JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

P39, a Novel Soybean Protein Allergen, Belongs to a Plant-Specific Protein Family and Is Present in Protein Storage Vacuoles

PING XIANG,[†] LISA M. BAIRD,[§] RUDOLF JUNG,[#] MICHAEL G. ZEECE,[⊥] JOHN MARKWELL,[△] AND GAUTAM SARATH*,[○]

British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada, Biology Department, University of San Diego, San Diego, California 92110, Pioneer-Hybrid International, Johnston, Iowa 50131, Department of Food Science and Technology, University of Nebraska, Lincoln, Nebraska 68583-0919, Department of Biochemistry, University of Nebraska, Lincoln, Nebraska 68583-0718, and Grain, Forage and Bioenergy Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Lincoln, Nebraska 68583-0737

Soybean lecithins are seeing increasing use in industry as an emulsifier and food additive. They are also a growing source of human food allergies, which arise principally from the proteins fractionating with the lecithin fraction during manufacture. A previous study (Gu, X.; Beardslee, T.; Zeece, M.; Sarath, G.; Markwwell, J. *Int Arch. Allergy Immunol.* **2001**, *126*, 218−225) identified several allergenic proteins in soybean lecithins and a soybean IgE-binding protein termed P39 was discovered. However, very little was known about this protein except that it was coded by the soybean genome. This paper investigates key biological and immunological properties of this potential soybean lecithin allergen. P39 is encoded by a multigene family in soybeans and in several other higher plants. The soybean P39-1 protein and its essentially indistinguishable homologue, P39-2, have been cloned and studied. These proteins and their homologues belong to a family of plant-specific proteins of unknown function. In soybeans, P39-1 is seed specific, and its transcript levels are highest in developing seeds and decline during seed maturation. In contrast, P39 protein was detectable only in the fully mature, dry seed. Subcellular fractionation revealed that P39 protein was strongly associated with oil bodies; however, immunolocalization indicated P39 was distributed in the matrix of the protein storage vacuoles, suggesting that association with oil bodies was an artifact arising from the extraction procedure. By the use of recombinant techniques it has also been documented that IgE-binding epitopes are present on several different portions of the P39-1 polypeptide.

KEYWORDS: P39; soybean lecithin; allergen; plant-specific protein family; protein; oil bodies; protein storage vacuoles

INTRODUCTION

Soybean lecithin is the most important industrial byproduct of oil processing. It is obtained through the water washing of the crude soybean oil followed by the separation and drying of the hydrated gums. The major components of soybean lecithins are phosphatides, phytoglycolipids, triglycerides, phytosterols, tocopherols and free fatty acids (*1*). It is an excellent food emulsifier and is widely used in processed foods, pharmaceuticals, and cosmetics. However, many commercial products that contain soybean lecithin are not labeled as containing soybean proteins.

Depending upon the sample and the detection methods, the reported protein content of different lecithin preparations varies widely. The highest reported protein concentration in a lecithin sample was 27200 ppm (*2*) and typically ranges between 100 and 500 ppm (*3*). Even though the overall protein level in soybean lecithins is not high, these proteins represent an important source of allergens (*3*–*5*). Several clinical cases of allergic reactions associated with soybean lecithin have been reported (*5*–*8*). In these cases, allergic responses were elicited by skin contact or ingestion of foods that contained soybean lecithin. This indicated that even a small amount of proteins present in soybean lecithin was able to cause allergic reactions in soybean-sensitive patients. Therefore, with the increased

* Author to whom correspondence should be addressed [e-mail Gautam.Sarath@ars.usda.gov; telephone (402) 472-4204; fax (402)472-4020].
[†] British Columbia Cancer Agency.
[§] University of San Diego.
[#] Pioneer-Hybrid International.
[⊥] Department of Food Science and Technology, University of Nebraska.
[△] Department of Biochemistry, University of Nebraska.
[○] U.S. Department of Agriculture.

10.1021/jf073292x CCC: $40.75
Published on Web 02/20/2008

usage of soybean lecithin, more emphasis should be placed on the study of protein allergens in soybean lecithin preparations.

Gel electrophoretic analyses of proteins isolated from lecithin samples have documented the presence of several polypeptides. Four major bands were observed at molecular masses of 39, 20, 12, and 7 kDa, respectively (*3*). Immunoblotting with sera from soybean-allergic individuals revealed that all four protein bands were IgE-binding proteins. An additional IgE-binding protein of 57 kDa was also detected (*3*). On the basis of N-terminal sequencing, the 12 kDa protein (P12) was identified as the large subunit of soybean 2S albumin, the 20 kDa protein (P20) was identified as soybean Kunitz trypsin inhibitor, and the 7 kDa protein (P7) was identified as a known allergen originating from dust mites (*3*). The 57 kDa protein (P57) was not identified because its N terminus was apparently blocked. The 39 kDa protein (P39) possessed the N-terminal sequence SDKEDSVFKGIN and was identified as a soybean protein with unknown function (*3*).

Because P39 was the major protein as well as the dominant IgE binder in soybean lecithin samples, more studies on both the biological and immunological characteristics of P39 were undertaken.

