

Analysis of the Distribution of the Major Soybean Seed Allergens in a Core Collection of *Glycine max* Accessions

R. W. Yaklich,* R. M. Helm, G. Cockrell, and E. M. Herman

ABSTRACT

P34 is an outlying member of the papain-superfamily of cysteine proteases that is expressed in developing soybean [*Glycine max* (L.) Merr.] seeds and may be involved in the defense against *Pseudomonas* infection. P34 is the major human allergen of soybean seeds and is present in processed food products that contain soybean protein. We have surveyed a core collection of soybean accessions using a monoclonal antibody against P34 and with human sera from soybean-sensitive individuals. We found that the accessions of soybean surveyed contain similar levels of P34 and that P34 is the major allergenic protein. We have also surveyed wild relatives of soybean and found that P34 was present in the examples assayed. These results indicate that it may not be possible to eliminate P34 from the food supply by breeding with an improved germplasm base.

AN OUTLYING MEMBER of the papain-superfamily of cysteine-proteases has been identified and cloned from maturing soybean seeds (Kalinski et al., 1990, 1992). This protein, termed P34, possesses most of the conserved characteristics of cysteine proteases including a large precursor domain that is posttranslationally processed. The primary sequence contains aligned and conserved amino acids that are critical in the conserved tertiary conformation of members of the papain superfamily. P34 exhibits some unique features that separate it from other members of the papain superfamily. Among these are replacement of the conserved cysteine in the active site found in all other papain family proteins with a glycine, suggesting that the protein is enzymatically inactive (Kalinski et al., 1990, 1992). Cysteine proteases are typically self-processed under acid-reducing conditions resulting in the cleavage of the large precursor domain. However, P34 is processed after an asparagine residue in a single step (Kalinski et al., 1992), most likely by the same enzyme that processes the 11S storage proteins (Hara-Nishimura et al., 1993, 1995). Sequence comparisons and alignments indicate that although P34 is a member of the papain superfamily, it is also quite dissimilar from the enzymatically active cysteine proteases including those identified in soybean.

P34 may have a function in defense against *Pseudomonas* infection by binding syringolide elicitors secreted by the bacteria (Cheng et al., 1998). P34 is very abundant in seeds (Kalinski et al., 1992), but it is also found in vegetative cells that are subject to bacterial infection (Cheng et al., 1998). Soybean cultivars that are resistant

to *Pseudomonas* possess higher levels of P34 in leaf tissue compared with sensitive cultivars (Cheng et al., 1998). P34 has been shown to be the major allergenic protein of soybean seeds (Ogawa et al., 1991, 1993; Burks et al., 1988). Assays of IgE binding using immunoglobulins from soybean sensitive individuals indicates that 65% of the total allergenic response can be accounted for by P34 (Ogawa et al., 1991, 1993; Helm et al., 1998). Detailed immunological analysis of the allergenicity of P34 by epitope mapping has shown that there are at least 12 distinct epitopes on the protein (Helm et al., 1998).

The use of soybean products in processed foods poses a significant health and food safety problem for sensitive individuals since the P34 protein can be detected in a wide variety of processed foods (Tsuji et al., 1995). The elimination of P34 from soybean seeds would enhance food safety and make the use of soybean products available to sensitive individuals. Therefore, using a monoclonal antibody against P34 and sera from soybean sensitive individuals, we surveyed a core collection of soybean accessions to examine the abundance of P34 and other allergenic proteins among these cultivars. In parallel, we assayed other members of the *Glycine* species for the presence of P34 and other sera allergenic proteins.

