Identification of Phaseolin Polypeptide Subunits in a Crystalline Food Protein Isolate from Large Lima Beans (*Phaseolus lunatus*)

Inteaz Alli,*,† Bernard F. Gibbs,‡ Monika K. Okoniewska,† Yasuo Konishi,‡ and France Dumas‡

Food Science and Agricultural Chemistry Department, McGill University, Macdonald Campus, P.O. Box 187, Ste. Anne de Bellevue, Quebec, Canada H9X 3V9, and Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec, Canada H4P 2R2

A crystalline protein isolate from large lima beans (*Phaseolus lunatus*) was found to contain phaseolin subunits not previously identified. Mass spectrometry and partial amino acid sequence analysis revealed that a glycosylated subunit of MW = 26240 was similar to a C-terminal segment of the phaseolin polypeptides of *Phaseolus vulgaris*, while a glycosylated subunit of MW = 26113 and its nonglycosylated variant of MW = 24249 were similar to an N-terminal segment of phaseolin polypeptides of *P. vulgaris*.

Keywords: *Phaseolin; large lima beans*

INTRODUCTION

Globulin proteins represent the major storage proteins in most legume seeds. In the common bean, *Phaseolus vulgaris*, the most extensively studied species of *Phaseolus*, globulins account for 50–75% of the total protein (Müller, 1983). Two principal globulins (a legumin-like fraction, G1, and a vicilin-like fraction, G2) have been identified. Phaseolin, the dominant component has been well characterized and is regarded as a legumin-like protein with possible homology with those of other legumes such as *Pisum sativum, Vicia faba, Arachis hypogea,* and *Glycine max* (Müller, 1983). The G2 protein component, which is not as well characterized as the G1 component, is considered to be a vicilin-like protein with possible homology with those of *Vicia sp., Pisum sp.,* and *Arachis sp.* (Müller, 1983). In *P. vulgaris*, the ratio of G1:G2 has been reported to be 6:1 (Hall *et al.*, 1977).

In comparison with *P. vulgaris*, the other common species of *Phaseolus* beans (e.g., *P. lunatus, P. aureus)* have received relatively little attention as far as the molecular characteristics of the major storage proteins are concerned. The phaseolin of *P. vulgaris* has been described in terms of its polypeptides, association-dissociation behaviour, the amino acid sequences of two of the polypeptides and the glycosylation characteristics (Paaren *et al.*, 1987; Slightom *et al.*, 1983, 1985; Romero *et al.*, 1975; Sun *et al.*, 1974). Ersland *et al.* (1983) reviewed some characteristics of the storage proteins of *P. vulgaris* in relation to those of the storage proteins of other genera of legume seeds.

In our previous studies using crystalline, acid extracted proteins from *P. vulgaris*, it was found that the major protein component showed characteristics similar to those reported for phaseolin; this suggested that phaseolin represented the major component of this crystalline protein (Alli *et al.*, 1993). From previous comparative studies of a crystalline, acid extracted protein prepared from *P. vulgaris* with that prepared from *P. lunatus*, it was evident that the polypeptides similar to the phaseolin polypeptides of *P. vulgaris* were not the major polypeptides (Li *et al.*, 1989; Alli and Baker, 1983). The present work was carried out to isolate and characterize the major polypeptides in the crystalline, acid extracted protein from the large lima bean, a species of *P. lunatus*.

MATERIALS AND METHODS

Preparation of Protein Sample. A bipyramidal crystalline protein preparation was obtained from dried seeds of large lima beans (*P. lunatus*) by extraction of the ground seeds with citric acid solution (0.1 N, pH 4.0); the extraction procedure has been described previously (Alli and Baker, 1980). The proteinaceous extract was refrigerated (4 °C, 18 h) to precipitate protein material that was recovered by centrifugation and lyophilized; details of the procedure used and the resulting crystalline structure (light microscope) obtained have been reported previously (Alli and Baker, 1980). The protein content of the lyophilized isolate was determined by the micro-Kjeldahl method (AOAC, 1980).