MATERIALS AND METHODS

Materials. All chemicals were of reagent grade or better and obtained Fisher Scientific (Chicago, IL), VWR Inc. (Kansas City, MO), or Sigma-Aldrich (St. Louis, MO). Sources of specialty chemicals are identified in the sections below. Soybean seeds (*Glycine max*. Merr. cv. Dunbar) were obtained from Dr. Thomas E. Clemente, University of Nebraska, Lincoln. Seeds were planted in a greenhouse in Lincoln, NE, under a 16 h, ~26–30 °C day/8 h, ~22–26 °C night growth regimen, using supplemental lighting from halide lamps (200 mol of photons m^{-2} s^{-1}) in a soil mixture consisting of 40% Canadian peat, 40% coarse vermiculite, 15% masonry sand, and 5% screened topsoil, amended with 4.45 kg of Waukesha fine lime m^{-3}. Plants were watered biweekly with a nutrient solution containing 200 ppm of N and with tap water as needed. Soybean seeds were first harvested at 8 days after anthesis, then at the sizes of approximately 5, 15, 50, 90, 180, 300, 420, and 200 mg of seed fresh weight. Leaves were a mixture of young and mature leaves. Root and stem tissues were harvested from 87-day-old plants. Flowers were harvested as needed.

Identification of P39 and Similar Proteins. An initial BLAST search of P39 based on the partial N-terminal sequence SDKEDSVFKGIN was performed against a soybean EST database possessed by Pioneer-Hybrid International (courtesy of Dr. Rudolf Jung). The first identified protein was named P39-1. Its cDNA clone was provided by Dr. Rudolf Jung. A more complete search based on the coding region of P39–-1 for similar proteins from soybean and other species was performed using a TBLASTN search against the Soybean EST database from Pioneer-Hybrid International, as well as other plant EST databases using freely available BLAST searching tools NCBI BLAST 2.0 (*9*). Sequence alignment was performed using CLUSTAL W [http://www.ebi.ac.uk/Tools/clustalw/index.html; (*10*)]. Aligned files were used to generate conserved boxes using the BoxShade 3.21 program (http://www.ch.embnet.org/software/BOX_form.html). A phylogenetic tree was built using Phylip [http://bioweb.pasteur.fr/cgi-bin/seqanal/drawtree.pl; (*11*)]. Plant sequences and accession numbers used for generating these data were as follows: **P39-1-6**, *Glycine max*, 42722638, 13788114, 16105921, 22932482, 6455812, and 58020550, respectively; **Atha1-5**, *Arabidopsis thaliana*, At1g54860, At3g06035, At5g19250, At5g19240, and At5g19230, respectively; **Ahyp-1**, *Arachis hypogaea*, 110810512; **Ghir-1**, *Gossypium hirsutum*, 109853380; **Ljap-1**, *Lotus japonicum*, 47953398; **Mtra-1**, *Medicago trancatulta*, 27232951; **Ntab-1**, *Nicotiana tabacum*, 94323945; **Slyc-1**, *Solanum lycopersicum*, 62933876; **Jreg-1**, *Juglans regia*, 52125082; **Hvul-1**, *Hordeum vulgare*, 57828708; **Sbic-1**, *Sorghum bicolor*, 13317695; **Zmay-1**, *Zea mays*, 67016248; **Osat-1**, *Oryza sativa*, 112786892; **Pvir-1**, *Panicum virgatum*, 59864141; **Taes-1**, *Triticum aestivum*, 39009049.

Northern Blot Analyses. Total RNA was extracted using the RNeasy plant RNA extraction system from Qiagen. Total RNA were separated on a 1.2% agarose−formaldehyde gel, and the gel was stained with ethidium bromide to visualize RNA and confirm loading. Subsequently, RNA was transferred to Hybond N+ membrane and probed with dig11-dUTP labeled P39-1 probe. Hybridization and washes were performed at 50 °C following the manufacturer-supplied protocols (Roche Applied Science, Indianapolis, IN). Hybridizing bands were detected as described for Southern blots. After detection, membranes were stripped with 50% formamide/5%SDS /50 mM Tris-HCl (pH 7.5) two times at 80 °C and stored in 2× SSC until reuse.

Recombinant P39-1 Protein Construction, Purification, and P39-1 Antibody Production. PCR primers (P39FNDEI, 5′-catatgagtgataaggaggacagt; P39RECORI, 5′-gaattcatcaaggtagttgatgag) were designed for read-through of the P39-1 mature sequence. PCR products were purified from 1% agarose gel and ligated into the pBAD/TOPO vector using the TA system (Invitrogen Inc.). The ligated DNA was transformed into *Escherichia coli* Top10 cells following manufacturer-recommended protocols (Invitrogen Inc.). Constructs containing inserts with the predicted sizes and right orientation were verified through endonuclease mapping and sequence verified in the Genomics Core Research Facility of the University of Nebraska, Lincoln. Sequence-verified inserts were subsequently obtained by double digestions using *Nde*I and *Eco*RI restriction endonucleases and ligated into the pET28a vector (Novagen) to produce recombinant P39 for antibody production. This recombinant protein was purified as described (*12*). Purified fractions were dialyzed against water, lyophilized, and used to raise antibodies in rabbits (Cocalico Biologicals, Inc., Reamstown, PA). Antibodies were partially purified by ammonium sulfate precipitation, followed by dialysis against TBS containing 0.02% azide. The partially purified P39 IgG antibodies were stored at −80 °C until use.

