ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 336, No. 1, December 1, pp. 10–18, 1996
Article No. 0526

# Amino Acid Sequence of a New 2S Albumin from *Ricinus communis* Which Is Part of a 29-kDa Precursor Protein

José Godinho da Silva Jr.,* Olga L. T. Machado,† Clarice Izumi,‡ Júlio César Padovan,‡,§ Brian T. Chait,¶ Urooj A. Mirza,¶ and Lewis J. Greene‡,‖,[1]

**Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, 21949-900, and Departamento de Química, Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Brazil; †Departamento de Bioquímica Médica, Instituto de Ciências Biomédicas, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, 21910-000, Brazil; §Departamento de Biofísica, Escola Paulista de Medicina, R. Botucatu, 740, São Paulo 04023-900, S.P., Brazil; ¶Mass Spectrometry Laboratory, Rockefeller University, 1230 York Avenue, New York, New York 10021-6399; and ‖Departamento de Ginecologia e Obstetrícia e ‡Centro de Química de Proteínas, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, Ribeirão Preto 14049-900, S.P., Brazil*

Received November 1, 1995, and in revised form August 19, 1996

**The isolation and sequence determination of a new 2S albumin storage protein from *Ricinus communis* seeds denoted 2S ASP-Ib are described. The fragment approach using selective enzymatic cleavage, Edman degradation, and mass spectrometry was used to demonstrate that the 11-kDa heterodimer protein linked by disulfide bridges has the following structure: short chain, GEREGSSSQQCRQEVQRKDLSSCERYLRQSSS; long chain, <QQQESQQLQQCCNQVKQVRDECQCEAIKYIAEDQIQQGQLHGEESERVAQRAGEIVSSCGVRCMR. The molecular weight of the intact protein, 11,140 ± 2, determined by matrix-assisted laser desorption mass spectrometry was consistent with the assigned structure. The S- and L-chains are identical to residues 18–49 and 66–130 of the precursor protein predicted by S. D. Irwin, J. N. Keen, J. B. C. Findlay, and J. M. Lord [(1990) *Mol. Gen. Genet.* 222, 400–408], on the basis of the structure of a cDNA isolated using probes based on the sequence of another 2S albumin, described by F. S. Sharief and S. S. L. Li [(1982) *J. Biol. Chem.* 257, 14753–14759], which we denote 2S ASP-Ia. Three of the four termini could have been produced by post-translational processing by endopeptidase(s) and carboxypeptidase(s) which utilized basic residues as the cleavage sites. Mass spectrometric evidence suggested that the protein presented microheterogeneity at its termini, i.e., truncated forms presumably due to processing heterogeneity. The present characterization of the 2S ASP-Ib protein, the second 2S albumin from *Ricinus communis* seeds, demonstrates that the 237-residue precursor protein codes for two different heterodimer proteins containing 97 and 99 residues each. This system should be useful for studying the posttranslational processing of plant storage proteins.** 

***Key Words:*** **2S albumin; *Ricinus communis;* storage protein; posttranslational processing; precursor protein; 2S ASP-Ia; 2S ASP-Ib.**

The 2S albumins are storage proteins present in *Ricinus communis* seeds. These proteins were isolated for the first time by Spies' group as an allergenic protein fraction (1, 2). Evidence that the castor bean allergens are the 2S albumin storage proteins was presented by Youle and Huang (3) and confirmed by Thorpe *et al.* (4).

Sharief and Li (5) reported the amino acid sequence of the first 2S albumin to be isolated from *R. communis* seeds. Irwin *et al.* (6) using probes based on this se-

[1] To whom correspondence should be addressed at Centro de Química de Proteínas, Faculdade de Medicina de Ribeirão Preto-USP, Av. dos Bandeirantes, 3.900, 14049-900 Ribeirão Preto, S.P., Brazil. Fax: 55-16-633-2119. E-mail: ljgreene@uhura.fmrp.usp.br.



0003-9861/96 $18.00

quence isolated a cDNA coding for a 29.3-kDa precursor. This precursor contained, as expected, the 4- and 7-kDa chains of the Sharief and Li 2S albumin. They also proposed that the albumin precursor contained a second 2S albumin on the basis of the presence of two additional glutamine-rich regions and the similar location of cysteine residues in the proprotein. Structural evidence for the existence of the putative second 2S albumin was presented in terms of sequence data for the small subunit which was isolated by SDS–PAGE. The first 10 residues were consistent with the proposed N-terminal sequence of the putative small subunit.

We have reported the isolation of a second 2S albumin from castor bean seeds (7), which differed from the one isolated by Li *et al.* in terms of isoelectric point (5, 8). On the basis of the amino acid composition of its small and large chains and partial sequence data we concluded (9) that it was the second albumin predicted from the cDNA data. In the present paper we describe the isolation and complete sequence determination of this new 2S albumin from *R. communis* and discuss the specificity of the post-translational enzymes which could be involved in processing these plant storage proteins.

## MATERIALS AND METHODS

*Materials.* Mono P (HR 5/20), Sephadex G-25F, Sephadex G-50M, PD-10 desalting column, IEF[2] precast Gel (pH 3.5–9.5), IEF p*I* standards (broad pH range), and polybuffer 74 were products of Pharmacia Biotechnology (Uppsala, Sweden). The following were products of Sigma Chemical Co. (St. Louis, MO): guanidine–HCl (2× crystallized), 4-vinylpyridine, carboxypeptidase A, carboxypeptidase B, and molecular weight standard kits (2500–17,000). Reagents for amino acid analysis by the postcolumn ninhydrin method and the PTC method, amino acid mixture H, PTH standards, dithiothreitol, and trifluoroacetic acid were from Pierce Chemical Co. (Rockford, IL). HPLC-grade acetonitrile was from Baker (Phillipsburg, NJ). L-Histidine was donated by Ajinomoto (São Paulo, Brazil). Trypsin, endoproteinase Lys-C, and endoproteinase Glu-C were products of Boehringer-Mannheim (Mannheim, Germany).

