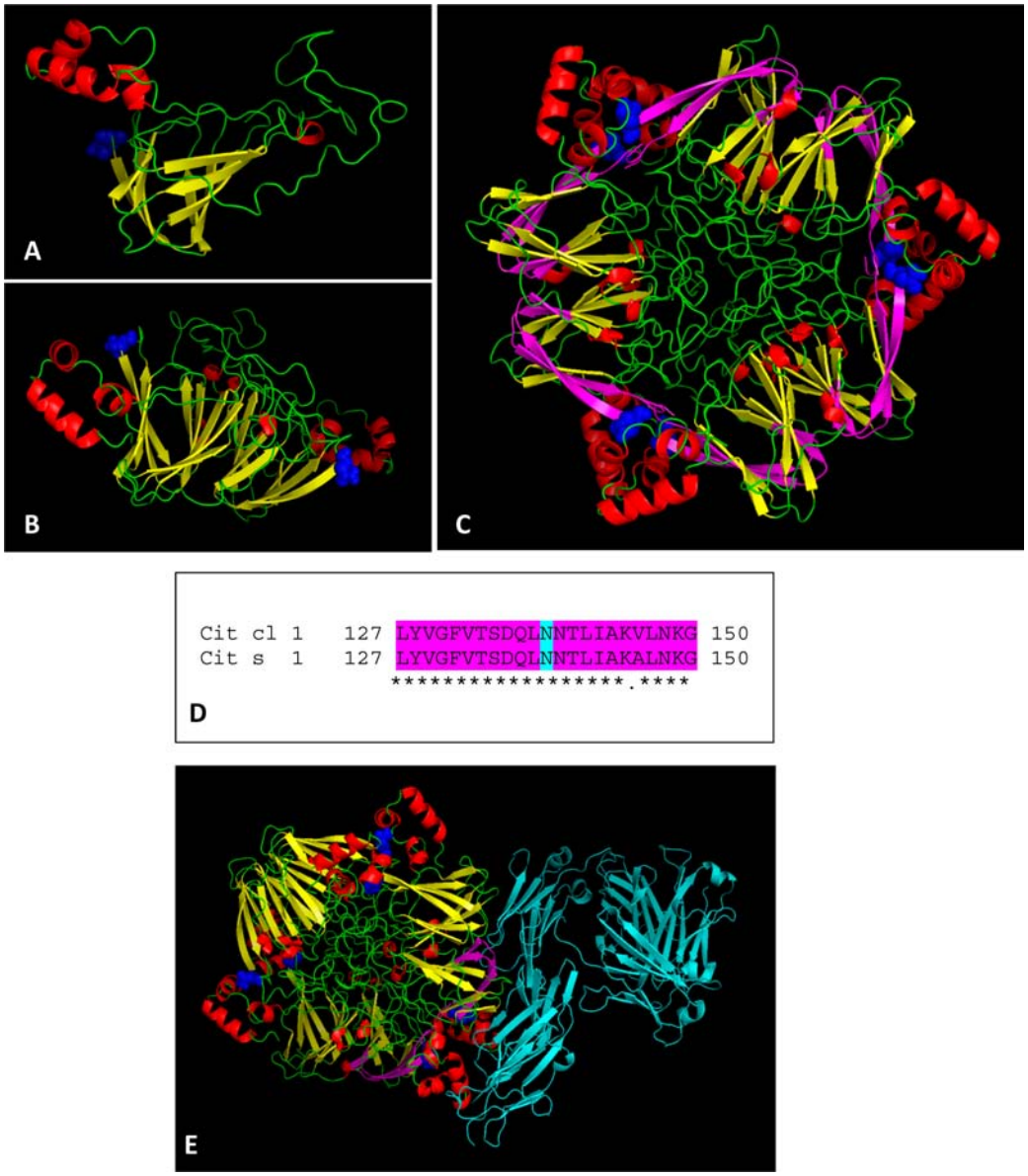

Figure 3. Structures of the Cit cl 1 revealed a typical alpha/beta fold. Secondary structure elements displayed α-helices (red), β-sheets (yellow), and loops and turns (green) in the monomer (A), dimer (B), and quaternary homohexamer structures (C). The blue spheres represent putative N-glycosylation site at Asn138. (D) Amino acid alignment between sequences of Cit cl 1 and Cit s 1 corresponding to the putative IgE epitope (purple) containing the N-glycosylation site (blue). The distribution of the Cit cl 1 epitopes over the homohexamer quaternary structure is reported in C. (E) Prediction of human IgE-fc (cyan) bound to the epitopes of the 3D Cit cl 1 allergen.

simulation occurs at this epitope (Figure 3E). As to future perspectives, it would be interesting to find experimentally the IgE epitope of Cit cl 1, and document the association of the structure with clinical phenomena.

■ ASSOCIATED CONTENT

Supporting Information

Tables listing proteins identified in gel slices a–e for each extraction procedure. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: s.mazzuca@unical.it.

Funding

This research was funded by the project "My Darling Clementine: un prodotto salutistico nuovo e innovativo dalle clementine e dal limone di Calabria" (POR FESR 2007-2013).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Authors are grateful to Dr. Ivano Eberini for his valuable advice on the structural analysis of proteins.

■ REFERENCES

(1) Guimarães, R.; Barros, L.; João, C. M.; Barreira, J. C. M.; Sousa, M. J.; Carvalho, A. M.; Ferreira, I. C. Targeting excessive free radicals with peels and juices of citrus fruits: grapefruit, lemon, lime and orange. *Food Chem. Toxicol.* **2010**, *48*, 99–106.

