A Novel Methionine-Rich Protein in Soybean Seed: Identification, Amino Acid Composition, and N-Terminal Sequence

Aswathi A. George and Benito O. de Lumen*

Department of Nutritional Sciences, University of California, Berkeley, California 94720

Using the powerful resolution obtainable with two-dimensional gel electrophoresis and an in vitro labeling method we developed for detecting methionine-containing proteins, we identified a new methionine-rich protein (MRP) in soybean seed, confirmed its methionine content by amino acid microanalysis, and sequenced its N terminus without using the usual protein purification procedures. The complete amino acid composition and N-terminal sequence of the first 20 amino acids are presented. The MRP has a methionine content of 12.1%, constitutes 0.6% of the total protein, and has a molecular weight of 10 800. This value is exceptionally high considering that the average overall methionine content of soybean seed protein is 1.4%. This is the first report of an exceptionally methionine-rich protein in soybean seed that is distinct from the methionine–cysteine protein we reported earlier and from the cysteine-rich proteins reported previously by other workers. The MRP presents a genetic engineering strategy to improve the nutritional quality of soybean protein and raises the question of its possible biological role in the seed.

Legumes are the richest sources of proteins among plant foods, but are deficient in the sulfur amino acids methionine and cysteine. Soybean (*Glycine max*) is the single largest source of protein in animal feed worldwide, and like most legumes, its protein is first limiting in the sulfur amino acids. The methionine level in soybean ranges from 1.1 to 1.6% of total protein, most samples being in the 1.2–1.4% range. Expressed as a percentage of total protein, the soybean protein contains only 56% as much methionine and cysteine as the nutritionally complete Food and Agriculture Organization (FAO) reference protein, egg protein (Kelley, 1973).

Although this problem can be overcome by supplementation with free methionine and complementation with cereals, there are good reasons to improve the amino acid profile of legumes. Leaching of methionine during processing and bacterial degradation leading to formation of undesirable volatile sulfides are associated with supplementation. The poultry and swine feed industry would derive considerable savings if an improved variety of soybean did away with the requirement for supplementation of corn/soy-based animal feed with exogenous methionine. Certain population groups whose main source of carbohydrates is root crops would benefit from legumes with complete protein, since the amino acid profile of root crops is not complementary with that of legumes.

The genes encoding seed storage proteins are under strict developmental control and are expressed in a tissue-specific manner (Goldberg et al., 1989). The expression of storage protein genes directly affects the nutritional quality of the seed protein. One of the possible approaches to genetically engineering the nutritional quality of soybean seed protein would be to identify endogenous methionine-rich proteins and increase their biosynthesis.

The sulfur-rich plant proteins reported so far can be classified as cysteine rich and methionine rich. The cysteine-rich group includes the Bowman–Birk protease inhibitors (Hwang et al., 1977; Odani and Ikenaka, 1977, 1978; Joudrier et al., 1987), the wheat endosperm purothionein (Mak and Jones, 1976), the pea low molecular weight albumins (Higgins et al., 1986), and a soybean low molecular weight albumin (Kho and de Lumen, 1988). These proteins contain very little or no methionine. Methionine-rich proteins (MRP) have been reported in Brazil nut (Altenbach et al., 1987; Sun et al., 1987), sunflower seed (Lilley et al., 1989), corn (Pedersen et al., 1986; Kirihara et al., 1988), and rice (Musumura et al., 1989).

Combining the powerful resolution obtainable with two-dimensional gel electrophoresis and an in vitro labeling method (de Lumen and Kho, 1987) we developed for detecting methionine-containing proteins, we were able to identify a soybean MRP directly without employing commonly used protein purification techniques. Additionally, electroblotting the proteins from the two-dimensional gel onto poly(vinylidene difluoride) (PVDF) membranes allowed amino acid analysis and N-terminal sequencing of the protein directly from the membrane. Considering that the average methionine content of soybean is 1.4%, this is the first report of an exceptionally methionine-rich protein in soybean seed. The 10.8-kDa MRP has a methionine content twice that of a 16.0-kDa methionine-cysteine rich protein we identified earlier in soybean seed (Kho and de Lumen, 1988) and is distinct from cysteine-rich proteins reported previously by other workers. We report here the identification of a MRP in soybean seed, its amino acid composition, and its N-terminal sequence as a first step toward cloning its gene, with a view of ultimately increasing its biosynthesis in the seed and determining its biological role, if any.