MATERIALS AND METHODS

Seed samples of *Glycine* species were procured from the National Germplasm Repositories or grown at Beltsville in 1996 by normal agronomic procedures (Yaklich et al., 1979). The *G. max* samples consisted of the core collection of soybean ancestors and first progeny that contributed at least 95% of the genes found in public cultivars released between 1947 and 1988 in North America (Gizlice et al., 1994), and 'Black Eyebrow', 'Blackhawk', 'Minsoy', 'Williams', 'Miles', 'Essex', and 'Kunitz'. Seed samples of nodulated and non-nodulated lines of 'Clark', BARC-14, BARC-15, BARC-16, BARC-17, and the high protein lines CX797-21, CX797-115, CX804-3, CX804-108, (J.R. Wilcox, USDA-ARS, Purdue University), 76-48773 (A. Matson, Soybean Research Foundation), D76-8070, D80-6931, D81-8259, D81-8498, (E.E. Hartwig, USDA-ARS, Jamie Whitten Delta States Research Center), NC-2-62, (J.W. Burton, USDA-ARS, North Carolina State University), BARC-6, BARC-7, BARC-8, and BARC-9, (R.C. Leffel, USDA-ARS, Beltsville, Maryland). In addition, plant introductions of the following *Glycine* species were included in the study: *G. soja* Siebold & Zucc., PI 65549, PI 81762, PI 101404; *G. arenaria* Tind., PI 505204; *G. argyrea* Tind., PI 505151, PI 509451; *G. canescens* F.J. Herm, PI 440932, PI 440942, PI 446934; *G. clandestina* Wendl., PI 440948, PI 440954, PI 440960; *G. curvata* Tind., PI 505164, PI 505166, PI 505167; *G. cyrtoloba* Tind., PI 440962, PI 440963, PI 509472; *G. falcata* Benth., PI 440975, PI 505179, PI 509473; *G. latifolia* (Benth.) Newell & Hymowitz, PI 378709, PI 253238, PI 440978; *G. microphylla* (Benth.) Tind., PI 339664, PI 440956, PI 446939; *G. tabacina* (Labill.) Benth., PI 339661, PI 343986, PI 373990; and *G. tomentella* Hayata, PI 373987, PI 440998, PI 441000, PI 446993, PI 446995, and PI 509499.

R.W. Yaklich, Soybean and Alfalfa Res. Lab.; E.M. Herman, Climate Stress Lab., USDA-ARS, Beltsville Agric. Res. Center, Beltsville, MD 20705; R.M. Helm and G. Cockrell, Dep. of Pediatrics, Univ. of Arkansas for Medical Sciences, Arkansas Children's Hospital Research Institute, Little Rock, AR 72202. Mention of proprietary products are included for the benefit of the reader and do not imply endorsement by the USDA. Received 29 Sept. 1998. *Corresponding author (ryaklich@asrr.usda.gov).

Seeds from each line were ground in a Wiley mill and passed through a 30-mesh screen. The seed powder was weighed in an Eppendorf tube and 41 μ L of extraction buffer [0.17 M sodium dodecyl sulfate, 6.0 M urea, 0.57 mM 2-mercaptoethanol, and 0.05 M tris(hydroxy methyl) aminomethane, pH 6.8] was added for each milligram of seed powder. The seed proteins were extracted by treatment for 45 min in an ultrasonic water bath. The samples were centrifuged and 8 μ L of the supernatant of *G. max* and *G. soja* and 15 μ L of the other *Glycine* spp. were added to each well of the polyacrylamide gel.