Chromatography of Protein Isolate. The lyophilized isolate was subjected to reversed-phase high-performance liquid chromatography (RP-HPLC) using the procedure described previously (Alli *et al.*, 1993). The separation was done on a diphenyl reversed-phase column (0.46 × 25 cm length, Vydac 219TP54). A quantity (3 mg) of protein sample was solubilized in trifluoroacetic acid (TFA) solution (1 mL, 0.1%) and injected into a liquid chromatograph equipped with a diode array, UV–visible detector (HP1090, Hewlett-Packard). Elution was done at 1 mL/min using the following two-buffer gradient system: buffer A, 0.1% TFA in water; buffer B, 0.1% TFA in 70/30 acetonitrile/water, starting at 10% B and increasing to 70% B in 30 min. The eluate was monitored at 210 nm. The protein fractions that separated were collected and dried in a Speed-vac Concentrator (Savant, NY) under vacuum. The dried fractions were redissolved in trifluoroacetic acid solution (200 μL/0.1%), rechromatographed using the identical conditions described above, and then recovered for further analyses.

Amino Acid Analysis. Approximately 100 μg of the protein isolate was placed in a Corning culture tube (catalog no. 9820, 6 × 50 mm) which was previously muffled at 450 °C overnight. The tube was placed in a reaction vial and the sample was dried in the Waters Pico-Tag Work Station (Waters, Division of Millipore). Constant boiling HCl (200 μL) containing 1% phenol was added to the vial that was alternately purged (with dried nitrogen) and evacuated. After three purges, the vial was heated at 150 °C for 1 h and cooled, and

* Author to whom correspondence should be addressed [fax (514) 398-7977].
† McGill University.
‡ National Research Council Canada.

0021-8561/94/1442-2679$04.50/0

the contents of the tube were dried in a Speed-vac concentrator. The residue was dissolved in sodium citrate buffer (pH 3.0) and the analysis was performed on a Beckman System 6300 high-performance analyzer according to the procedures of Spackman *et al.* (1958) as modified by Veeraragavan *et al.* (1990).

N-Terminal Amino Acid Sequence. Automated Edman degradation was performed in a gas-phase sequencer (Model 470A) equipped with an on-line phenylthiohydantoin (PTH) analyzer (Model 120A from Applied Biosystems Inc.) employing the general protocol of Hewick *et al.* (1981). Samples were applied to TFA-treated cartridge filters coated with 1.5 mg of polybrene and 0.1 mg of NaCl (Biobrene Plus, ABI); a standard program (03RPTH, ABI) was employed for sequencing. The PTH-amino acid were identified by comparing the retention times with those of the standard amino acid derivatives.

SDS Electrophoresis. SDS–polyacrylamide gel electrophoresis, according to the procedure of Laemmli (1970), was used for separation of the polypeptide components; 4% stacking gels and 7% running gels of polyacrylamide were used. Slab gels (0.75 mm thickness) were run at 125 V until the tracking dye (bromophenol blue) reached the 1 cm mark at the bottom of the slab; the electrophoresis was performed with a Bio-Rad Mini Trans blotting system using a Tris–glycine electrode buffer (pH 8.3). Protein was detected by staining the gels with Coomassie Brilliant Blue followed by destaining with methanol/acetic acid.

Mass Spectrometric Determination of MW. Molecular weights of the crystalline protein and the HPLC-separated fractions were determined by ion spray mass spectrometry using a triple quadrupole mass spectrometer (API III LC/MS/MS system, Sciex, Thornhill, ON, Canada) according to the procedure described previously (Feng *et al.*, 1991). Multiply charged protein ions were generated by spraying the sample solution through a stainless steel capillary held at high potential. The voltage on the sprayer was set at 5.2 kV for position ion production. A coaxial air flow along the sprayer was provided to assist the liquid nebulization; the nebulizer pressure was adjusted in the range of 25–35 psi. The sample was delivered to the sprayer by a syringe infusion pump (Model 22, Harvard Apparatus, South Natick, MA) through a fused silica capillary of 100 μm i.d. The liquid flow rate was set at 1.0 μL/min for sample introduction. The interface between the sprayer and the mass analyzer was made of a small conical orifice of 100 μm diameter. The potential on the orifice was set at 35 V during calibration and was raised to 80–150 V for proteins to enhance ion signals. A gas curtain formed by a continuous flow (1.2 L/min) of nitrogen in the interface region served to evaporate the aerosol droplets and to break up the cluster formation from supersonic expansion. For MW determination, only the first rod set (Q1) of the triple quadrupole system was used, and the other two (Q2 and Q3) served as ion guidance lenses. The instrument mass-to-charge ratio scale was calibrated with the ammonium adduct ions of poly(propylene glycol)s (PPG). The unit resolution was maintained across the entire mass range for singly charged PPG calibrant ions, according to the 50% valley definition (i.e., in the group of two adjacent peaks the smaller peak was resolved to at least 50% of its full height). The same resolution setting was used for protein MW measurement to avoid possible interference on accuracy due to resolution adjustment. All protein mass spectra shown were obtained from signal averaging of multiple scans.