Oil body Extraction and Western Blots. Oil bodies were extracted of mature seeds according to the method of Tzen et al. (*13*). Two grams of seeds from different plant species or soybean varieties were incubated in water at 4 °C overnight and used for the oil body extraction. The extracted oil body fractions were treated with diethyl ether, and the gelatin-like protein layer between the diethyl ether and water phases was collected, precipitated with acetone, and used for the SDS-PAGE (*14*). Proteins were transferred to a nitrocellulose membrane (Bio-Rad, Hercules, CA) in a tank apparatus overnight at 10 V in transfer buffer (10 mM Tris-HCl, 100 mM glycine, 10% methanol) at 4 °C. Membranes were blocked overnight with TBST (25 mM Tris Cl, pH 7.5, 150 mM NaCl, 0.05% Tween-20), containing 5% nonfat dry milk and subsequently incubated with partially purified rabbit antiP39 IgG antibodies (1:5000) in blocking buffer for 2 h. After extensive washes (3 × 10 min) with TBST, membranes were incubated for 1 h with horseradish peroxidase conjugated to goat anti-rabbit IgG antibody (Sigma, St. Louis, MO) diluted 1:10000 in TBST. Membranes were washed as above with TBST, and antigen–antibody complexes were detected by chemiluminescence using chemicals obtained from Pierce Chemical Co. (Rockford, IL) or ECL plus reagent (Amersham Pharmacia, Piscataway, NJ) and imaged using Biomax ML film (Kodak). Films were developed using an automatic developer (Kodak M35A, X-OMAT Processor).

Electron Microscopy of Soybean Seeds. To process soybean seeds for electron microscopy, mature seeds were first imbibed overnight in water, and fresh green seeds were processed directly. The seed coat was removed, and seeds were sliced into sections <1 mm thick. Sections were fixed with 4% glutaraldehyde in 100 mM sodium phosphate buffer, pH 7.2, for 2 h and washed three times in phosphate buffer and two times with water. Sections were then dehydrated in a graded ethanol series (10, 20, 30, 50, and 70%, 1 h in each) and stored in 70% ethanol. For embedding, sections were dehydrated through 100% ethanol and infiltrated in LR white resin for 3 days. The resin was polymerized overnight in a 60 °C oven. Sections were cut at 60−90 nm on a Reichert Ultracut II ultramicrotome and mounted on nickel grids. Antibody to P39 was used to localize the antigen according to standard procedures (*15*), and sections were examined on a Zeiss 900T electron microscope at 80 kV.

Recombinant Fusion Fragment Protein Constructs of P39-1. Three thioredoxin fusion proteins representing overlapping fragments

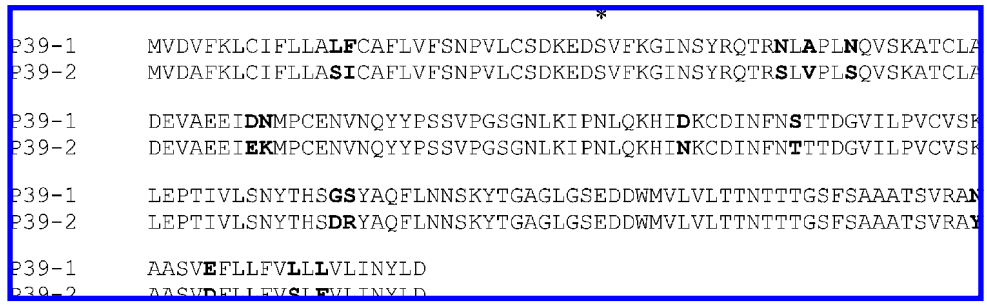

Figure 1. Sequence comparison of soybean P39-1 and P39-2 proteins. The mature sequences of both P39-1 and P39-2 proteins start with S28 and are denoted with an asterisk.

of P39-1 were constructed. PCR reactions were performed using additional P39-1 specific primers (P39F1NDEI, 5′-catatggtggacgttttcaa; P39F1ECORI, 5′-gaattctgcctttgaaacttgg; P39F2ECORI, 5′-gaattcggttggttccaattt; P39F3NDEI, 5′-catatgcccgtttgtgtctc). The expression of these recombinant P39-1 fragments proteins was in pBAD/Thio system as described (*12*) and purified accordingly. The purified proteins were lyophilized and stored at −80 °C until needed.

Immunoblotting Using Human Soybean Allergic Sera. Soybean-allergic sera and control human serum used in this study were obtained from Plasma Laboratories [Seattle, WA; (*16*)]. Aliquots of recombinant proteins (2 μg for each protein) were separated by SDS-PAGE, and immunoblots were processed essentially as described by Xiang et al. (*16*).

RESULTS

P39 Is Encoded by a Multigene Family in Soybean. A database search with the N-terminal 12 amino acid sequence of the lecithin P39 yielded two closely related ESTs encoding proteins, named P39-1 and P39-2 (accession no. 42722638 and 13788114, respectively), of 200 amino acids and a putative 27 amino acid signal peptide. The predicted mature sequence of both proteins probably starts at S28 (**Figure 1**; S28 is marked with an asterisk) and the first 12 amino acids of these proteins match exactly the reported N-terminal sequence for the protein isolated from soy lecithin (*3*).

Continued searches of public and a proprietary soybean EST database from Pioneer-Hybrid International (courtesy of Dr. Rudolf Jung) identified three other proteins with high sequence homology to P39-1 and P39-2. These proteins were named P39-3 through P39-5 (accession no. 16105921, 22932482, and 6455812; respectively). We also found a partial clone, termed P39-6 (accession no. 158020550), with some sequence homology to the soybean P39 family of proteins.