MATERIALS AND METHODS

Materials. Soybean plants (*G. max* cv. Hodgson 78) were grown in the greenhouse, and seeds were collected at maturity (60 days after flowering).

Ampholytes were obtained from Serva (Westbury, NY), urea was from Fisher Scientific (Fair Lawn, NJ), and all the electrophoresis reagents were from Bio-Rad (Richmond, CA). Iodo-[1-^{14}C]acetic acid was from ICN Radiochemicals (Irvine, CA), and PVDF membranes (Immobilon) were from Millipore (Bedford, MA).

Protein Extraction. The seeds were harvested and frozen at −80 °C until required. Following removal of the seed coat, 1.0 g of the seeds was ground in a homogenizer with 10.0 mL of 0.1 M sodium phosphate buffer (pH 7.5) containing 0.5 M NaCl, 0.001 M phenylmethanesulfonyl fluoride, and 0.05 M β-mercap-

* Author to whom correspondence should be addressed.

0021-8561/91/1439-0224$02.50/0

toethanol for 3 min. The extract was centrifuged at 12000*g* for 20 min at 4 °C by using a SS-34 (Sorvall) rotor. The supernatant was saved, and the extraction was repeated once more. The combined supernatants were filtered through eight layers of cheesecloth to obtain the total protein extract and frozen in aliquots at −20 °C.

Two-Dimensional Electrophoresis. The total proteins were separated in the first dimension by isoelectric focusing (IEF), essentially as described by O'Farrell (1975). The total concentration of ampholytes used in the IEF gels was 2%, of which 1.8% consisted of ampholytes 5–7 and 0.2% of ampholytes 3–10. Tube gels were poured to a height of 12.0 cm, overlayed with water, and allowed to polymerize for 2 or more hours. Samples were prepared for loading by mixing an equal volume of the protein extract with the lysis buffer. Four hundred micrograms of protein was loaded per tube gel. The IEF gels were not prefocused before the samples were loaded since no *N,N,N′,N′*-tetramethylethylenediamine (TEMED) was used in the polymerization mixture. One other change to the O'Farrell procedure was that 0.5% ethanolamine and 0.2% sulfuric acid were used as the upper and lower reservoir buffers, respectively, instead of 0.02 M NaOH and 0.01 M H_3PO_4. The isoelectric focusing was carried out at 400 V for 17 h, and the voltage was then increased to 800 V for another hour to focus the bands more sharply.

Separation in the second dimension was carried out by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) using an exponential gradient of 17–27% polyacrylamide.

Detection of Methionine-Containing Proteins. After electrophoresis, the proteins were electroblotted onto Immobilon by using a semidry electroblotter (Milliblot, Millipore) with a three-buffer transfer buffer system: anode buffer 1, 300 mM Tris, 10% MeOH, pH 10.4; anode buffer 2, 25 mM Tris, 10% MeOH, pH 10.4; cathode buffer, 25 mM Tris, 40 mM glycine, 10% MeOH, pH 9. Blotting was carried out at 2.5 mA/cm² of gel for 30 min.

The blot was stained with Coomassie Blue, the transferred proteins were reacted with iodo[1-^{14}C]acetic acid and autoradiographed, and the quantitation of methionine in the proteins was carried out as described in de Lumen and Kho (1987).

Amino Acid Analysis and N-Terminal Sequencing. Following two-dimensional electrophoresis, electroblotting, and staining with Coomassie Blue, the blot was washed with water for 5 min, dried, and stored at −20 °C in sealed plastic bags until needed.