*Amino acid analysis.* Peptides were hydrolyzed in constant boiling 6 N HCl for 22 h at 110°C under vacuum. Most of the samples, 2–10 nmol, were analyzed by the method of Spackman *et al.* (10) on an automatic instrument (11), using a 4-nmol standard. Peptides obtained in small quantities (0.5 to 2.0 nmol) were analyzed after vapor-phase hydrolysis by the PTC method (12) using a PicoTag column with a 100-pmol standard. The recoveries of peptides reported here are based on the previous step and are not cumulative.

*Edman degradation.* Automatic Edman degradation was carried out on 250 to 3000 pmol peptide using a Porton Model 2020/2090 liquid-phase instrument. The HPLC system was standardized with 100 pmol of each PTH derivative. Manual Edman degradation was carried out on 1 to 4 nmol peptide as described by Fushitani *et al.* (13). PTH amino acids were separated by $C_{18}$ reverse-phase HPLC (14) using 100 pmol of each PTH derivative to standardize the system.

*Electrophoresis.* Isoelectric focusing was carried out in a 2117 Multiphor II electrophoresis system on an ampholine PAG plate (245 × 110 × 1 mm), pH range 3.5–9.5, at 10°C, 20 W, for 1.5 h, as recommended by the manufacturer.

SDS–Tricine PAGE was performed using stacking, spacer, and separating gels as described by Schägger and Von Jagow (15) with a 16.5% separating gel. Protein was detected with Coomassie brilliant blue R-250 (10 mg/ml) in methanol/acetic acid/water (45/10/45) at 8°C (cold room).

*Isolation of 2S protein by chromatofocusing.* The protein was isolated from fraction SF-III which was prepared from detoxified extracts of *R. communis* seeds (2) by SP-Sephadex chromatography followed by gel filtration on Sephadex G-75 (7). One- to 5-mg batches of SF-III were submitted to chromatofocusing on a Mono P column (HR 5/20) previously equilibrated with 25 mM L-His, pH 6.0 (16). After sample application (0.5 ml) in the same buffer, the column was eluted isocratically for 7.5 ml. The column was developed with 50 ml aqueous polybuffer 74, 1:20 (v/v), pH 4.0, to establish the pH gradient. Effluent pH was monitored with a flow cell or on a tube by tube basis, and protein was detected by absorbance at 280 nm (see legend to Fig. 1A for operating conditions). Effluent containing the major protein peak and eluting near pH 4.80 was pooled and the protein was separated from polybuffer and L-His by gel filtration on Sephadex G-50M (0.9 × 130 cm) eluted with 0.2 M acetic acid.

Since the material in Fig. 1A presented two major bands on IEF–PAGE (pH 4.7–4.9) and discontinuous PAGE (12.5%, data not shown), the pool was submitted to a second preparative chromatofocusing step designed to provide more resolution near pH 4.8. The second elution system utilized 25 mM L-His, pH 5.85, as starting buffer and polybuffer 74 (1:20, v/v), pH 4.25. The column was equilibrated with starting buffer and the sample (0.5–1 mg) was applied in 0.5 ml starting buffer 5 min (2.5 ml) after elution with polybuffer had begun. The column was eluted with 40 ml polybuffer. Effluent containing protein near pH 4.80 was separated from polybuffer and His as described above.

*Reduction and pyridylethylation of the 2S protein.* The protein (360 nmol) was denatured in 6 M guanidine–HCl containing 0.1 M Tris–HCl buffer, pH 8.4, and 20 mM EDTA for 8 h at 38°C. After reduction with 16 mg (105 μmol) DTT for 3 h at 38°C, the protein was reacted with 60 μl (560 μmol) 4-vinylpyridine for 4 h (17). The reaction was stopped by lyophilization and the mixture was filtered through a PD-10 column (Pharmacia) and then Sephadex G-25F (0.9 × 120 cm) both eluted with 0.2 M acetic acid. Protein was detected by absorbance at 280 nm and the product, a mixture of S- and L-chains, was dried in a SpeedVac system.

*Separation of S- and L-chains.* The chains were separated, rechromatographed when necessary, and tested for homogeneity by $C_{18}$ reverse-phase HPLC. Because of the low-load capacity of the separations illustrated in Figs. 1C and 1D, the runs were repeated 10 times.

$C_{18}$ reverse-phase chromatography was carried out using two LKB-HPLC 2248 pumps and an LKB-VWM 2141 dual-wavelength detector at room temperature (22°C) at 0.5 ml/min. The conditions used to equilibrate and develop the column are indicated in the legend to Figs. 1C and 1D. Samples were injected in 50- to 100-μl volumes and effluent was collected by hand on the basis of absorbance and dried in a SpeedVac system.

*Endoproteinase Lys-C hydrolysis.* The S-chain, 150 nmol, and the L-chain, 40 nmol, were each hydrolyzed with the enzyme as described by the manufacturer at a 1/50 (w/v) enzyme/substrate ratio in 25 mM Tris–HCl buffer, pH 8.4, for 8 h, at 38°C. The reactions were stopped by lyophilization and the products separated by HPLC.

*Endoproteinase Glu-C hydrolysis of fragment 1-16 from the L-chain.* The peptide (5.0 nmol) was hydrolyzed as described by the

[2] Abbreviations used: DTT, dithiothreitol; IEF, isoelectric focusing; PEC, pyridylethyl-S-cysteine; PTC, phenylthiocarbamyl; PTH, phenylthiohydantoin; TFA, trifluoroacetic acid.

manufacturer with the enzyme at a 1/36 (w/w) enzyme/substrate ratio in 25 mM ammonium bicarbonate, pH 7.8, for 6 h at 25°C at which time the same amount of enzyme was added and hydrolysis was continued for 12 h at 25°C. The reaction was stopped by lyophilization and the products were separated by HPLC.

*Trypsin hydrolysis of fragment 29-65 from L-chain.* Peptide 29-65 (4.3 nmol) was hydrolyzed with bovine trypsin, at a 1/50 (w/w) enzyme/substrate ratio, in 100 mM Tris–HCl buffer, pH 8.4, for 24 h at 38°C. The reactions were stopped by lyophilization and the products separated by HPLC.