P39-1 and P39-2 share a very high sequence similarity (92% identity). In addition, they contain similar signal peptides and havethe same predicted cleavage site at S28 and an identical N-terminal sequence (**Figure 1**). Therefore, it is likely that probes generated to P39-1 will also detect P39-2. For instance, at least two bands were detected in all lanes of a Southern blot of endonuclease-digested soybean genomic DNA using dig11-dUTP-labeled P39-1 cDNA as probe, which possibly represented the P39-1 and P39-2 gene locus (data not shown). In the results described and discussed below, we use P39-1 for convenience and assume that our probes would not have differentiated between P39-1 and P39-2.

On the other hand, P39-3, P39-4, and P39-5 share much less similarity with P39-1 and P39-2 (sequence identity around 31–35%), even though they share greater identity to each other (92% identity between P39-3 and P39-4; 71 and 67% identity between P39-5 and P39-3 and between P39-5 and P39-4, respectively). Therefore, they apparently form a separate group, and are less likely to cross-react to the probes generated to detect P39-1.

P39-1 Is a Member of a Unique Plant Protein Family. Over 100 related ORFs and/or ESTs from other plant species were found in the existing databases. These proteins appeared to belong to the same protein family. All of these proteins have a similar size, between 170 and 200 amino acids in length, and exhibit >29% identity among the group. An alignment of selected sequences indicated the presence of conserved domains and several invariant residues (**Figure 2a**). Of the 12 invariant residues, 4 were conserved cysteines that occurred at similar points in all of these proteins, strongly suggesting the presence of intramolecular disulfides in the folded molecule.

A phenogram generated from the available protein sequences is shown in **Figure 2b**. In this P39 family tree, soybean P39-1 and P39-2 share most similarity to each other (**Figure 2b**, circle) and to one protein from *A. thaliana* Atha1 than to the four other P39-like proteins from soybean. Other closely related sequences include proteins from two other legumes, *M. trancatula* Mtra-1 and *A. hypogaea* Ahyp-1. Sequences from other dicots appear to share various levels of similarity to each other and the soy P39-1 protein. Interestingly, all of the monocot P39-like sequences analyzed cluster together in a separate branch with differentiation between P39-like sequences from C_3 species such as rice, *O. sativa* Osat-1, and barley, *H. vulgare* Hvul-1, with respect to the C_4 species such as *S. bicolor* Sbic-1, *P. virgatum* Pvir-1, and maize *Z. mays* Zmay-1 (**Figure 2b** (square).

Expression of P39-1 mRNA. To determine the abundance and expression patterns of P39-1 mRNA, total RNA was isolated from soybean roots, stems, leaves, flowers, seeds, and root nodules and detected with a dig11-dUTP-labeling P39-1 DNA probe (**Figure 3**). P39 mRNA was specifically detected in soybean seed and not in the other tissues, indicating seed-specific expression of the P39-1 and P39-2 genes (**Figure 3a**). These results agreed with the limited expression information available in public databases, where P39-1 and P39-2 cDNAs were indicated to be enriched in seeds, with limited abundance in somatic embryos and low expression in the germinating seeds. This type of expression pattern is similar to the major seed storage proteins (Dr. Rudolf Jung, Pioneer-Hybrid International, personal communication).

The expression level of P39-1 varied during seed development (**Figure 3b**). Its mRNA was not detectable in very young seeds, 8 days after flowering; however, P39-1 transcripts were detectable in seeds that weighed approximately 5 mg. The mRNA level for P39-1 increased during early seed development and peaked when seeds were around 50–90 mg in fresh weight, and then started to decline in more mature seeds, and was nearly absent in fully mature seeds (**Figure 3b**).

To verify the presence of P39-1 and related proteins in soybeans, we generated an antibody to a recombinant P39-1 protein. This polyclonal antibody preparation specifically recognized a 39 kDa protein in soybean lecithin. After longer