Amino acid analysis was carried out by the Protein Structure Laboratory (University of California, Davis) on a Beckman 6300 analyzer after hydrolysis with 6 N HCl at 110 °C using identical spots carefully excised from two Immobilon membranes. Under these hydrolysis conditions, tryptophan was not determined, and the entries for Glu and Asp in Table I represent the sum of the acid and amide forms. Methionine and cysteine were determined as methionine sulfone and cysteic acid after performic acid oxidation.

The N-terminal sequence was obtained from identical spots carefully excised from five Immobilon membranes by using an Applied Biosystems 477A protein sequencer. This analysis was carried out at the Microchemical Facility, University of California, Berkeley.

RESULTS AND DISCUSSION

Figure 1 shows the total proteins separated by two-dimensional electrophoresis and visualized by Coomassie Blue staining. Molecular weight markers and total proteins electrophoresed in one dimension only have been included for reference. Figure 1 demonstrates the resolving power of two-dimensional electrophoresis and the heterogeneity of soybean seed proteins. On the basis of comparisons with the electrophoregrams of Lei et al. (1983), it was possible to identify, as shown in Figure 1, the several subunits of β-conglycinin, the acidic polypeptides of glycinin, lectin, and the Kunitz trypsin inhibitor. In addition, there were many polypeptide spots of unknown identity.

A comparison of the Western blot of the total proteins and its corresponding autoradiograph (Figure 2) helps detect the methionine-containing proteins. Of particular

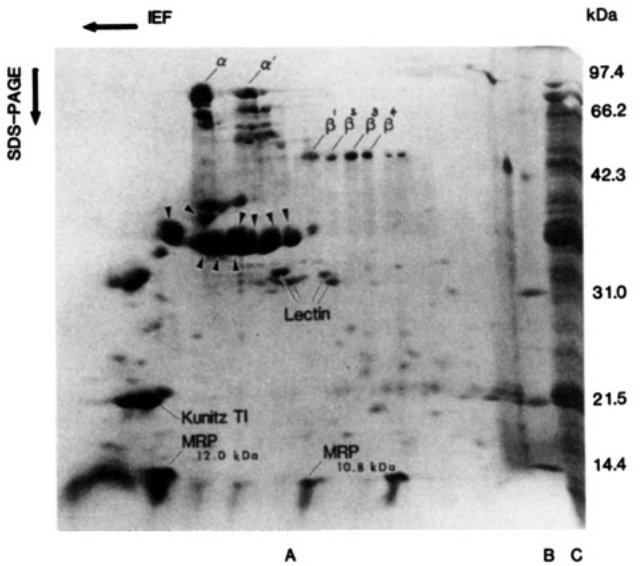

Figure 1. Two-dimensional electrophoresis of total soybean proteins. (A) The subunits of β-conglycinin are labeled α, α′, β^1, β^2, β^3, and β^4. The arrowheads point to the acidic subunits of glycinin. The four components of soybean lectin are identified, and the Kunitz trypsin inhibitor is labeled Kunitz TI. The methionine-rich proteins identified in this study are labeled MRP. (B) Molecular weight markers. (C) One-dimensional electrophoresis, SDS–PAGE of total soybean proteins.

importance are the two low molecular weight spots in the 10 000–12 000 region that show intense signals on the autoradiograph. From the densitometric scans of the Western blot and the autoradiograph, the more basic of the two spots (10.8 kDa, indicated by arrow) was estimated to contain 11.0% methionine and the second, larger spot 9.3% methionine. Whether the second spot represented two overlapping spots remains to be established. These proteins most likely belong to a family of methionine-containing proteins we have previously identified in soybean seed (Kho and de Lumen, 1988).