*Carboxy-terminal analysis.* Carboxypeptidase A and B hydrolysis was carried out in 20 mM ammonium bicarbonate buffer, pH 8.12, at 40°C (18). The reaction was stopped by the addition of 30 $\mu$l glacial acetic acid and the material dried and applied to the amino acid analyzer, calibrated with a 4-nmol standard solution of amino acids. Peptide fragment 1-18 from the S-chain (1.0 nmol) was hydrolyzed with 0.25 $\mu$g carboxypeptidase B for 0.5 h. Another sample of the same peptide (3.6 nmol) was incubated with 0.25 $\mu$g carboxypeptidase B for 3 h and carboxypeptidase A (0.25 $\mu$g) was added for 0.5 h. The L-chain (2.0 nmol) was incubated with 0.25 $\mu$g carboxypeptidase B for 3 h and then with 0.25 $\mu$g carboxypeptidase A for 0.5 h.

*Mass spectrometry.* Matrix-assisted laser desorption mass spectrometry (Fig. 1B) was carried out on a time-of-flight spectrometer constructed at Rockefeller University (19). A matrix of $\alpha$-cyano 4-hydroxycinnamic acid was used to prepare the samples as described previously (20).

Electrospray analysis was performed with a Finnigan-MAT TSQ-700 triple-quadrupole mass spectrometer. The negative-ion MS–MS spectrum of the tetrapeptide corresponding to residues 1–4 of the L-chain was obtained from a solution containing 30 pmol/$\mu$l of peptide in 50:50 (v/v) water/isopropanol. The solution was infused at a rate of 3 $\mu$l/min through a 100-$\mu$m-i.d. fused silica capillary. The spectrum given in Fig. 3 is an average of 16 scans obtained at a rate of 3 s/scan. The deprotonated singly charged intact peptide ($m/z$ 513) was selected by the first quadrupole analyzer. Collision-induced dissociation of the $m/z$ 513 ion was performed in the second stage (rf only octapole) region of the mass spectrometer (argon gas pressure 2.6 mTorr). The resulting fragments were analyzed in the third-stage quadrupole of the mass spectrometer.

## RESULTS

### *Isolation of the Protein and Its Polypeptide Chains*

The 2S protein was purified from batches of 1–5 mg of fraction SF-III by chromatofocusing as illustrated in Fig. 1A. The major peak eluted at pH 4.80 from several chromatofocusing runs was combined and submitted to a second preparative run using a flatter pH gradient. The pools of protein thus obtained were homogeneous by analytical chromatofocusing and also by IEF–PAGE (p*I* 4.90) (data not shown). The overall yield for repetitive preparative chromatofocusing was 3.6% (5 mg from 140 mg of fraction SF-III).

The laser desorption mass spectra of 2S protein is shown in Fig. 1B. The major component observed in this spectrum had a mass of 11,140 ± 2, which is consistent with the heterodimer structure proposed here with the amino-terminal glutamine of the long chain being in the pyrrolidone carboxylic acid form. Note that the sample contained components with mass up to 540 units below the major component, which probably correspond to truncated forms.

The most abundant species in the spectrum of the S-chain had a mass of 3915.8 which corresponds to the assigned structure with its two cysteine residues alkylated with PEC (calculated mass = 3.915.3, $\Delta$ = +0.5), and the spectra also contained the small peak with a measured mass of 3573.1 which corresponds to the sequenced alkylated S-chain minus $G^{18}$-$E^{19}$-$R^{20}$ at the amino terminus (calculated mass = 3543.0, $\Delta$ = +0.1) (data not shown).

Two polypeptide chains were isolated after reduction and pyridylethylation of the 2S protein when the desalted reaction mixture was submitted to reverse-phase chromatography on a $\mu$Bondapack $C_{18}$ column (data not shown). The effluent corresponding to the major S-chain peak obtained in 59% yield was combined and on rechromatography on a Vydac $C_{18}$ column behaved like a single component (Fig. 1C). The L-chain, which had been eluted with a steep acetonitrile gradient, presented two peaks when rechromatographed on a Vydac $C_{18}$ column under near-equilibrium conditions (data not shown). Effluent corresponding to the major peak behaved like a single component upon rechromatography (Fig. 1D). The lower yield for the L-chain (26%) was due to the need to carry out a second RP-HPLC purification step.

Amino acid composition data are reported in Table I for the 2S protein (column 1), S-chain (column 2), and L-chain (column 3). The integral values given in parentheses were obtained from the sequence data. The amino acid composition of the 2S protein is the same as that obtained from the sequence data except for uncorrected serine (9.4 vs 12), uncorrected cysteine (3.8 vs 8), and Glx (28.9 vs 32). This was probably due to the instability of the first two amino acids and the high content of Glx for which ±8% (error of analysis) is equivalent to two to three residues. Another indication of the homogeneity of 2S protein was the absence of the amino acids Thr, Pro, and Phe.

The data for the chains obtained after reduction and S-pyridylethylation in columns 2 and 3 are also within ±10% of the integral values obtained by sequencing, except for residues sensitive to acid hydrolysis such as Ser, Met, PEC and sterically hindered residues $Ile^{55}$-$Val^{56}$ of the long chain. Indeed, only Lys and Asp of the L-chain differed by more than ±10% from the sequence data but they were obtained in the expected amounts in the tryptic peptides.

### *Primary Structure*

The sequence of the S-chain was determined by automatic and manual Edman degradation of intact S-