RESULTS AND DISCUSSION

The protein preparation contained 85.5% protein and 6.1% total carbohydrate with mannose representing the principal hexose unit (DiLollo *et al.*, 1993), showed bipyramidal crystalline microstructure when examined under the light microscope (Musakhanian and Alli, 1990), and was heterogeneous when subjected to native PAGE (Alli and Baker, 1983). RP-HPLC resulted in the separation of three major fractions (Figure 1). The early eluting fraction (F1) was relatively well separated compared to the remaining two fractions (F2, F3).

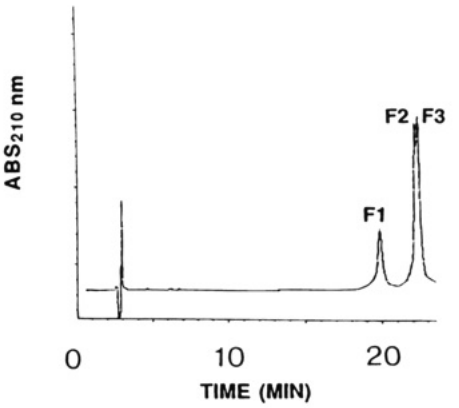

Figure 1. Reversed-phase HPLC of crystalline protein isolated from large lima beans.

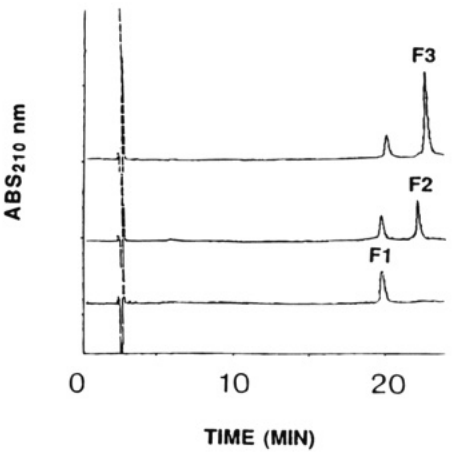

Figure 2. Rechromatography by reversed-phase HPLC of three fractions (F1, F2, and F3) recovered from the first chromatography of large lima bean crystalline protein.

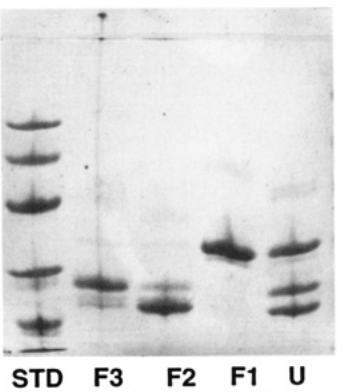

Figure 3. SDS–PAGE of fractions F1, F2, and F3 and unfractionated crystalline protein from large lima beans.

Further attempts to improve the separation of fractions F2 and F3, by modifying the chromatographic conditions, were unsuccessful. Figure 2 shows the chromatograms which were obtained from rechromatography of the three fractions, F1, F2, and F3. F1 was obtained in the relative absence of F2 and F3, while F2 and F3 were well resolved (Figure 2) but contained small quantities of F1. The fractions obtained from rechromatography were used in all subsequent analyses.