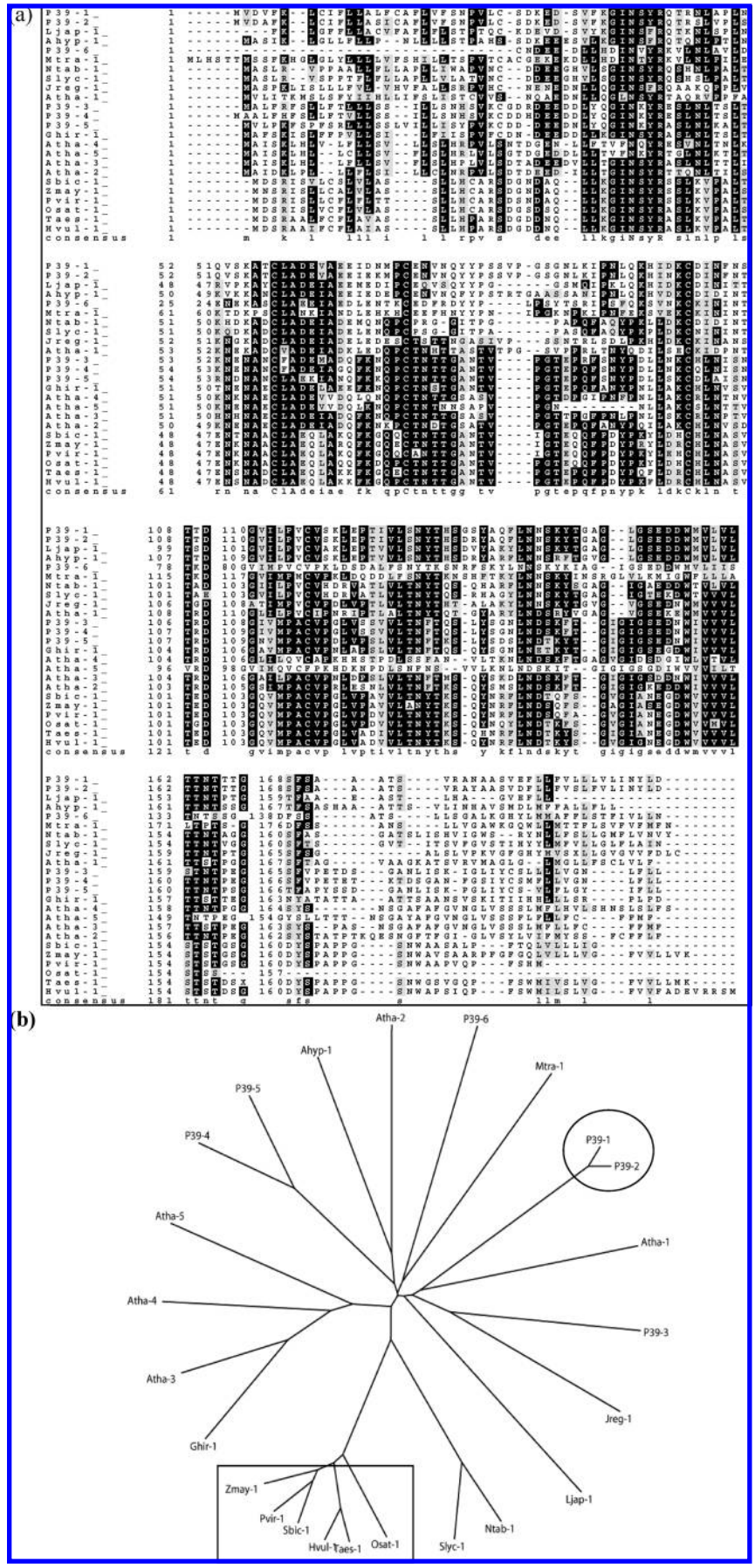

Figure 2. Sequence alignment (**a**) of several P39-related protein sequences from the databases, and (**b**) phylogenetic tree of these proteins. The soybean P39-1 and P39-2 are circled. The monocot P-39 protein clade is boxed.

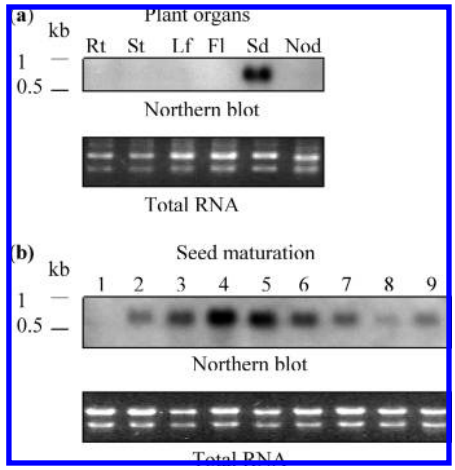

Figure 3. Expression of P39-1 and P39-2 genes in soybean tissues: (**a**) seed-specific expression pattern of P39 mRNA [(Rt), stems (St), leaves (Lf), flowers (Fl), young seeds (Sd) (weight about 90 mg), and mature root nodules (Nod)]; (**b**) P39 mRNA expression during seed maturation (lane 1, young pods 8 days after flowering; lanes 2−8, green seeds with approximate fresh weights of 5, 15, 50, 90, 180, 300, and 420 mg; lane 9, mature seeds}.

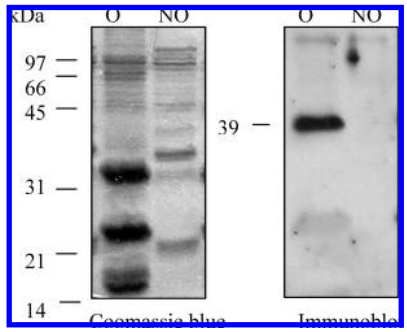

Figure 4. P39 fractionates with oil bodies: lane O, oil body fraction; lane NO, proteins from the corresponding oil body depleted soluble fractions.

exposure, a protein with an approximate molecular mass of 57 kDa was also detected. Having confirmed the specificity of the antibodies to P39 in soybean extracts, we used these antibodies to study P39 localization and abundance in soybean seed extracts, in extracts from other plants, and during soybean seed maturation.

P39-1 Is Concentrated in Soybean Oil Body Fraction. Initial experiments using a variety of extraction protocols failed to detect P39 protein in total seed extracts, suggesting a low abundance of this allergen in seeds. Because P39 was first identified from soybean lecithin, it was possible that P39-1 was associated with soybean oil and could be concentrated in the oil body fraction. To test this possibility, soybean oil bodies were separated specifically from other soluble fractions of soybean seed extracts using the protocol of Tzen et al. (*13*). The final oil body preparation was enriched for four major proteins with apparent molecular masses of 34, 24, 18, and 17 kDa (**Figure 4**). This result is similar to previous data reported by Herman (*17*). Antibodies raised against recombinant P39-1 strongly recognized one band at approximately 39 kDa in soybean oil body fractions prepared from soybean, but not proteins in the soluble fraction depleted of oil bodies (**Figure 4**).