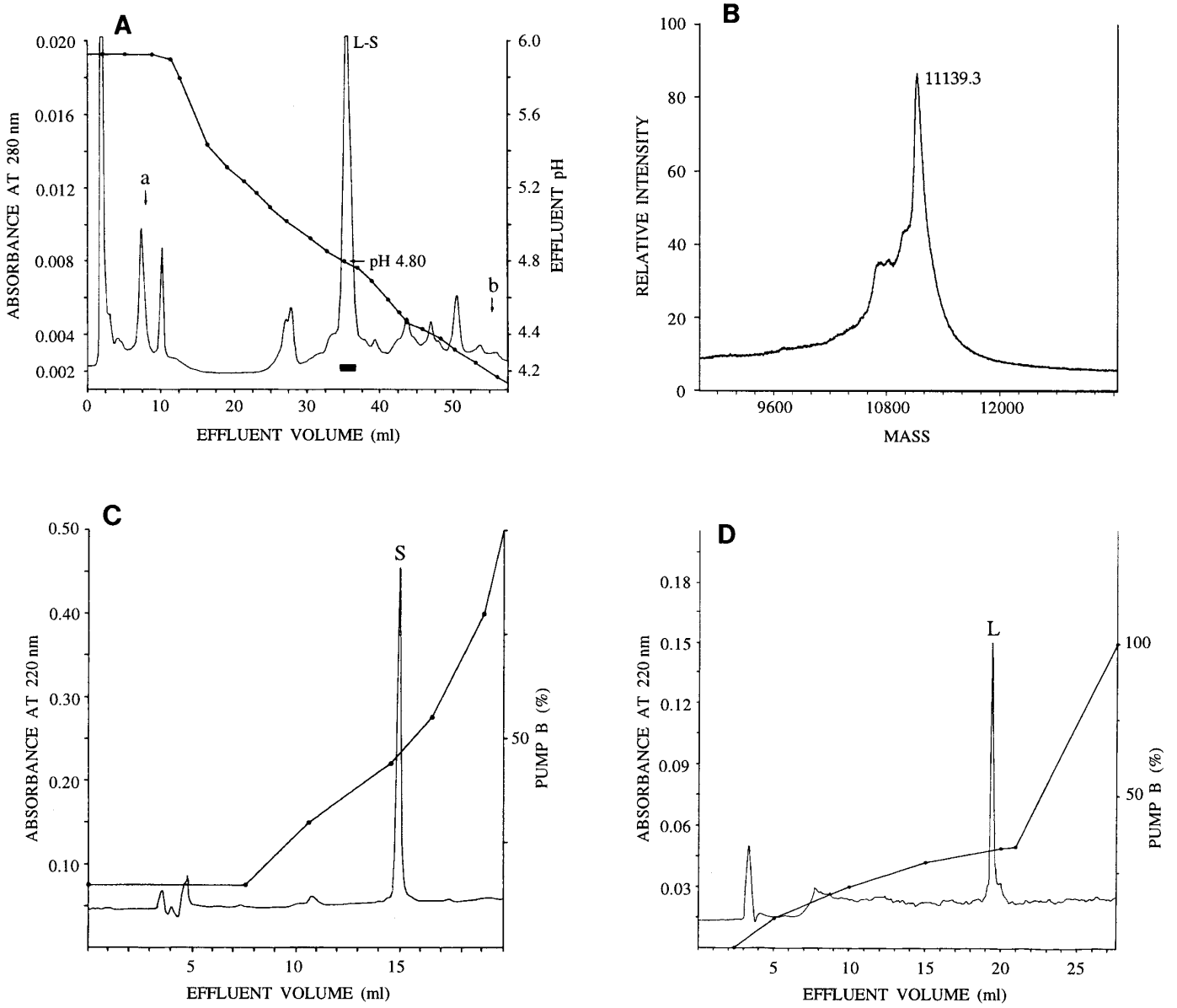


**FIG. 1.** Isolation of the 2S protein and the S- and L-chains. (A) Preparative chromatofocusing of fraction SF-III on MonoP resin. The sample (1 mg SF-III/0.5 ml starting buffer) was applied to the column (HR 5/20), which was first eluted with starting buffer, 25 mM His, pH 6.0, for 7.5 ml, and then dilute polybuffer, pH 4.0 (arrow a). The column was developed at 0.5 ml/min, 22°C and regenerated with 1 M NaCl (arrow b). The effluent was combined on the basis of absorbance at 280 nm or in volumes of 2.5 ml. Effluent pH was measured manually. (B) Matrix-assisted laser desorption spectra of 2S protein. (C) Analytical rechromatography of S-chain after reduction and pyridylethylation of 2S protein. The peptide (2 nmol) was rechromatographed on a Vydac $C_{18}$ column (4.3 × 200 mm) using 0.1% TFA in pump A and 80% acetonitrile (v/v) containing 0.08% TFA (v/v) in pump B, using the gradient indicated in the figure. Absorbance was measured at 220 and 254 nm and 220 nm is reported. (D) Analytical rechromatography of the L-chain after reduction and pyridylethylation of 2S protein. The peptide (2 nmol) was rechromatographed under the same conditions as described in C, using the gradient indicated in the figure.

chain and of the carboxy-terminal peptide obtained after hydrolysis with endoproteinase Lys-C. The sequence of the L-chain was determined by automatic Edman degradation of peptides obtained after hydrolysis with endoproteinase Lys-C and endoproteinase Glu-C. The amino-terminal sequence was determined by mass spectrometry of the peptide obtained after hydrolysis with endoproteinase Glu-C.

**TABLE I**

**Amino Acid Composition of a 2S Protein Extracted from *Ricinus communis* and S- and L-Chains Obtained after Reduction and Vinylpyridylethylation**[a]

| Amino acid | 1<br>2S protein | 2<br>S-chain | 3<br>L-chain |
|---|---|---|---|
| Lys | 3.00 (3)[b] | 1.18 (1) | 1.48 (2) |
| His | 0.98 (1) | (0) | 1.08 (1) |
| PEC[c] | | 1.84 (2) | 4.99 (6) |
| Arg | 10.14 (10) | 5.00 (5) | 4.36 (5) |
| Asp | 3.82 (4) | 1.07 (1) | 3.57 (3) |
| Ser | 9.43 (12) | 5.73 (8) | 3.84 (4) |
| Glu | 28.89 (32) | 8.27 (9) | 24.70 (23) |
| Gly | 5.57 (6) | 1.85 (2) | 4.20 (4) |
| Ala | 3.82 (4) | 0.12 (0) | 4.20 (4) |
| Cys | 3.83 (8) | (0) | (0) |
| Val | 5.61 (6) | 0.98 (1) | 3.92 (5) |
| Met | 1.00 (1) | (0) | 0.41 (1) |
| Ile | 3.83 (4) | (0) | 2.88 (4) |
| Leu | 4.11 (4) | 1.79 (2) | 1.97 (2) |
| Tyr | 2.11 (2) | 0.87 (1) | 0.96 (1) |
| Total | (97) | (32) | (65) |
| Yield[d] | | 59% | 26% |

[a] Data are reported as mol/mol peptide for two to three hydrolysates whose differences were ⩽8%. Amino acids ⩽0.1 mol/mol peptide are not reported. No corrections were made for destruction of amino acids during hydrolysis or incomplete peptide bond hydrolysis.