SDS–PAGE resulted in the separation of the original crystalline protein into three principal components (Figure 3). Molecular weight estimations were performed by mass spectrometry and were not attempted by SDS–PAGE. Electrophoresis of the three fractions which were separated by RP-HPLC along with the

Table 1. N-Terminal Amino Acid Sequence of Fractions Separated from Lima Bean (*P. lunatus*) Protein Isolate

fraction	N-terminal sequence
F1 from *P. lunatus* protein	SL DISNE _G_LT DIVYA SLDVV LTYVV IK_GG LFVPH YNSKA
fractment of α-type phaseolin[a] precursor starting at 249D	249 DN TIGNE FGNLT ERTDN SLNVL ISSIE MKEGA LFVPH YYSKA
fragment of β-type phaseolin[a] precursor starting at 243D	243 DN TIGNE FGNLT ERTDN SLNVL ISSIE MEEGA LFVPH YYSKA
F2 from *P. lunatus* protein	25 30 35 T SLREE EESQD NPF
F3 from *P. lunatus* protein	25 30 35 T SLREE EESQD NPF
α-type and β-type phaseolin[a] precursors	25 30 35 T SLREE EESQD NPF

[a] Amino acid sequence deduced from nucleotide sequence (Slightom *et al.*, 1983, 1985).

original protein, indicated that these fractions showed migration distances corresponding to those of the three components separated from the original protein (Figure 3); this suggests that the fractions which were separated by RP-HPLC are in all likelihood the same components as those separated by electrophoresis. Fraction F1 which was the first eluting fraction from RP-HPLC was the slowest migrating component and was relatively free of fractions F2 and F3; it will be recalled that RP-HPLC resulted in complete separation of F1 from F2 and F3 (Figure 1). SDS–PAGE of F2 and F3 indicated that each component contained a relatively minor proportion of the other component due to the close retention times of these two fractions when separated by RP-HPLC.

Published information on the amino acid sequence of *Phaseolus* beans is limited to *P. vulgaris*. Consequently, the information obtained on the present work with *P. lunatus* can be compared for homology with the proteins of *P. vulgaris*. Table 1 shows the determined N-terminal amino acid sequence of fractions F1, F2, and F3 separated from the protein isolated from *P. lunatus*, along with the sequence (deduced from nucleotide sequence, Slightom *et al.*, 1983, 1985) of α-type and β-type phaseolin precursors of *P. vulgaris*. Fractions F2 and F3 show identical N-terminal sequences in the region of residues 25–38 (residues 1–24 represent the signal peptide); this sequence is identical to the reported N-terminal of the α-type and β-type phaseolin polypeptide precursors of *P. vulgaris* (Slightom *et al.*, 1983, 1985). This indicates that fractions F2 and F3 of *P. lunatus* are similar to some N-terminal segments of the phaseolin polypeptide of *P. vulgaris*. The N-terminal sequence of fraction F1 was similar to a segment of the α-type precursor starting at residue 249 and of the β-type precursor starting at residue 243; the degree of homology is 50% for identical amino acids, and 62% for identical and homologous amino acids with respect to the α-type precursor, and 48% for identical amino acids and 60% for identical and homologous amino acids with respect to the β-type precursor. Amino acid analysis (Table 2) of the HPLC-separated fractions indicate similarity between F2 and F3 and a distinct difference between these two fractions and fraction F1. On the basis of the similarity in their amino acid compositions, the identical N-terminal amino acid sequence of F2 and F3, as well as their close retention characteristics during RP-HPLC (Figure 1), it is likely that F2 and F3 are variants of the same molecular species.

Figure 4A shows the mass spectra of the unfractionated protein; spectral interpretation reveals the presence of three components of MW = 26118, 25956, and 24252. The lowest MW component, MW = 24252, gave the highest intensity followed by the components of MW = 26118 and MW = 25956, respectively. Figure 4B shows the mass spectra of Fraction 1; spectral interpretation reveals the presence of four components which differ by a mass of 160–164 indicating the presence of a hexose sugar residue. Component B has a mass corresponding to one hexose residue less than that of component A; similarly component C has a mass corresponding to one hexose residue less than component B. The results suggest that F1 has a MW of approximately 26240 and variants with different degrees of glycosylation can be identified from its mass spectra. The components identified from the spectra of fraction F1 were not identified in the spectra of the unfractionated protein from which F1 was obtained; a likely explanation is that the signal from F1 in the unfractionated protein was suppressed by the presence of the relatively intense signals of the other components (Figure 4A).