To test the potential cross-reaction of the anti-P39-1 antibodies across species, oil body fractions were prepared from sesame, sunflower, and peanut seeds (**Figure 5**). No band migrating at

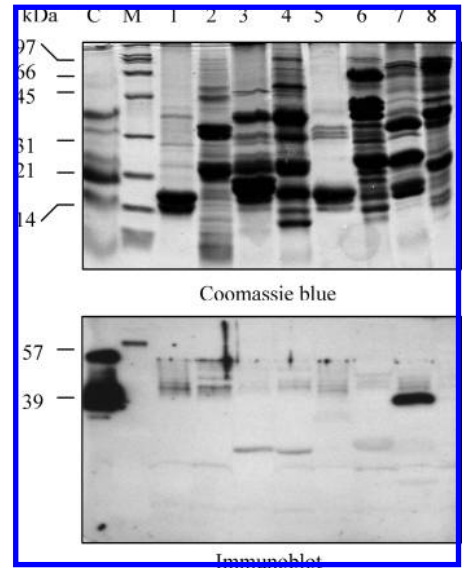

Figure 5. Detection of soybean P39-like proteins in oil body preparations from other plants: lane C, soybean lecithin; lane M, molecular mass markers; lanes 1, 3, 5, and 7, proteins from sesame, sunflower, peanut, and soybean oil body fractions, respectively; lanes 2, 4, 6, and 8, proteins from the corresponding soluble extracts depleted of oil bodies.

the 39 kDa position was detected in the oil body fractions from any of these plants. However, several higher molecular mass bands in the 42−52 kDa region were detected. After a longer exposure of the X-ray film or upon increased protein loading, several additional immunoreacting bands between 39 and 57 kDa were detected in the soybean oil body fractions. The relationships of these proteins cross-reacting to anti-P39-1 antibodies to P39 are not clear.

P39-1 Is Possibly Processed during Development. Because soybean P39-1 was concentrated in the oil body fractions, oil bodies from soybean seeds of different sizes were prepared and subjected to Western blotting with the anti-P39-1 antibody. To obtain reasonable yields of oil bodies, only seeds of >150 mg of fresh weight were used. Our results indicated that a protein with an apparent molecular mass of 39 kDa was detected only in dried mature seeds. In extracts from seeds of other ages, several immunoreactive bands between 42 and 55 kDa were detected and could represent differentially processed forms of P39-1 (**Figure 6**). Alternatively, these results could be artifacts and might indicate that P39-1 is best extracted from fully mature seeds.

P39 Is Localized in Protein Storage Vacuoles. Electron microscopic immunocytochemical analysis indicated that P39-1 is primarily localized in the protein storage vacuoles of soybean seeds (**Figure 7**). There was also some labeling around oil bodies; however, this phenomenon was also observed in the preimmune control. Therefore, the P39-1 protein is apparently associated with protein storage vacuoles. The labeling density of the young (100–200 mg of fresh weight) and older soybean seeds (300–400 mg of fresh weight) did not exhibit significant differences (**Figure 7**). These results are consistent with the Northern blot data on the relative abundance of P39-1 mRNA in developing soybean seeds (**Figure 3b**), which showed that P39-1 transcripts begin to accumulate in young seeds.

IgE Immunoblots of Recombinant P39-1 and Its Fusion Fragments. Potential allergenicity of P39-1 was evaluated using thioredoxin fusions of the full-length P39-1 protein and fragments derived from the entire molecule. Polyclonal antibodies raised to the intact P39-1 protein bound efficiently to all of the

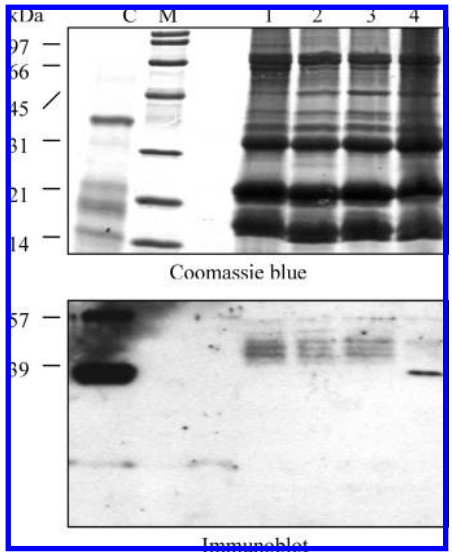

Figure 6. Accumulation of P39-1 protein during seed maturation: lane C, soybean lecithin; lane M, molecular mass markers; lanes 1–3, proteins from oil body fractions of seeds during maturation; lane 4, oil bodies from harvested seeds.

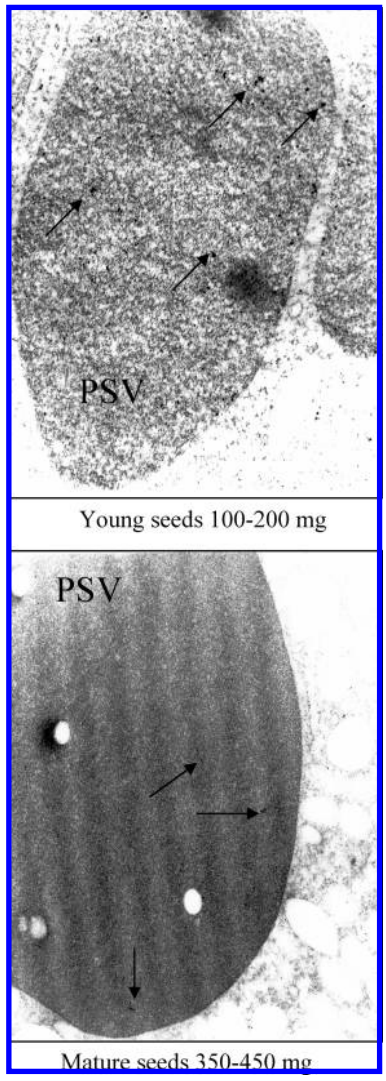

Figure 7. Immunogold localization of P39-1 in soybean seeds. Arrows show positions of immunogold complexes.