[b] The numbers within parentheses are integral values obtained from sequence data reported in this paper as were the total numbers of residues/mol.

[c] PEC, pyridylethyl-S-cysteine.

[d] The yields for the S- and L-chains are for the purified peptides (see Fig. 3).

The sequence data for the S-chain and L-chain are summarized in Fig. 2. The amino acid composition of each peptide was consistent with the Edman, carboxypeptidase, and/or mass spectrometric data used to assign the position of each amino acid.

*S-Chain*

Hydrolysis of the S-chain which contained only one Lys with endoproteinase Lys-C provided two fragments, corresponding to residues 1–18 and 19–32, which were separated by reverse-phase chromatography (data not shown). The sum of the amino acids of the fragments corresponded to the amino acid composition of the starting material, and the recoveries were 64 and 61%, respectively. Fragment 1-18 was identified as the amino-terminal fragment on the basis of its Edman degradation (residues 1–5) which corresponded to the amino terminal of the parent peptide which was sequenced until residue 20. Complementary data were obtained by digestion with carboxypeptidases A and B, which released 1 mol each of Lys and Arg, indicating the sequence $Arg^{17}$-$Lys^{18}$, which is consistent with the Edman data.

The Edman degradation of fragment 19-32 of the S-chain defined its sequence (data not shown). Note that the sequence $D^{19}$-$L^{20}$ was demonstrated in both the parent peptide and the 19-32 fragment.

The S-chain corresponds to residues 18 to 49 of the precursor protein predicted from cDNA data[3] (cf. Fig. 4).

*L-Chain*

Hydrolysis of the L-chain which contained two lysine residues with endoproteinase Lys-C provided the expected three peptide fragments 1-16, 17-28, and 29-65 which were recovered in 60, 55, and 51% yield, respectively. Peptide 1-16 was assigned the amino-terminal position because it was Edman negative as was the parent L-chain and because it contained one Lys. Peptide 17-28, whose sequence was determined by Edman degradation, was assigned to the central position because it contained lysine. Peptide 29-65 was assigned to the C-terminal position because it did not contain lysine but did contain Met (0.99 mol/mol) and Arg (0.99 mol/mol) which were shown to be the C-terminal residues of the parent L-chain by carboxypeptidases A plus B. The fragment alignment is shown in Fig. 2B.

Peptide 1-16, when treated with endoproteinase Glu-C, provided two peptides: 1-4 (76% yield) was Edman negative and contained only Glx upon acid hydrolysis. On the basis of the amount of Glx recovered and assuming it was a tetrapeptide (see Fig. 3 for mass spectrometry data), its recovery was 76%. No amino acid composition data were obtained for peptide 5-16 but its amino acid sequence was determined by Edman degradation. When four Glx were subtracted from the amino acid composition of fragment 1-16 an "expected" amino acid composition for 5-16 was obtained which was consistent with the automatic Edman data.

The amino acid sequence of fragment 29-65 was determined by Edman degradation, except for residues 49, 51, 52, 63, and 65.

Four peptides were obtained when fragment 29-65 was digested with trypsin. These fragments from the L-chain were ordered on the basis of the following considerations. Peptide 63-65, PEC-Met-Arg recovered in 76% yield, was assigned to the carboxyl terminal because it contained the only Met in the molecule which was shown to occupy the penultimate position of the L-

[3] The amino acid residues are numbered in Fig. 4 as they occur in the precursor protein reported by Irwin *et al.* (6). Thus, residues 1 to 32 of the sequenced S-chain correspond to residues 18 to 49 of the precursor protein and the sequenced L-chain residues 1 to 65 correspond to residues 66 to 130 of the precursor protein.

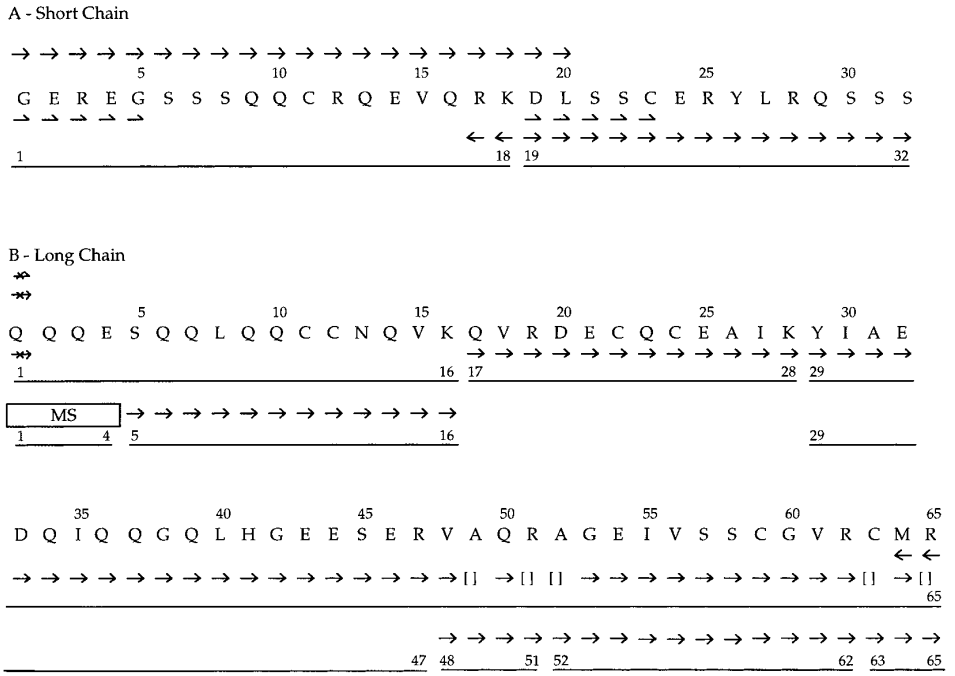


**FIG. 2.** Summary of the data used to assign the sequences of the S- and L-chains of the 2S albumin storage protein (2S ASP-Ib). The S-chain was cleaved by selective hydrolysis with endoproteinase Lys-C and the mixture was separated by HPLC. The L-chain was cleaved by selective hydrolysis with endoproteinase Lys-C, endoproteinase Glu-C, and trypsin. The fragments were separated by HPLC. The arrows above the sequences of the S- and L-chains indicate data obtained for the intact chains. The remaining arrows indicate data obtained for the peptide fragment indicated below the arrows. →, data obtained by automatic Edman degradation; ⇀, data obtained by manual Edman degradation; −x→, unsuccessful automatic Edman degradation; and ←, carboxypeptidase digestion.