(2) Goulas, V.; Manganaris, G. A. Exploring the phytochemical content and the antioxidant potential of Citrus fruits grown in Cyprus. *Food Chem.* **2012**, *131*, 39−47.
(3) Ibáñez, M. D.; Sastre, J.; San Ireneo, M. M.; Laso, M. T.; Barber, D.; Lombardero, M. Different patterns of allergen recognition in children allergic to orange. *J Allergy Clin. Immunol.* **2004**, *113*, 175−177.
(4) Ahrazem, O.; Ibáñez, M. D.; López-Torrejón, G.; Sanchez-Monge, R.; Sastre, J.; Lombardero, M.; Barber, D.; Salcedo, G. Lipid transfer proteins and allergy to oranges. *Int. Arch. Allergy Immunol.* **2005**, *137*, 201−210.
(5) Iorio, R. A. S.; Calamelli, E.; Pula, C.; Lodolini, M.; Scamardella, F.; Pession, A.; Ricci, G. Citrus allergy from pollen to clinical symptoms. *PLoS One* **2013**, *8*, e53680 DOI: 10.1371/journal.-pone.0053680.
(6) Ebo, D. G.; Ahrazem, O.; López-Torrejón, G.; Bridts, C. H.; Salcedo, G.; Stevens, W. J. Anaphylaxis from mandarin (*Citrus reticulata*): identification of potential responsible allergens. *Int. Arch. Allergy Immunol.* **2007**, *144*, 39−43.
(7) Naruse, A.; Osako, J.; Tsuruta, D.; Yanagihara, S.; Ishii, M.; Kobayashi, H. A case of anaphylaxis caused by lemon sorbet. *J. Allergy Ther.* **2012**, *3*, 112−117.
(8) Navarro, L. A.; Pastor-Vargas, C.; Liñana, J. J.; Martínez, I .; Maroto, A. S.; Vivanco, F.; Bartolomé, B. Anaphylaxis due to orange soft drinks. *J. Invest. Allergol. Clin. Immunol.* **2012**, *22*, 297−299.
(9) Glaspole, I. N.; de Leon, M. P.; Rolland, J. M.; O'Hehir, R. E. Anaphylaxis to lemon soap: citrus seed and peanut allergen cross-reactivity. *Ann. Allergy Asthma Immunol.* **2007**, *98*, 286−289.
(10) Ahrazem, O.; Ibáñez, M. D.; López-Torrejón, G.; Sanchez-Monge, R.; Sastre, J.; Lombardero, M.; Barber, D.; Salcedo, G. Orange germin-like glycoprotein Cit s 1: an equivocal allergen. *Int. Arch. Allergy Immunol.* **2006**, *139*, 96−103.
(11) Crespo, J. F.; Retzek, M.; Foetisch, K.; Sierra-Maestro, E.; Cid-Sanchez, A. B.; Pascual, C. Y.; Conti, A.; Fellu, A.; Rodriguez, J.; Vieths, S.; Scheurer, S. Germin-like protein Cit s 1 and profilin Cit s 2 are major allergens in orange (*Citrus sinensis*) fruits. *Mol. Nutr. Food. Res.* **2006**, *50*, 282−290.
(12) Pöltl, G.; Ahrazem, O.; Paschinger, K.; Ibáñez, M. D.; Salcedo, G.; Wilson, I. B. H. Molecular and immunological characterization of the glycosylated orange allergen Cit s1. *Glycobiology* **2007**, *17*, 220−230.
(13) Pignataro, V.; Canton, C.; Spadafora, A.; Mazzuca, S. Proteome from lemon fruit flavedo reveals that this tissue produces high amounts of the Cit s 1 germin-like isoforms. *J. Agric. Food Chem.* **2010**, *58*, 7239−7244.