Table 2. Amino Acid Composition of Fractions F1, F2, and F3 of *P. lunatus* Protein Isolate

	mol %		
	F1	F2	F3
Asp	14.20	12.55	12.78
Thr	3.10	2.95	3.03
Ser	5.95	8.75	8.63
Glu	14.58	16.45	16.20
Gly	9.63	5.20	5.98
Ala	5.75	5.40	5.33
Val	8.38	4.75	4.48
Met	0.50	0.75	0.73
Ile	5.38	6.20	6.05
Leu	9.00	12.65	12.23
Tyr	3.15	2.60	2.53
Phe	3.88	6.00	5.85
His	2.40	2.80	2.90
Lys	7.13	5.00	5.33
Arg	2.15	4.45	3.85
Pro	4.93	4.80	4.00

Interpretation of the mass spectra of fraction F2 (Figure 4C) reveals the presence of four components having MW of 26113, 25952, 25789, and 25627. Each component differs from the next lower MW component by 161–163, which corresponds to the MW of a hexose residue. The results suggest that fraction F2 has a molecular weight corresponding to 26113 and at least three components having different degrees of glycosylation can be identified in the mass spectra. The components of MW 26113 and 25952 identified in F2 were also identified as components in the spectra of the unfractionated protein (Figure 4A) from which F2 was obtained. Interpretation of the mass spectra of fraction F3 reveals the presence of a single component with MW 24249 (Figure 4D); there is no indication that this component is glycosylated.

Comparison of the RP-HPLC retention characteristics, amino acid composition and sequence, and MW data of fractions F2 and F3 suggest that these two

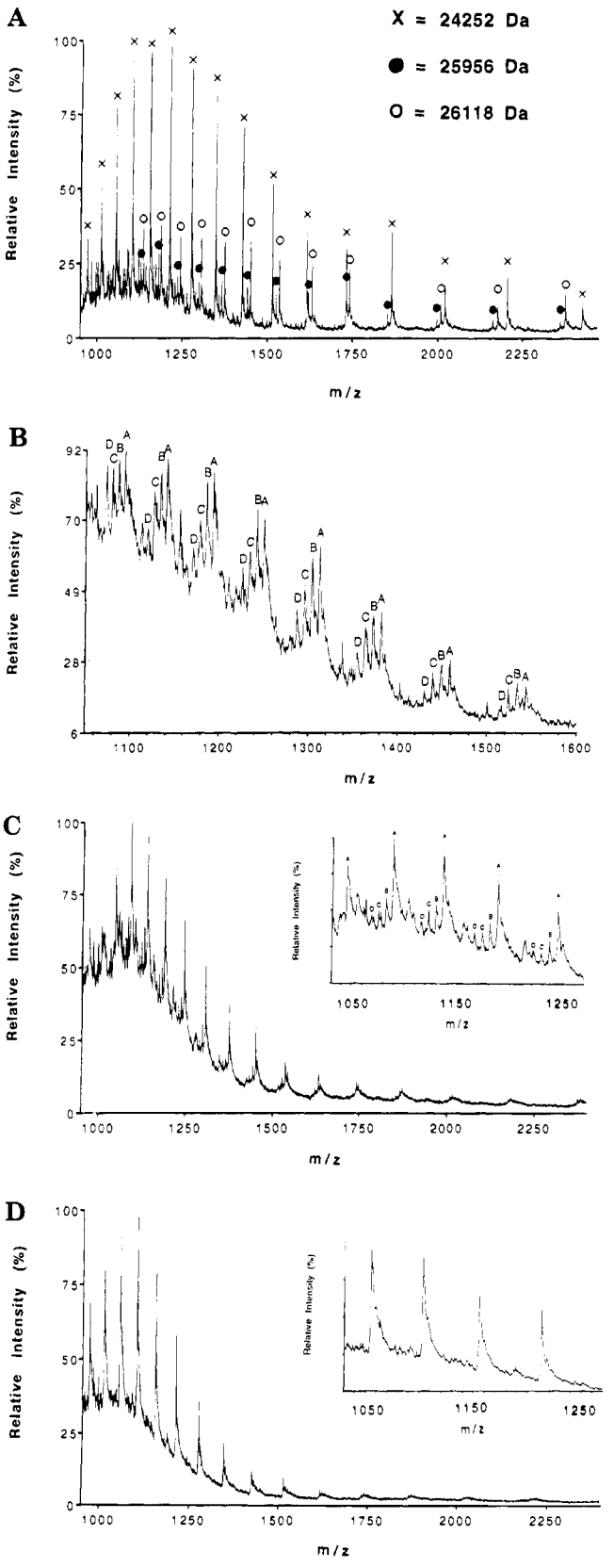

Figure 4. (A) Interpreted mass spectra of unfractionated crystalline protein from large lima beans and interpreted mass spectra of fractions F1 (B), F2 (C), and F3 (D) obtained from crystalline protein of large lima beans.