The complete amino acid composition of the 10.8-kDa protein is shown in Table I. Compared with defatted soymeal and the FAO reference protein, it has an exceptionally high level of methionine. The high levels of glutamic acid/glutamine and lysine are typical of soybean seed proteins (Rackis, 1961; Koshiyama, 1968). The methionine content of 12.1% confirmed the high value obtained by the in vitro labeling method and the methionine richness of this protein. This value is 2-fold higher than the 6.2% methionine for a 16.0-kDa soybean seed protein we reported earlier (Kho and de Lumen, 1988) and the highest ever reported for a soybean seed protein. It is likely that the 16.0-kDa protein we purified by differential solubilization in aqueous methanol and high-performance liquid chromatography (HPLC) in our previous work was slightly contaminated with the 10.8-kDa MRP since a faint band corresponding to about 11.0 kDa appeared together with the major 16.0-kDa band in single-dimension SDS–PAGE (Kho and de Lumen, 1988). Consequently, this could have contributed to an overestimation of the methionine content of the 16.0-kDa protein. Thus, it is very likely that the 16.0-kDa soybean protein is really a cysteine-rich protein (18.8% cysteine) with modest amounts of methionine.

The N-terminal sequence of the first 20 amino acids of the 10.8-kDa protein is shown in Figure 3. The sequence M-M-M-K in positions 15–18 is consistent with the methionine richness of the protein. An identical sequence is found in sunflower seed 2S albumin protein that contains 15% methionine (Lilley et al., 1989) and a

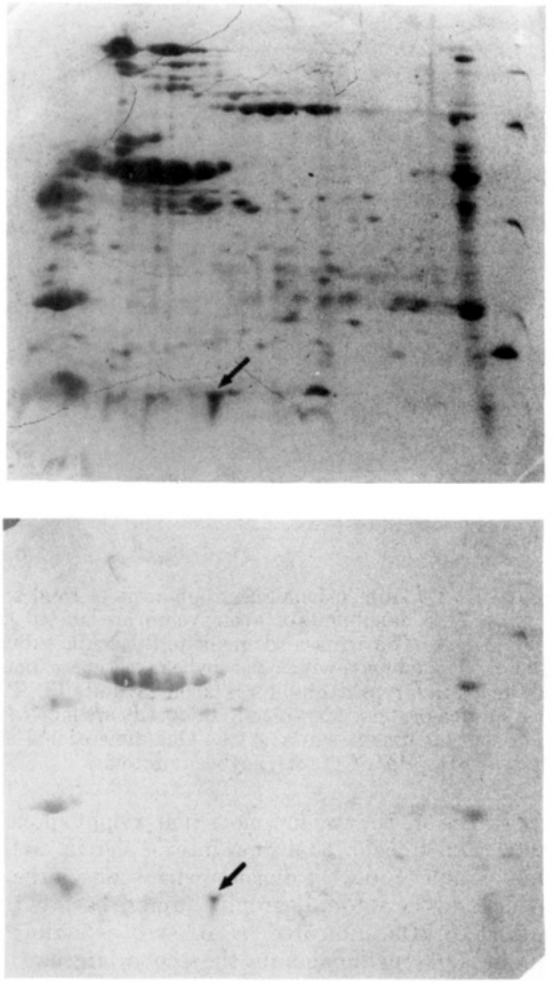

Figure 2. Identification of methionine-rich proteins. (A) Total soybean proteins separated by two-dimensional electrophoresis and electroblotted onto Immobilon. (B) Autoradiograph of the electroblotted proteins incubated with iodo[1-^{14}C]acetic acid, identifying methionine-rich proteins. The methionine-rich 10.8-kDa protein is marked with an arrow in both (A) and (B).