(14) López-Torrejón, G.; Ibáñez, M. D.; Ahrazem, O.; Sánchez-Monge, R.; Sastre, J.; Lombardero, M.; Barber, D.; Salcedo, G. Isolation, cloning and allergenic reactivity of natural profilin Cit s 2, a major orange allergen. *Allergy* **2005**, *60*, 1424−1429.
(15) Dunwell, J.; Gibbings, J.; Mahmood, T.; Naqvi, S. S. Germin and germin-like proteins: evolution, structure, and function. *Crit. Rev. Plant Sci.* **2008**, *27*, 342−375.
(16) Membre, N.; Bernier, F.; Staiger, D.; Berna, A. *Arabidopsis thaliana* germin-like protein: common and specific features point to a variety of functions. *Planta* **2000**, *21*, 345−354.
(17) Carter, C.; Thornburg, R. W. Tobacco Nectarin I: purification and characterization as a germin-like, manganese superoxide dismutase implicated in the defense of floral reproductive tissue. *J. Biol. Chem.* **2000**, *275*, 36726−36733.
(18) Hauser, M.; Egger, M.; Wallner, M.; Wopfner, N.; Schmidt, G.; Ferreira, F. Molecular properties of plant food allergens: a current classification into protein families. *Open Immunol J.* **2008**, *1*, 1−12.
(19) Kandasamy, M. K.; McKinney, E. C.; Meagher, R. B. Plant profilin isovariants are distinctly regulated in vegetative and reproductive tissues. *Cell Motil. Cytoskeleton* **2002**, *52*, 22−32.
(20) Ferreira, F.; Hawranek, T.; Gruber, P.; Wopfner, N.; Mari, A. Allergic cross reactivity: from gene to the clinic. *Allergy* **2004**, *59*, 243−267.
(21) Hauser, M.; Roulias, A.; Ferreira, F.; Egger, M. Panallergens and their impact on the allergic patient. *Allergy Asthma Clin. Immunol.* **2010**, *6*, 1.
(22) Salcedo, G.; Sanchez-Monge, R.; Barber, D.; Díaz-Perales, A. Plant non-specific lipid transfer proteins: an interface between plant defence and human allergy. *Biochim. Biophys. Acta* **2007**, *1771*, 781−791.
(23) Carvalho, A.; Gomes, V. M. Role of plant lipid transfer proteins in plant cell Physiology-A concise review. *Peptides* **2007**, *28*, 1144−1153.
(24) Hyun, T. K.; Kim, J. S. Genomic identification of putative allergen genes in woodland strawberry (*Fragaria vesca*) and mandarin orange (*Citrus clementina*). *Plant Omics J.* **2011**, *4*, 428−434.
(25) Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L. Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. *Anal. Chem.* **2010**, *82*, 2272−2281.
(26) Petersen, T. N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785−786.
(27) Wang, W.; Scali, M.; Vignani, R.; Spadafora, A.; Sensi, E.; Mazzuca, S.; Cresti, M. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis* **2003**, *24*, 2369−2375.
(28) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **1996**, *68*, 850−885.