chain by carboxypeptidases A and B (data not shown). Peptide 29-47 (obtained in 60% yield) was identified as the amino-terminal fragment because it contained the only residues of Tyr, Asp, Leu, and His present in fragment 29-65 and these residues were shown by automatic Edman degradation to occupy positions 29, 33, 40, and 41 of the parent peptide 29-65. Peptide 52-62 was assigned to the penultimate position because it contained the second residue of PEC present in peptide 29-65 at position 59 (data not shown). Therefore, peptide 48-51 (74% yield) must occupy the second position of the proposed alignment of the four fragments obtained by tryptic hydrolysis of fragment 29-65.

This alignment of the peptides is consistent with the 37-cycle partial Edman sequence of fragment 29-65. The sequence data for the fragments provide primary data for residues 49, 51, 52, 63, and 65 which were not identifiable by automatic Edman degradation of peptide 29-65. The amino acid compositions of the fragments of peptide 29-65 are also consistent with the sequence data. Figure 2B summarizes the data for sequence assignment for the L-chain which corresponds to residues 66 to 130 of the precursor protein (cf. Fig. 4).

*Mass Spectrometry*

Fragment 1-4 obtained after hydrolysis of the L-chain with endoproteinase Lys-C and endoproteinase Glu-C contained >98% Glx on acid hydrolysis. It was Edman negative as was the L-chain from which it was prepared with endoproteinase Glu-C (Fig. 2B). The *m/z* of the deprotonated single-charged intact peptide, 512.8 (Fig. 3), corresponded to <Glu-Gln-Gln-Glu-$O^-$ (calculated *m/z* 513.4, $\Delta$ = 0.6) but not to <Glu-Glu-Gln-Glu-$O^-$ (calculated *m/z* 514.4, $\Delta$ = 1.6) or <Glu-Glu-Glu-Glu-$O^-$ (calculated *m/z* 515.4, $\Delta$ = 2.6). Furthermore, all of the other major negative ions obtained by collision-induced decomposition were consistent with the structure <Glu-Gln-Gln-Glu-OH as indicated in the insets in Fig. 3.

## DISCUSSION

The most important result of the present study is that we have isolated and sequenced a new heterodimeric 2S albumin storage protein from *R. communis* seeds whose structure was predicted by Irwin *et al.* (6) on the basis of the structure of a cDNA. The cDNA was isolated with probes constructed using the

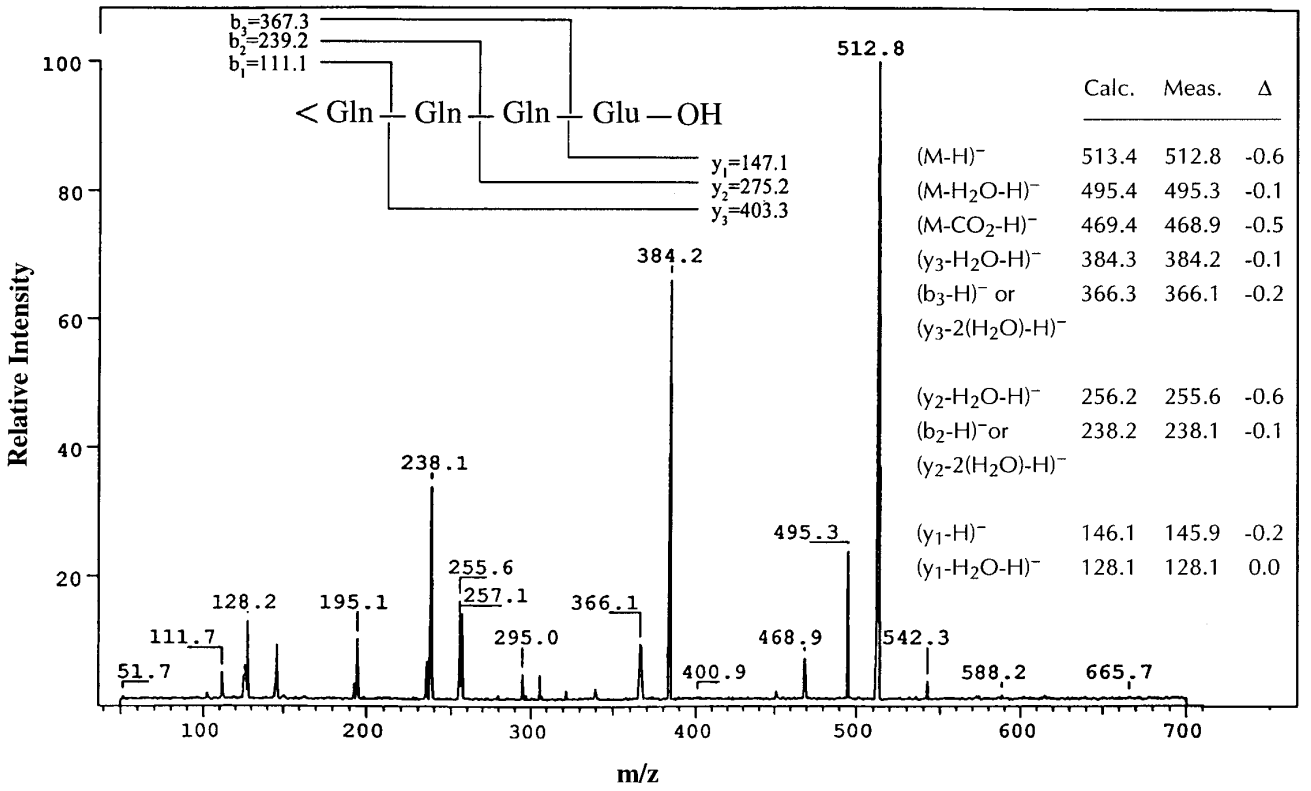


**FIG. 3.** Negative-ion electrospray ionization mass spectrum of the collision-induced fragment ions obtained from the *m/z* 513 deprotonated intact tetrapeptide (M-H). The peptide was introduced into the spectrometer at 90 pmol/min The collision energy was 26 eV and the argon gas pressure in the collision cell was 2.6 mTorr. The insets indicate the sequences of the peptide, calculated *m/z* values for major fragments, and observed *m/z* values. M, intact molecule.