Electrophoresis SDS-PAGE was carried out according to the procedures of Laemmli (1970) in 1.5-mm thick mini-gels with a 12.5% (w/v) separating gel and 4% (w/v) stacking gel. The gels contained 15 wells and an outer well received the same volume of molecular weight standard. The amount of seed extract loaded to each well were equivalent with respect to weight of the seed. Electrophoresis was performed in a buffer solution [0.19 M glycine, 0.1% (w/v) sodium dodecyl sulfate, 0.025 M tris(hydroxy methyl aminomethane, pH 8.3] for 2 to 3 h to completion. The fractionated proteins were electroblotted to Immobilon-P transfer membranes (Millipore Corp., Bedford, MA) with a 10 mM Tris-sodium borate 20% (v/v) methanol, pH 8.2 transfer buffer (Melroy and Herman, 1991). Blots were stained with amido-black to visualize proteins or were immunolabeled. For immunolabeling, air-dried blots were wetted in methanol and washed several times in transfer buffer; washed in TBS [Tris-buffered saline, 100 mM tris(hydroxy methyl) aminomethane, 150 mM NaCl, pH 7.5] and finally washed and blocked in TBS containing 5% (w/v) powdered milk for 3 h. Monoclonal antibodies (designated P4B5) and ascites expansion of the cell lines secreting antibodies elicited against P34 were previously described in Herman et al., (1990) and Kalinski et al., (1992). For immunolabeling, a 1:5000 dilution of the clarified ascites fluid was added to the blots in TBS containing 5% powdered milk overnight and washed in TBS. Immunoreactive bands were visualized by labeling the blots in TBS containing 5% powdered milk with a 1:5000 dilution of anti-mouse immunoglobulin-G (IgG) alkaline phosphatase and development with chromogenic substances 5-bromo-4-chloro-3-indolylphosphate and nitroblue tetrazolium. For the serum IgE antibody, patients with a convincing clinical history of soybean allergy having atopic dermatitis, a positive immediate skin test response to soybean extracts or confirmed to be sensitive to soybean protein by double-blind, placebo-controlled food challenge were used as sources of soybean-specific IgE (Helm et al., 1998). Equal volumes of sera and or plasma (PlasmaLab International, Everett, WA) from these individuals were combined to prepare a serum pool containing soybean-specific IgE antibody. The serum pool was diluted 1:20 (v/v) in Tris-buffered saline (25 mM Tris-HCl, pH 7.2, 150 mM NaCl) with 0.05% Tween 20 (TBST) and added to the blots overnight at 4°C, washed 3 times in 20 mL TBST, and radiolabeled with anti-IgE (Sanofi Pasteur Diagnostics, Chaska, MN) diluted 1:5 (v/v) in TBST for 2 to 4 h. Blots were then washed 3 times in TBST, air dried and exposed to X-Omat x-ray film at -70°C for 18 to 72 h and developed in an automatic developer (Konoca SRX-10). High titer IgE antibodies are specifically produced in people with allergies. In order to test the specificity of the soybean IgE antibody, control blots were incubated with a high titer IgE serum (control serum, 1:10 v/v) derived from people who have allergies (for example, to pollen, dust, etc.) but not to soybean or other food proteins. The control serum was treated in a similar manner as the food allergy serum with the exception that the blots were exposed to the X-Omat x-ray film for 5 d.

RESULTS

To examine the distribution of P34, the main soybean seed storage protein allergen, we surveyed the core collection representing at least 95% of the genes found in public cultivars released between 1947 and 1988 in North America. The representative samples of three replicate blots presented in Fig. 1 shows that all contained the P34 protein (Panel B) and likewise, all of the samples of *G. max* that we surveyed contained P34 (data not presented). Panel C shows that the pooled human sera reacted mainly with P34 and this also was observed in all the *G. max* samples that we assayed. There were visual differences in intensity of the reaction to the P34 and IgE antibodies that may indicate differences in tissue concentration of the allergen. The samples of *G. soja* that we analyzed gave results similar to those of *G. max* (data not presented).