fractions could be variants which differ primarily in glycosylation. The MW and other differences between these two components strongly suggest glycosylation of F2. The actual MW difference between F2 (MW = 26113) and F3 (MW = 24249) is 1864.

An explanation for the observed differences in MW of the subunits, in relation to α-phaseolin polypeptide (MW = 49271, 436 amino acid residues and glycosylation sites at residues 258 and 347) and β-phaseolin polypeptide (MW = 47566, 421 amino acid residues and glycosylation sites at residues 252 and 341), is provided below. The MW of F1 from MS analysis was 26240; N-terminal sequence analysis indicate that F1 is homologous to a C-terminal segment of *P. vulgaris* α-phaseolin or β-phaseolin polypeptide starting at residue 249 or 243, respectively. Based on the primary structure of α- and β-phaseolin polypeptides (Slightom *et al.*, 1983, 1985), the MW of the C-terminal segments starting at residue 249 (for α-phaseolin polypeptide) and 243 (for β-phaseolin polypeptide) are 20574 and 19540, respectively. Two phaseolin polypeptides (polypeptides A and C) have been reported (Sturm *et al.*, 1987) to have two glycosylation sites with the glycan contributing a MW of 3443. Based on this glycan contribution, the calculated MW of the 249 C-terminal segment (MW = 20574) of α-phaseolin polypeptide containing the glycan would be 24104, and that of the 243 C-terminal segment (MW = 19540) of β-phaseolin polypeptide containing the glycan would be 23070; both are lower than the observed MW of F1 (MW = 26240). Because of the high degree of homology between F1 and the C-terminal segments of α- and β-phaseolin polypeptides of *P. vulgaris*, it is likely that the observed difference in MW could be due to the presence of an additional glycosylation of F1, the *P. lunatus* phaseolin polypeptide subunit. Further work is in progress to confirm this hypothesis.

Fractions F2 and F3, which have identical N-terminal sequences and are homologous with a N-terminal segment of α- and β-phaseolin polypeptides, showed MW corresponding to 26113 and 24249, respectively; MS results indicate that the MW difference between F2 and F3 was glycosylation of F2. The N-terminal segment of phaseolin polypeptide of *P. vulgaris* having MW of 26113 and 24249 does not have a potential glycosylation site. This suggests that F2 of *P. lunatus* is glycosylated and is similar to a *P. vulgaris* phaseolin polypeptide N-terminal segment, which is not glycosylated. Hence, F3 represents a nonglycosylated variant of F2.

ACKNOWLEDGMENT

Financial support for this work was provided by the Natural Sciences and Engineering Research Council of Canada.

LITERATURE CITED

Alli, I.; Baker, B. E. Constitution of leguminous seeds: VIII. The microscopic structure of proteins isolated from *Phaseolus* beans. *J. Sci. Food Agric.* **1980**, *31*, 1316–1322.

Alli, I.; Baker, B. E. Constitution of leguminous seeds. Polyacrylamide-disc gel electrophoresis of acid soluble and acid precipitated proteins prepared from baby lima beans. *J. Sci. Food Agric.* **1983**, *34*, 737–742.

Alli, I.; Gibbs, B. F.; Okoniewska, M. K.; Konishi, Y.; Dumas, F. Identification and characterization of phaseolin polypeptides in a crystalline protein isolated from white kidney beans (*Phaseolus vulgaris*). *J. Agric. Food Chem.* **1993**, *41*, 1830–1834.

AOAC. Automated Kjeldahl method for total nitrogen determination. In *Official Methods of Analysis of the Association of Official Analytical Chemists*; Horwitz, W., Ed.; AOAC: Washington, DC, 1980.