P39-1-derived fragments (**Figure 8a**). Interestingly, IgE in pooled sera from soybean-sensitive individuals (*16*) also readily

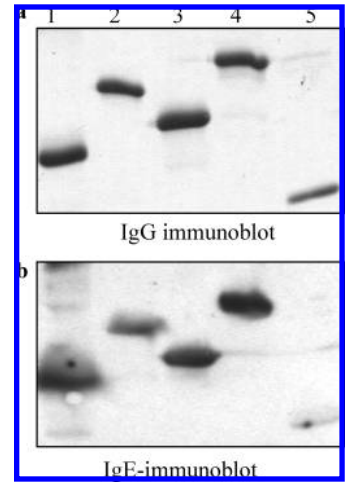

Figure 8. IgE binding to recombinant P39 and its fragments: purified recombinant fragment, P39f1 (lane 1), P39f2 (lane 2), P39f3 (lane 3), full-length P39 protein (lane 4), and thioredoxin fusion partner (lane 5) were separated by a 12.5% SDS gel and stained with Coomassie blue.

recognized all four constructs of P39-1 (**Figure 8b**), suggesting that there were potentially several epitopes spread throughout the entire length of the protein.

DISCUSSION

Soybeans contain a number of well-characterized seed protein allergens (*12, 18–21*). The relative allergenicity appears to be determined by several factors related to amino acid sequences of these proteins, which result in specific structural motifs (*22*). For example, many of the plant allergens such as the peanut Ara h 1, Ara h 3, soybean β-conglycinin, and soybean glycinin G1, G2, and Gly m Bd 28K proteins are members of the cupin protein superfamily (*16, 23*). Other structural features associated with many allergens are hydrophobicity and intramolecular disulfides, which can make them stable to a number of physical and chemical denaturants (*22, 24, 25*). Soybean P39 does not belong to the cupin superfamily, but is hydrophobic as judged from its original detection in soy lecithin (*3*) and its fractionation with oil bodies isolated from soybean seeds (**Figure 4**), and possesses four conserved cysteines (**Figure 2a**) that could potentially form intramolecular disulfide bonds. These characteristics point to its potential as an allergen. Furthermore, both the recombinant protein and its fusion fragments (**Figure 8**) efficiently bound IgE to sera obtained from soy-sensitive adults, indicating that putative allergenic epitopes could be present on this molecule. What is not clear is if this binding arises from specific or nonspecific interactions, including the presence of hidden epitopes. Hidden epitopes appear to be present in many soybean seed proteins and exhibit cross-reactivity with IgE elicited in patients to a related food source, for example, people with allergies to peanut (*12, 20, 26, 27*). The lack of discrimination in IgE binding to P39 and its fusion fragments (**Figure 8**) precluded a detailed dissection of epitopes; it is also possible that observed IgE binding was artifactual. These data suggest caution in interpreting the actual allergenic potential of this protein. The likelihood of P39 and P39 homologues acting as allergens has not been directly evaluated, but our data suggest that it could be a potential allergen, especially in lecithin.

Enrichment of P39 in soybean oil bodies appears to be an artifact generated during the disruption of soybean seeds. This property is similar to that of another soybean allergen, Gly m Bd 30K (*28*), although it was initially identified as an oil body

associated protein (*17*) and named not as P30 but as P34, but later proved to be localized in protein bodies (*28, 29*). However, unlike Gly m Bd 30K, the gold labeling of P39 protein overall is not very dense, confirming that this protein is of low abundance, but chiefly present in protein bodies.

In conclusion, the full sequences of two highly similar soybean 39 kDa proteins, P39-1 and P39-2, were documented, and it was shown that P39-1 and its homologues from other plants constitute a unique protein family. In soybeans, P39-1 and P39-2 are seed specific and are transcribed early in seed development. Gene expression but not protein production appears to be down-regulated in more mature seeds. Both P39-1 and P39-2 proteins are accumulated in low levels and are extracted in oil-body-enriched fractions. Ultrastructural immunohistochemistry demonstrated P39 localization predominantly in the matrix of the protein bodies. IgE from several soybean-sensitive individuals efficiently recognized recombinant polypeptide fragments derived from the P39-1 sequence, suggesting that this protein may possess several allergenic epitopes.

ACKNOWLEDGMENT

We thank Uyen Chu for technical assistance.