Figure 2 is a representative sample of three replicate blots containing some of the high protein soybean lines (Lanes 1–11). Normal protein lines contain approximately 410 g kg⁻¹ protein whereas the high protein lines represented have a range in protein of 480 to 550 g kg⁻¹. Some of the staining of the P34 bands for the high protein lines in panel B appear to be less intense than the three named cultivars in Lanes 12 through 14. P34 is the main human allergen observed in the cultivars assayed in Panel C with some other reactive bands to the human sera that were also observed in Panel C of Fig. 1. This indicates that selecting for enhanced protein content in the high protein lines did not result in the

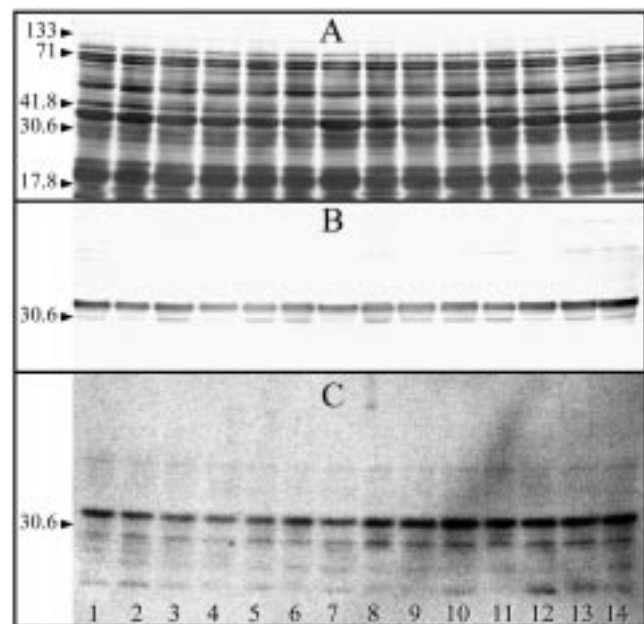


Fig. 1. Immunoblot analysis of the seed extracts of soybean, *Glycine max*. Upper panel (A), amido-black stained blot of the total proteins. Middle panel (B), replicate blot that was treated with anti-P34 serum. Lower panel (C), replicate blot that was treated with human serum. Arrowheads (far left) are the molecular weight rainbow standards. Lanes 1–14 represent the lines 'AK Harrow', 'Arksoy', 'Bansei', Black Eyebrow, Blackhawk, 'Capital', 'CNS', 'Dunfield', 'Flambeau', 'Haberlandt', 'Illini', 'Improved Pelican', 'Jackson', and 'Jogun'.

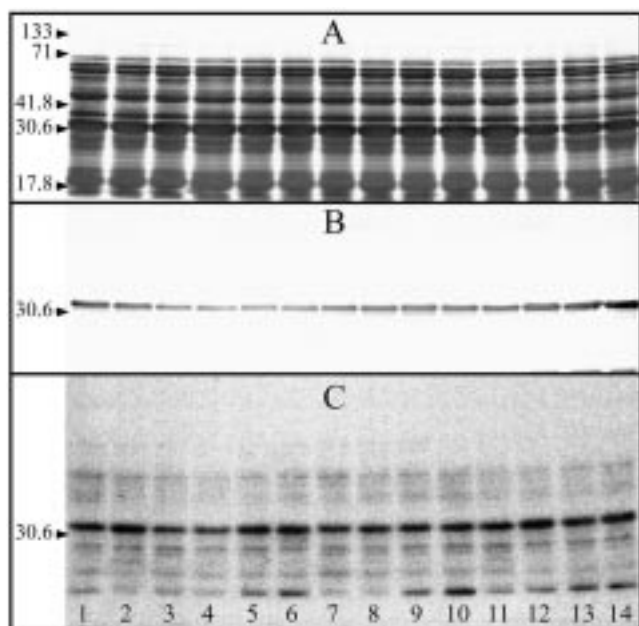


Fig. 2. Immunoblot analysis of the seed extracts of soybean, *Glycine max*. Upper panel (A), amido-black stained blot of the total proteins. Middle panel (B), replicate blot that was treated with anti-P34 serum. Lower panel (C), replicate blot that was treated with human serum. Arrowheads (far left) are the molecular weight rainbow standards. Lanes 1–14 represent the lines D81-8359, CX804-108, NC-2-62, CX797-115, CX797-21, 76-48773, D81-8498, BARC-6, BARC-7, BARC-8, BARC-9, Williams, Miles, and Clark, respectively.

parallel selection of enhanced levels and/or additional human allergens.