Table I. Amino Acid Composition of the 10.8-kDa Protein, Defatted Soymeal, and the FAO Reference Protein

	10.8-kDa protein, g/100 g of protein	defatted soymeal,[a] g/100 g of protein	FAO reference protein,[b] g/100 g of protein
Asp	7.7	10.4	
Ser	8.1	4.6	
Gly	7.3	3.4	
Val	1.7	4.6	5.0
Leu	6.6	6.7	7.0
Phe	1.0	4.5	6.0 (Phe + Tyr)
Lys	8.8	6.0	5.5
Arg	5.6	7.6	
Cys	2.5	1.3	
Thr	1.8	3.7	4.0
Glu	23.4	18.4	
Ala	3.3	3.6	
Ile	4.0	4.4	4.0
Tyr	1.6	3.5	
Trp	nd	1.2	
Pro	1.7	5.3	
His	3.0	2.2	
Met	12.1	1.4	2.2[c] 3.5 (Met + Cys)

[a] Rackis (1961). [b] FAO/WHO (1973). [c] FAO (1957). [d] nd, not determined.

homologous sequence of M-M-M-R in Brazil nut 9.0-kDa protein that has 18% methionine (Altenbach et al., 1987). It is interesting that the Brazil nut and sunflower seed

```
1                                                   10
Tyr - Arg - Lys - Lys - Glu - Gly - Gly - Glu - Glu - Glu

11                                                  20
His - Gly - Gly - Gly - Met - Met - Met - Lys - Gln - Lys
```

Figure 3. N-Terminal amino acid sequence of the 10.8-kDa methionine-rich protein.

proteins have extensive homologies with the 2S albumins from castor bean (Sharief and Li, 1982) and rapeseed (Crouch et al., 1983) which, however, contain much less methionine (2.1 and 2.6%, respectively). The reason for this is that the homology between Brazil nut protein, castor bean, and rapeseed proteins is mostly in 24 of 28 amino acids, which include all the cysteines, and the nonhomology is found in regions that contain methionine. A 10.0-kDa MRP (22.5% methionine) and a 15.0-kDa MRP (11% methionine) have been characterized in maize (Kirihara et al., 1988; Pedersen et al., 1986) as has been a 10.0-kDa MRP (20% methionine) from rice (Musumura et al., 1989). The primary structures of the maize MRP are characterized by doublets of M-M separated by two to three amino acids, and in rice the doublets are 10-30 amino acids apart. Since plant proteins are generally low in methionine, as predicted by the theory of molecular evolution (Ohta and Kimura, 1971), the high methionine contents of MRP in plants raises an interesting question on whether they have any biological role in seed development other than providing S reserves for the germinating seed.

The other class of sulfur-rich proteins in plants are the low molecular weight cysteine-rich proteins, which generally contain low levels of methionine. These include a pea seed albumin (PA1) (Higgins et al., 1986), the Bowman–Birk trypsin inhibitors and related family of isoinhibitors from soybean seed (Hwang et al., 1977; Odani and Ikenaka, 1977, 1978, Joudrier et al., 1987), and the 16.0-kDa soybean protein we identified (Kho and de Lumen, 1988). Comparison of DNA and protein sequences of the sulfur-rich proteins in rye and wheat to trypsin and α-amylase inhibitors from cereals has led to a proposal that they all belong to a superfamily of homologous proteins with a common ancestral gene (Kreis et al., 1985). Except for the trypsin inhibitors, the possible biological role of these proteins other than providing S reserves for the germinating seed is interesting and remains to be established. Certain cysteine-rich proteins in animal cells have regulatory roles, including differentiation (Hursch et al., 1987), which makes one wonder if their plant protein counterparts play any regulatory role in seed development.

The methionine plus cysteine content of the subunits of glycinins and β-conglycinins, the major storage proteins in soybean, ranges from only 0.6 to 3.0% and from virtually zero to 0.48%, respectively (Gayler and Sykes, 1985). The identification of a 12.1% methionine protein opens up the possibility of increasing the expression of the gene encoding this protein through genetic engineering and improving the nutritional quality of soybean protein. We think that introduction of a homologous gene coding for an endogenous protein could lead to a more stable integration into the genome and that enhancing the biosynthesis of an endogenous protein, such as the MRP, would be less likely to interfere with the normal biology of the seed compared with the introduction of a heterologous gene coding for a foreign protein. A 17-fold increase in the level of this protein is estimated to bring the methionine content of soybean seed up to that of the FAO reference protein (FAO/WHO, 1973), since the 10.8-kDa protein was estimated to constitute about 0.6% of the total proteins.