amino acid sequence of another 2S albumin described by Sharief and Li (5). Thus, we have demonstrated the validity of their prediction that the putative 237-residue large protein precursor contains two different heterodimeric 2S albumins with chains containing 65 and 32/34 residues and having cysteine residues in the same relative positions and thus presumably identical disulfide bridges.

The primary structure of the 2S albumin storage protein documented here and denoted 2S ASP-Ib[4] indicated by the light areas of Fig. 4 corresponds to residues 18 to 49 (S-chain) and 66 to 130 (L-chain) in the putative precursor protein. Similarly, the 2S albumin storage protein (2S ASP-Ia) described by Sharief and Li (5) corresponds to residues 136 to 169 (S-chain) and residues 173 to 237 (L-chain). Presumably the shaded sequences 1 to 17, 50 to 65, 131 to 135, and 170 to 172 are required for formation of the disulfide bonds and the conformation of the precursor before posttranslational processing.

The 2S storage protein whose sequence is reported here (2S ASP-Ib) presents conserved cysteine residues in both the S- and L-chains typically found in seed proteins of the 2S albumin superfamily (6). Sequence comparison of 2S ASP-Ib and other 2S storage albumins (21–23), some protease inhibitors (24), and sweet proteins (25) shows homology.

A combination of the selective enzymatic cleavage fragment approach and automatic and manual Edman degradation was used because of the high content of Glx, 9/32 (28%), of the S-chain and 23/65 (35%) of the L-chain and the fact that the L-chain was Edman negative. The amino acid composition of fragments, as well as the results of manual Edman degradation of fragments and treatment with carboxypeptidases, provided complementary confirmatory evidence for most of the results of automatic Edman degradation of the L-chain and primary evidence for the final residues of the automatic Edman degradation which could not be identified. The structure of the amino-terminal peptide of the

[4] We propose to call the protein described here 2S albumin storage protein I-b (2S ASP-Ib), where I indicates the cDNA described by Irwin *et al.* (6) and b indicates the second protein coded by I to be isolated and sequenced. On this basis, the protein described by Sharief and Li (5) would be denoted 2S albumin storage protein I-a (2S ASP-Ia).

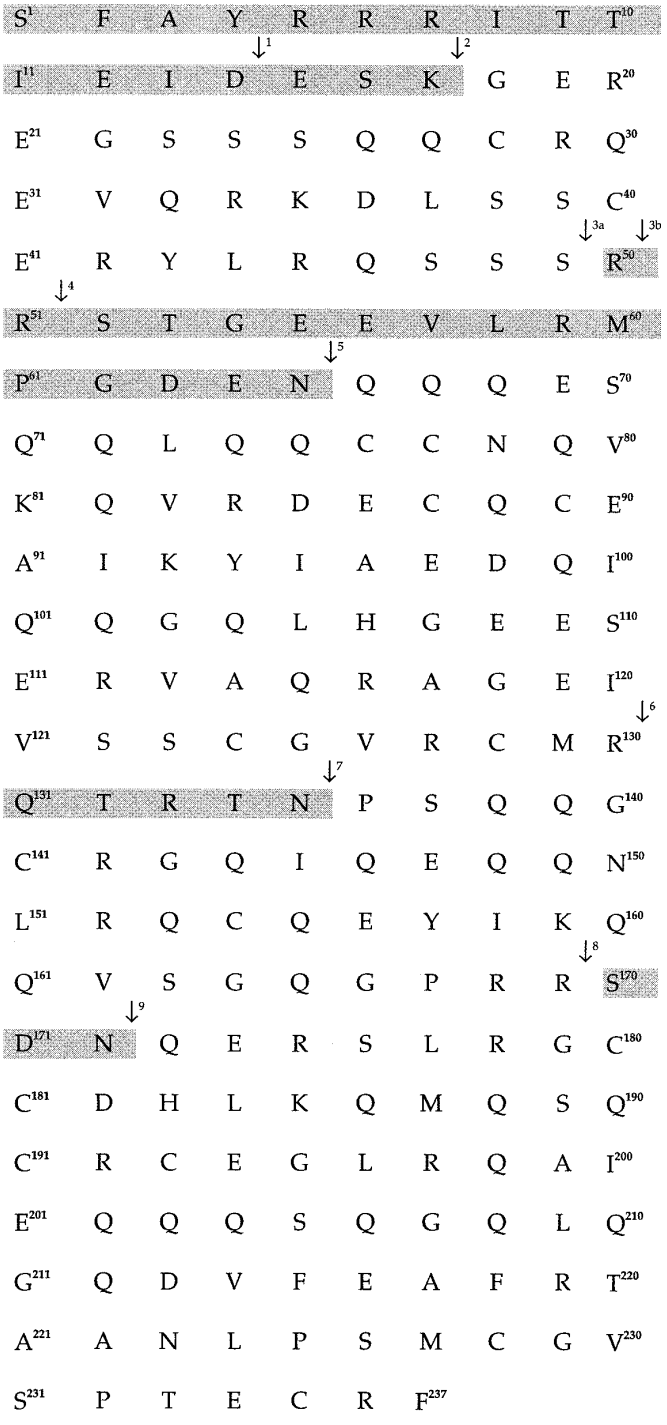


**FIG. 4.** Primary sequence of the putative 2S albumin storage precursor protein. The structure is based on cDNA data reported by Irwin *et al.* (6). Residues 18 to 49 and 66 to 130 correspond to the 2S albumin storage protein, 2S ASP-Ib documented here, and residues 136 to 169 and 173 to 237 correspond to 2S ASP-Ia sequenced by Sharief and Li (5) which was used to obtain probes for isolation of the cDNA. The shaded residues are parts of the putative precursor protein which are not present in the 2S albumin storage proteins after processing. The minimum number of cleavages using basic residues as hydrolysis sites to produce 2S ASP-Ib is indicated by arrows 2, 3a, 3b, 5, and 6. The cleavage sites suggested by Irwin *et al.* (6) correspond to arrows 1, 4, 5, and 7.