Representative samples of three replicate blots of the *Glycine* spp. (Fig. 3) show that they all reacted to the anti-P34 serum (Panel B) indicating that P34 is highly conserved in the *Glycine* family. P34 also was present in the rest of the accessions tested (data not presented). However, unlike *G. max* and *G. soja*, the rest of the *Glycine* spp. that we tested have multiple, highly reactive protein bands with the IgE antibody (Panel C) that may be the acidic and basic polypeptides of glycinin and the subunits of *B*-conglycinin. The intensity and number of bands differed among the accessions that were assayed. This was observed when comparing the two different accessions of *G. argyrea* (Lanes 2 and 3), where PI 505151 has a high molecular weight band not present in PI 509451. Differences in band number and/or staining intensity were also evident in accessions of *G. canescens*, *G. clandestina*, and *G. cyrtoloba*. The three accessions of *G. curvata* appear similar in band number and staining intensity. We also found differences in band number and/or staining intensity to the IgE antibody between accessions in *G. falcata*, *G. microphylla*, *G. tabacina*, and *G. tomentella* (data not presented). Control blots treated with the high titer IgE serum from people who are not allergic to soybeans or other food proteins revealed no labeled protein bands.

DISCUSSION

Food allergies are a wide-spread problem that includes most of the major plant and animal foods (Ander-

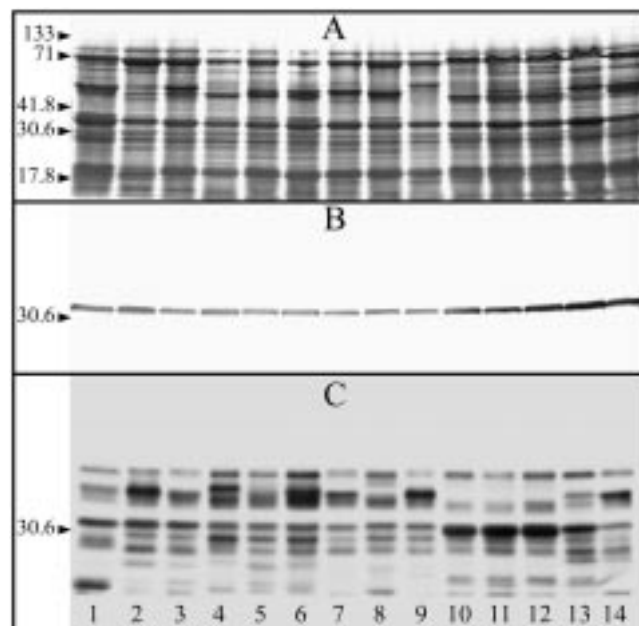


Fig. 3. Immunoblot analysis of the seed extracts of *Glycine* species. Upper panel (A), amido-black stained blot of the total proteins. Middle panel (B), replicate blot that was treated with anti-P34 serum. Lower panel (C), replicate blot that was treated with human serum. Arrowheads (far left) are the molecular weight rainbow standards. Lanes 1–14 represent *G. arenaria* (PI 505204), *G. argyrea* (PI 505151, PI 509451), *G. canescens* (PI 440932, PI 440942, PI 446934), *G. clandestina* (PI 440948, PI 440954, PI 440960), *G. curvata* (PI 505164, PI 505166, PI 505167), and *G. cyrtoloba* (PI 440962, PI 440963), respectively.

son and Sogn, 1984; Bernhisel-Broadbent and Sampson, 1989; Herian et al., 1990). Soybean protein allergies pose a significant problem for large numbers of people because soybean protein is now a common constituent of many processed foods. For sensitive individuals, avoiding soybean products is difficult and poses significant limitations in choosing processed and convenience foods. The repression of P34 accumulation in the soybean seed would result in elimination of the large majority of allergenicity of soybean seeds.