It is worth noting that the combination of two-dimensional gel electrophoresis and our in vitro labeling

technique for detecting methionine-containing proteins allowed us to identify the 10.8-kDa MRP, confirm its amino acid composition, and determine its N-terminal sequence without using the usual protein purification techniques. This principle should be generally applicable to cloning the gene for proteins that can be identified in the gel after electrophoresis or in transfer membranes by specific chemical reactions or biological assays. In particular, the use of PVDF membranes as the transfer support proved to be especially useful because it permits the direct amino acid analysis and N-terminal sequence analysis of as little as 1–10 μg of protein (Kennedy et al., 1988).

We have chosen the derived nucleotide sequence of the amino acid residues 3–8 and 13–17 (Figure 3) for the synthesis of mixed oligonucleotides to be used as probes for screening a cDNA library as a first step toward cloning the gene for the 10.8-kDa MRP.

ABBREVIATIONS USED

FAO, Food and Agriculture Organization; MRP, methionine-rich protein; PVDF, poly(vinylidene difluoride); IEF, isoelectric focusing; TEMED, *N,N,N′,N′*-tetramethylethylenediamine; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography.

ACKNOWLEDGMENT

We thank Natalie Saldou, Nancy Waltz, and Shirley Wong for technical assistance and Dr. Al Smith, John Gardner, and Michael Moore for amino acid microanalysis and N-terminal sequencing.

LITERATURE CITED

Altenbach, S. B.; Pearson, K. W.; Leung, F. W.; Sun, S. S. M. Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Mol. Biol.* **1987**, *8*, 239–250.

Crouch, M. L.; Tenbarge, K. M.; Simon, A. E.; Ferl, R. cDNA clones for *Brassica napus* seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. *J. Mol. Appl. Genet.* **1983**, *2*, 273–283.

de Lumen, B. O.; Kho, C. J. Identification of methionine-containing proteins and quantitation of their methionine contents. *J. Agric. Food Chem.* **1987**, *35*, 688–691.

FAO. *Protein Requirements*; FAO Nutrition Studies 16; Food and Agriculture Organization of the United Nations: Rome, 1957.

FAO/WHO. Technical Report Series 522; Food and Agriculture Organization of the United Nations: Rome, 1973; p 63.

Gayler, K. R.; Sykes, G. Effects of nutritional stress on the storage proteins of soybeans. *Plant Physiol.* **1985**, *78*, 582–585.

Goldberg, R. B.; Barker, S.; Perez-Grau, L. Regulation of gene expression during plant embryogenesis. *Cell* **1989**, *56*, 149–160.

Higgins, T. J.; Chandler, P. M.; Spencer, D.; Beach, L. R.; Blagrove, R. B.; Kortt, A. A.; Inglis, A. S. Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds. *J. Biol. Chem.* **1986**, *261*, 11124–11130.

Hursch, D. A.; Andrews, M. E.; Raff, R. A. A sea urchin gene encodes a polypeptide homologous to epidermal growth factor. *Science* **1987**, *237*, 1487–1489.

Hwang, D. L. R.; Davis Lin, K. T.; Yang, K. W.; Foard, D. E. Purification, partial characterization and immunological relationship of multiple low molecular weight protease inhibitors of soybean. *Biochim. Biophys. Acta* **1977**, *495*, 369–382.

Joudrier, P. E.; Foard, D. E.; Floener, L. A.; Larkin, B. A. Isolation and sequence of a cDNA encoding the soybean protease inhibitors PI IV and C II. *Plant Mol. Biol.* **1987**, *10*, 35–42.