LITERATURE CITED

(1) Suzuhaj, B. F. Lecithin production and utilization. *J. Am. Oil. Chem. Soc.* **1983**, *60*, 306–309.
(2) Porras, O.; Carlsson, B.; Fallstrom, S. P.; Hanson, L. A. Detection of soy protein in soy lecithin, margarine and, occasionally, soy oil. *Int. Arch. Allergy Appl. Immunol.* **1985**, *78*, 30–32.
(3) Gu, X.; Beardslee, T.; Zeece, M.; Sarath, G.; Markwell, J. Identification of IgE-binding proteins in soy lecithin. *Int. Arch. Allergy Immunol.* **2001**, *126*, 218–225.
(4) Awazuhara, H.; Kawai, H.; Baba, M.; Matsui, T.; Komiyama, A. Antigenicity of the proteins in soy lecithin and soy oil in soybean allergy. *Clin. Exp. Allergy* **1998**, *28*, 1559–1564.
(5) Muller, U.; Weber, W.; Hoffmann, A.; Franke, S.; Lange, R.; Vieths, S. Commercial soybean lecithins: asource of hidden allergens. *Z. Lebensm. Unters. Forsch. A* **1998**, *207*, 341–351.
(6) Fine, S. R. Possible reactions to soya lecithin in aerosols. *J. Allergy Clin. Immunol.* **1991**, *87*, 600.
(7) Lavaud, F.; Perdu, D.; Prevost, A.; Vallerand, H.; Cossart, C.; Passemard, F. Baker's asthma related to soybean Lecithin exposure. *Allergy* **1994**, *49*, 159–162.
(8) Renaud, C.; Cardiet, C.; Dupont, C. Allergy to soy lecithin in child. *J.Pediatr. Gastroenterol. Nutr.* **1996**, *22*, 328–329.
(9) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
(10) Pearson, W. R. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **1990**, *183*, 63–98.
(11) Felsenstein, J. PHYLIP (Phylogeny Inference Package), version 3.5c, distributed by the author, Department of Genetics, University of Washington, Seattle, WA, 1993.
(12) Xiang, P.; Beardslee, T. A.; Zeece, M. G.; Markwell, J.; Sarath, G. Identification and analysis of a conserved immunoglobulin E-binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. *Arch. Biochem. Biophys.* **2002**, *408*, 51–57.
(13) Tzen, J. T.; Peng, C. C.; Cheng, D. J.; Chen, E. C.; Chiu, J. M. A new method for seed oil body purification and examination of oil body integrity following germination. *J. Biochem. (Tokyo)* **1997**, *121*, 762–768.
(14) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–685.
(15) Herman, E. M. Electron microscopic immunogold localization. In *Methods in Plant Electron Microscopy and Cytochemistry*; William, D., Ed.;Humana Press: Totowa, NJ, 2000.
(16) Xiang, P.; Haas, E. J.; Zeece, M. G.; Markwell, J.; Sarath, G. C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen. *Planta* **2004**, *220*, 56–63.
(17) Herman, E. M. Immunogold-localization and synthesis of an oil-body membrane protein in developing soybean seeds. *Planta* **1987**, *172*, 336–345.
(18) Cordle, C. T. Soy protein allergy: incidence and relative severity. *J. Nutr.* **2004**, *134*, 1213S–1219S.
(19) Koppelman, S. J.; Lakemond, C. M.; Vlooswijk, R.; Hefle, S. L. Detection of soy proteins in processed foods:literature overview and new experimental work. *J. AOAC. Int* **2004**, *87*, 1398–407.
(20) L'Hocine, L.; Boye, J. I. Allergenicity of soybean: new developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Crit. Rev. Food Sci. Nutr.* **2007**, *47*, 127–143.
(21) Batista, R.; Martins, I.; Jeno, P.; Ricardo, C. P.; Oliveira, M. M. A proteomic study to identify soya allergens − the human response to transgenic versus non-transgenic soya samples. *Int. Arch. Allergy Immunol.* **2007**, *144*, 29–38.
(22) Breiteneder, H.; Mills, C.E. N. Plant food allergens − structural and functional aspects of allergenicity. *Biotechnol. Adv.* **2005**, *23*, 395–399.
(23) Mills, E. N.; Jenkins, J.; Marigheto, N.; Belton, P.S.; Gunning, A.P.; Morris, V. J. Allergens of the cupin superfamily. *Biochem. Soc. Trans.* **2002**, *30*, 925–929.
(24) Roychaudhuri, R.; Sarath, G.; Zeece, M.; Markwell, J. Reversible denaturation of the soybean Kunitz trypsin inhibitor. *Arch. Biochem. Biophys.* **2003**, *412*, 20–26.
(25) Roychaudhuri, R.; Sarath, G.; Zeece, M.; Markwell, J. Stability of the allergenic soybean Kunitz trypsin inhibitor. *Biochim. Biophys. Acta* **2004**, *1699*, 207–212.
(26) Beardslee, T. A.; Zeece, M. G.; Sarath, G.; Markwell, J. P. Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *Int. Arch. Allergy Immunol.* **2000**, *123*, 299–307.
(27) Anibarro, B.; Seoane, F. J.; Mugica, M. V. Involvement of hidden allergens in food allergic reactions. *J. Invest. Allergol. Clin. Immunol.* **2007**, *17*, 168–172.
(28) Ogawa, T.; Tsuji, H.; Bando, N.; Kitamura, K.; Zhu, Y. L.; Hirano, H.; Nishikawa, K. Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci., Biotechnol., Biochem.* **1993**, *57*, 1030–1033.
(29) Kalinski, A.; Melroy, D. L.; Dwivedi, R. S.; Herman, E. M. A soybean vacuolar protein (P34) related to thiol proteases is synthesized as a glycoprotein precursor during seedmaturation. *J. Biol. Chem.* **1992**, *267*, 12068–12076.

Received for review November 9, 2007. Revised manuscript received December 18, 2007. Accepted December 19, 2007. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

JF073292X