Our results indicate that P34 is found in soybean and the other *Glycine* spp. assayed. The results also indicate that P34 would be present in all of the cultivars released in the public domain. *G. soja*, which is cross compatible, producing fertile F_1 hybrids with *G. max* and its closest ancestor, gave results similar to *G. max*. The similarity of staining between the P34 blots of all of the *G. max* samples indicated that the high protein lines did not contain more P34 or any other allergens than the other *G. max* lines. We know that the high protein lines contain more glycinin and *B*-conglycinin than normal soybean lines (unpublished data) and when compared with P34, that the acidic and basic polypeptides of glycinin and the subunits of *B*-conglycinin did not produce strong allergenic protein bands from the pooled human sera. This indicates that P34 was found at similar levels in all the *G. max* samples, and that the specific major storage proteins which are increased in high protein lines are no more allergenic than those found in normal protein lines.

The *Glycine* spp. that we tested contained the P34 allergen and many other allergenic proteins. Glycinin and *B*-conglycinin have been shown to have some allergenic properties (Shibasaki et al., 1980). Some of the observed allergenic bands of the *Glycine* spp. to the human sera have relative mobilities similar to the acidic and basic polypeptides of glycinin and the subunits of *B*-conglycinin. Research by Staswick et al. (1983) on the 11S fraction (glycinin) from *G. tomentella* revealed that a group of proteins were present that resembled the A_{1b}-polypeptide components of glycinin from *G. max*. They were of similar size, had identical NH₂-terminal sequences, and similar amino acid composition to A_{1b}. The rest of the 11S fraction had other similarities to glycinin of *G. max*. Our results indicate that the major storage proteins in the *Glycine* spp. that are similar to those found in *G. max*, may be as allergenic as P34. This indicates that there are differences in the major storage proteins in *G. max* and *G. soja* that have amino acid sequences that do not elicit an adverse immunological response. The results with the wild soybean relatives are particularly interesting and indicate that major storage proteins of other legumes can be potent allergens. Peanut (*Arachis hypogaea* L.) storage proteins can induce severe and lethal reactions in sensitive individuals and have been extensively analyzed (Burks et al., 1996). The cross-reactivity of the human sera with some wild soybean storage proteins compared with the lack of response in domestic and other wild soybeans suggests that a comparison of glycinin between the *Glycine* spp. including *G. max* and *G. soja* could prove a useful model to further analyze the allergenicity of legume storage proteins.

Our results show that P34 is the major allergen in *G. max* and also is a major allergen in the other *Glycine* spp. These results indicate that the elimination of P34 through selective breeding is likely to be difficult. P34 appears to be encoded by only a few genes (Kalinski et al., 1990), so the prospects of eliminating the expression of P34 through either induced mutation or by cosuppression would appear to be possible. Experiments are currently in progress to suppress P34 synthesis in soybean seeds using biotechnological approaches. Such an approach offers the prospect of directly introducing suppression into elite germplasm through backcrossing.

ACKNOWLEDGMENTS

We thank Drs. R. Nelson and T. Hymowitz for selection and samples of germplasm used in this study. Mr. Christopher Pooley prepared the figures. This work was supported in part by the Arkansas Children's Nutrition Center (USDA-ARS 0524-0022).

REFERENCES

- Anderson, J.A., and D.D. Sogn. 1984. Adverse reactions to food. National Institutes of Health, publication No. 84-2442. Bethesda, MD.