Kelley, J. F. Increasing protein quality and quantity. In *Nutritional Improvement of Food Legumes by Breeding*; Milner, M., Ed.; Protein Advisory Group of the United Nations System, UN: New York, 1973; pp 179–184.

Kennedy, T. E.; Wager-Smith, K.; Barzilai, A.; Kandel, E. R.; Sweatt, J. D. Sequencing proteins from acrylamide gels. *Nature* **1988**, *336*, 499–500.

Kho, C. J.; de Lumen, B. O. Identification and isolation of methionine-cysteine rich proteins in soybean seed. *Plant Foods Hum. Nutr.* **1988**, *38*, 287–296.

Kirihara, J. A.; Petri, J. B.; Messing, J. Isolation and sequence of a gene encoding a methionine-rich 10 kDa zein protein from maize. *Gene* **1988**, *71*, 359–370.

Koshiyama, I. Chemical and physical properties of a 7S protein in soybean globulin. *Cereal Chem.* **1968**, *45*, 394–404.

Kreis, M.; Forde, B. G.; Rahman, S.; Miflin, B. J.; Shewry, P. R. Molecular evolution of the seed storage proteins of barley, rye and wheat. *J. Mol. Biol.* **1985**, *183*, 499–502.

Lei, M. G.; Tyrell, D.; Bassette, R.; Reeck, G. R. Two-dimensional gel electrophoresis of soybean proteins. *J. Agric. Food Chem.* **1983**, *31*, 963–968.

Lilley, G. G.; Caldwell, J. B.; Kortt, A. A.; Higgins, T. J.; Spencer, D. Isolation and primary structure for a novel methionine-rich protein from sunflower seeds (*Helianthus annus* L.). In *Proceedings of the World Congress on Vegetable Protein Utilization in Human Foods and Animal Feedstuffs*; Applewhite, T. H., Ed.; American Oil Chemists Society: Champaign, IL, 1989; pp 497–502.

Mak, A. S.; Jones, B. L. The amino acid sequence of wheat β-purothionein. *Can. J. Biochem.* **1976**, *22*, 835–842.

Musumura, T.; Shibata, D.; Hibino, T.; Katom, T.; Kawabe, K.; Takeba, G.; Tanaka, K.; Fujii, S. cDNA cloning of an mRNA encoding a sulfur-rich 10 kDa prolamin polypeptide in rice seeds. *Plant Mol. Biol.* **1989**, *12*, 123–130.

O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, *250*, 4007–4021.

Odani, S.; Ikenaka, T. Studies on soybean trypsin inhibitors X. Isolation and partial characterization of four soybean double-headed proteinase inhibitors. *J. Biochem. (Tokyo)* **1977**, *82*, 1513–1522.

Odani, S.; Ikenaka, T. Studies on soybean trypsin inhibitors XIII. Preparation and characterization of active fragments from Bowman-Birk proteinase inhibitor. *J. Biochem. (Tokyo)* **1978**, *83*, 747–753.

Ohta, T.; Kimura, M. Amino acid composition of proteins as a product of molecular evolution. *Science* **1971**, *174*, 150–153.

Pedersen, K.; Argos, P.; Naravana, S. V. L.; Larkins, B. A. Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of M_r 15000. *J. Biol. Chem.* **1986**, *261*, 6279–6284.

Rackis, J. Amino acids in soybean hulls and oilmeal fractions. *J. Agric. Food. Chem.* **1961**, *9*, 409–412.

Sharief, F. S.; Li, S. S. L. Amino acid sequence of small and large subunits of seed storage protein from *Ricinus communis*. *J. Biol. Chem.* **1982**, *257*, 14753–14759.

Sun, S. S. M.; Leung, F. W.; Tomic, J. C. Brazil nut (*Bertholletia excelsa H.B.K.*) proteins: Fractionation, composition and identification of a sulfur-rich protein. *J. Agric. Food Chem.* **1987**, *35*, 232–235.

Received for review April 13, 1990. Accepted July 16, 1990.