(29) Mintseris, J.; Pierce, B.; Wiehe, K.; Anderson, R.; Chen, R.; Weng, Z. Integrating Statistical Pair Potentials into Protein Complex Prediction. *Proteins* **2007**, *69*, 511−520.
(30) Pierce, B. G.; Hourai, Y.; Weng, Z. Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. *PLoS One* **2011**, *6*, 246−257.
(31) Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **2004**, *32* (Web server issue), W96−W99, DOI: 10.1093/nar/gkh354.
(32) Kozakov, D.; Hall, D. R.; Beglov, D.; Brenke, R.; Comeau, S. R.; Shen, Y.; Li, K.; Zheng, J.; Vakili, P.; Paschalidis, I. C.; Vajda, S. Achieving reliability and high accuracy in automated protein docking: Cluspro, PIPER, SDU, and stability analysis in CAPRI rounds 13−19. *Proteins* **2010**, *78*, 3124−3130.
(33) Zaman, U.; Abbasi, A. Isolation, purification and characterization of a non specific lipid transfer protein from *Cuminum cyminum*. *Phytochemistry* **2009**, *70*, 979−987.
(34) Egger, M.; Hauser, M.; Mari, A.; Ferreira, F.; Gadermaier, G. The Role of Lipid Transfer Proteins in Allergic Diseases. *Curr. Allergy Asthma Rep.* **2010**, *10*, 326−335.
(35) Berger, I. J.; Freitas-Astua, J.; Reis, M. S.; Targon, M. L. P. N.; Machado, M. A. *In silico* prediction of gene expression patterns in Citrus flavedo. *Genet. Mol. Biol.* **2007**, *30*, 752−760.
(36) Yvon, M.; Chabanet, C.; Pélissier, J. P. Solubility of peptides in trichloroacetic acid (TCA) solutions. Hypothesis on the precipitation mechanism. *Int. J. Pept. Protein Res.* **1989**, *34*, 166−176.
(37) Jensen-Jarolim, E.; Schmid, B.; Bernier, F.; Berna, A.; Kinaciyan, T.; Focke, M.; Ebner, C.; Scheiner, O.; Boltz-Nitulescu, G. Allergologic exploration of germins and germin-like proteins, a new class of plant allergens. *Allergy* **2002**, *57*, 805−810.
(38) Requena, L.; Bornemann, S. Barley (*Hordeum vulgare*) oxalate oxidase is a manganese-containing enzyme. *Biochem. J.* **1999**, *343*, 185−190.
(39) Lane, B. G. Oxalate, germin, and higher-plant pathogens. *IUBMB Life* **2002**, *53*, 67−75.
(40) Hee, J. K.; Pesacreta, T. C.; Triplett, B. A. Cotton-fiber germin-like protein. II: Immunolocalization, purification, and functional analysis. *Planta* **2004**, *218*, 525−535.
(41) Woo, E. J.; Dunwell, J. M.; Goodenough, P. W.; Marvier, A. C.; Pickersgill, R. W. Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat. Struct. Biol.* **2000**, *7*, 1036−1040.

(42) Opaleye, O.; Rose, R. S.; Whittaker, M. M.; Woo, E. J.; Whittaker, J. W.; Pickersgill, R. W. Structural and spectroscopic studies shed light on the mechanism of oxalate oxidase. *J. Biol. Chem.* **2006**, *281*, 6428−6433.
(43) Baker, D.; Sali, A. Protein structure prediction and structural genomics. *Science* **2001**, *294*, 93−96.