- Bernhisel-Broadbent, J., and H.A. Sampson. 1989. Cross allergenicity in the legume botanical family in children with food hypersensitivity. *J. Allergy Clin. Immunol.* 83:435-440.
- Burks, A.W., J.R. Brooks, and H.A. Sampson. 1988. Allergenicity of major component proteins of soybean determined by enzyme linked immunosorbent assay (ELISA) and immunoblotting in children with atopic dermatitis and positive soy challenges. *J. Allergy Clin. Immunol.* 81:1135-1142.
- Burks, A.W., D. Shin, G. Cockrell, J.S. Stanley, R.M. Helm, and G.A. Bannon. 1996. 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.* 245:334-339.
- Cheng, J.I., C. Boyd, D. Slaymaker, Y. Okinaka, E.M. Herman, and N.T. Keen. 1998. Purification and characterization of a 34 kDa syringolide binding protein from soybean. *Proc. Natl. Acad. Sci. (USA)* 95:3306-3311.
- Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* 34:1143-1151.
- Hara-Nishimura, I., T. Shimada, N. Hiraiwa, and M. Nishimura. 1995. Vacuolar processing enzyme responsible for maturation of seed proteins. *J. Plant Physiol.* 145:632-640.
- Hara-Nishimura, I., Y. Takeuchi, and M. Nishimura. 1993. Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Shistosoma mansoni*. *Plant Cell* 5:1651-1659.
- Helm, R.M., G. Cockrell, E. Herman, A.W. Burks, H.A. Sampson, and G.A. Bannon. 1998. Cellular and molecular characterization of a major soybean allergen. *Int. Arch. Allergy Immunol.* 117:29-37.
- Herian, A.M., S.L. Taylor, and R.K. Bush. 1990. Identification of soybean allergens by immunoblotting with sera from soy-allergic adults. *Int. Arch. Allergy Appl. Immunol.* 92:193-198.
- Herman, E.M., D.L. Melroy, and T.J. Buckhout. 1990. Apparent processing of a soybean oil body membrane protein accompanies the onset of oil mobilization. *Plant Physiol.* 94:341-349.
- Kalinski, A., J.M. Weisemann, B.F. Matthews, and E.M. Herman. 1990. Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. *J. Biol. Chem.* 265:13843-13848.
- Kalinski, A., D.L. Melroy, R.S. Dwivedi, and E.M. Herman. 1992. A soybean vacuolar protein (P34) related to thiol proteases is synthesized as a glycoprotein precursor during seed maturation. *J. Biol. Chem.* 267:12068-12076.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685.
- Melroy, D.L., and E.M. Herman. 1991. TIP, an integral membrane protein of the protein-storage vacuoles of the soybean cotyledon undergoes developmentally regulated membrane accumulation and removal. *Planta* 184:113-122.
- Ogawa, T., N. Bando, H. Tsuji, H. Okajima, K. Nishikawa, and K. Sasoka. 1991. Investigations of the IgE-Binding proteins in soybeans by immunoblotting with the sera of soybean-sensitive patients with atopic dermatitis. *J. Nutri. Sci. Vitamino* 37:555-565.
- Ogawa, T., H. Tsuji, N. Bando, K. Kitamura, Y.L. Zhu, H. Hirano, and K. Nishikawa. 1993. Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci. Biotech. Biochem.* 57:1030-1033.
- Shibasaki, M., S. Suzuki, S. Tajima, H. Nemoto, and T. Kuroume. 1980. Allergenicity of major component proteins of soybean. *Int. Arch. Allergy App. Immun.* 61:441-448.
- Staswick, P.E., P. Broue, and N.C. Nielsen. 1983. Glycinin composition of several perennial species related to soybean. *Plant Physiol.* 72:1114-1118.
- Tsuji, H.N. Okada, R. Yamanishi, N. Bando, M. Kimoto, and T. Ogawa. 1995. Measurement of Gly m Bd 30K, a major soybean allergen, in soybean products by a sandwich enzyme-linked immunosorbent assay. *Biosci. Biotech. Biochem.* 59:150-151.
- Yaklich, R.W., M.M. Kulik, and C.S. Garrison. 1979. Evaluation of vigor in soybean seeds: Influence of date of planting and soil type on emergence, stand, and yield. *Crop Sci.* 19:242-246.