DiLollo, A.; Alli, I.; Biliarderis, C.; Barthakur, N. Thermal and surface active properties of citric acid extracted and alkali extracted proteins from *Phaseolus* beans. *J. Agric. Food Chem.* **1993**, *41*, 24–29.

Ersland, D. R.; Brown, N. W. S.; Casey, R.; Hall, T. C. The storage proteins of *Phaseolus vulgaris L.*, *Vicia faba L.* and *Pisum sativum L.* In *Seed Proteins: Biochemistry, Genetics, Nutritive Value*; Gottschalk, W., Müller, H. P., Eds.; Martines Nijhoft/Dr. W. Junk Publishers: The Hague, The Netherlands, 1983; pp 355–376.

Feng, R.; Konishi, Y.; Bell, A. W. High accuracy molecular weight determination and variation characterization of proteins ups to 80 ku by Ion Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1991**, *2*, 387–401.

Hall, T. C.; McLeester, R. C.; Bliss, F. A. Equal expression of the maternal and paternal alleles for the polypeptide subunits of the major storage protein of the bean *Phaseolus vulgaris* L. *Plant Physiol.* **1977**, *59*, 1122–1124.

Hewick, R. M.; Hunkapiller, M. W.; Hood, L. E.; Dryer, W. J. A gas-liquid solid peptide and protein sequenator. *J. Biol. Chem.* **1981**, *256*, 7990–7997.

Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of Bacteriophage T4. *Nature* **1970**, *227*, 680–685.

Li, Z.; Alli, I.; Kermasha, S. Tryptic hydrolysis *(in vitro)* of crystalline and noncrystalline proteins from *Phaseolus* beans. *J. Agric. Food Chem.* **1989**, *37*, 643–647.

Müller, H. P. The genetic control of seed protein production in legumes. In *Seed Proteins: Biochemistry, Genetics, Nutritive Value*; Gottschalk, W., Müller, H. P., Eds.; Martines Nijhoff/Dr. W. Junk Publishers: The Hague, The Netherlands, 1983; pp 309–354.

Musakhanian, J.; Alli, I. Crystalline nature of acid extracted proteins from *Phaseolus* beans. *Can. Inst. Food Sci. Technol. J.* **1990**, *23*, 47–52.

Paaren, H. E.; Slightom, J. L.; Hall, T. C.; Inglis, A. S.; Blagrove, R. J. Purification of a seed glycoprotein: N-terminal and deglycosylation analysis of phaseolin. *Phytochemistry* **1987**, *26*, 335–343.

Romero, J.; Sai-Ming, M. Sun; McLeester, R. C.; Bliss, F. A.; Hall, T. C. Heritable variation in a polypeptide subunit of the major storage protein of the bean, *Phaseolus vulgaris* L. *Plant Physiol.* **1975**, *56*, 776–779.

Slightom, J. L.; Sun, S. M.; Hall, T. C. Complete nucleotide sequence of a French bean storage protein gene: phaseolin. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1887–1901.

Slightom, J. L.; Drong, R. F.; Klassy, R. C.; Hoffman, L. M. Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. *Nucleic Acids Res.* **1985**, *13*, 6483–6498.

Spackman, D. H.; Stein, W. H.; Moore, S. Automatic recording apparatus for use in the chromatography of amino acids. *Anal. Chem.* **1958**, *30*, 1190–1200.

Sturm, A.; Van Kuik, A. J.; Vliegenthart, J. F. G.; Chrispeels, M. J. Structure, position, and biosynthesis of the high mannose and the complex oligosaccharide side chains of the bean storage protein phaseolin. *J. Biol. Chem.* **1987**, *262*, 13392–13403.

Sun, S. M.; McLeester, R. C.; Bliss, F. A.; Hall, T. C. Reversible and irreversible dissociation of globulins from *Phaseolus vulgaris* seed. *J. Biol. Chem.* **1974**, *249*, 2118–2121.

Veeraragavan, K.; Colpitts, T.; Gibbs. B. F. Purification and characterization of two distinct lipases from *Geotrichum candidum*. *Biochim. Biophys. Acta* **1990**, *1044*, 26–33.

Received for review June 7, 1994. Accepted September 19, 1994.[⊗]

[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1994.