L-chain (<Glu-Gln-Gln-Glu released by endopeptidase Glu-C), a challenge for wet chemistry, was determined easily by mass spectrometry.

Thus, the primary sequence and especially the amino- and carboxyl-terminal residues assigned here were obtained from two or three independent sets of data. Yet, the mass spectra data suggest sequence microheterogeneity by the presence of truncated forms of the 2S protein (Fig. 1B) and by the presence of the desG$^{18}$-E$^{19}$-R$^{20}$ form of the S-chain (data not shown). We can rationalize these apparently contradictory results which indicate a unique structure from the sequencing data of the chains but microheterogeneity of the parent protein by mass spectrometry by suggesting that after reduction and alkylation we isolated essentially only one form of each chain which provided unambiguous sequencing data.

Figure 4 provides a hypothetical processing scheme for the structure of 2S ASP-Ib which would include (i) specificity for monobasic sites (26, 27) at arrow 2 (K$^{17}$-G$^{18}$) and arrow 6 (R$^{130}$-Q$^{131}$), (ii) specificity for paired basic residues (28), arrow 3b (R$^{50}$-R$^{51}$), followed by (iii) a carboxypeptidase B-like cleavage at arrow 3a (S$^{49}$-R$^{50}$), and finally (iv) an asparagine (29, 30)-specific cleavage at arrow 5 (N$^{65}$-Q$^{66}$). Similarly, 2S ASP-Ia could have been produced by cleavage at R$^{169}$-S$^{170}$ (arrow 8), N$^{135}$-P$^{136}$ (arrow 7), and N$^{172}$-Q173 (arrow 9). The fact that other basic and N-X residues are intact in the product protein can be explained on the basis of the conformation of the protein and the specificity of the processing enzymes.

The structure proposed by Irwin *et al.* (6) for 2S ASP-Ib contained 10 residues more than those demonstrated here. It is important to note that mass spectrometry could have detected higher molecular weight species differing in only one more amino acid residue, if they had been present. It is beyond the scope of this discussion to survey the literature on plant proteolytic enzymes. The only structural information for precursor processing enzymes in castor bean seeds that we are aware of is the description of the processing of ricin which was recently reviewed by Lord *et al.* (31). An N-G specific processing enzyme that acts selectively on soybean glycinin has been described (32). The possibility cannot be excluded that the processing of the 2S albumin storage protein involves first selective steps and then other nonselective steps as this storage protein is mobilized by the seedling.

Finally, it should be emphasized that the processing of plant precursor protein having 237 residues coding

for two heterodimers containing approximately 97 residues demonstrates a high level of efficiency for the utilization of precursor proteins comparable to that in the animal system, which was predicted by Irwin *et al.* (6) and demonstrated in the present study.

## ACKNOWLEDGMENTS

Research supported by FINEP (FNDCT), CNPq (PADCT), CNPq, and CAPES.

## REFERENCES

1. Spies, J. R., and Coulson, E. J. (1943) *J. Am. Chem. Soc.* **65,** 1720–1725.
2. Spies, J. R. (1974) *J. Agric. Food Chem.* **22,** 30–36.
3. Youle, R. J., and Huang, A. H. C. (1978) *Plant Physiol.* **61,** 1040–1042.
4. Thorpe, S. C., Kemeny, D. M., Panzani, R. C., McGurl, B., and Lord, J. M. (1988) *J. Allergy Clin. Immunol.* **82,** 67–72.
5. Sharief, F. S., and Li, S. S. L. (1982) *J. Biol. Chem.* **257,** 14753–14759.
6. Irwin, S. D., Keen, J. N., Findlay, J. B. C., and Lord, J. M. (1990) *Mol. Gen. Genet.* **222,** 400–408.
7. Costa e Silva, M. C., Machado, O. L. T., and Silva, J. G., Jr. (1991) *Braz. J. Med. Biol. Res.* **24,** 141–143.
8. Li, S. S.-L., Lin, T. T.-S., and Forde, M. D. (1977) *Biochim. Biophys. Acta* **492,** 364–369.
9. Machado, O. L. T., and Silva, J. G., Jr. (1992) *Braz. J. Med. Biol. Res.* **25,** 567–582.
10. Spackman, D. H., Stein, W. H., and Moore, S. (1958) *Anal. Chem.* **30,** 1190–1206.
11. Alonzo, N., and Hirs, C. W. W. (1968) *Anal. Biochem.* **23,** 272–288.
12. Bidlingmeyer, B. A., Cohen, S. A., and Tarvin, T. L. (1984) *J. Chromatogr.* **336,** 93–104.
13. Fushitani, K., Matsuura, M. S. A., and Riggs, A. F. (1988) *J. Biol. Chem.* **263,** 6502–6517.
14. Rosa, J. C., Ressiguié, R., Matsuura, M. S. A., and Greene, L. J. (1992) High Performance Liquid Chromatography Identification of the PTH-Amino Acids Using a Step-Gradient Solvent System. XXI Meeting of the Brazilian Society for Biochemistry and Molecular Biology, Caxambu, MG, Abstract P12–P59.
15. Schägger, H., and Von Jagow, G. (1987) *Anal. Biochem.* **166,** 368–379.
16. Ion Exchange Chromatography: Principles and Methods (1994) Pharmacia Biotech, Uppsala, Sweden.
17. Hawke, D., and Yuan, P. (1987) Pyridylethylation on PVDF-Blotted Proteins, ABI User Bulletin 28B, Applied Biosystems Inc., Fullerton, CA.
18. Ambler, R. (1972) *in* Methods in Enzymology (Hirs, C. H. W., and Timasheff, S. N., Eds.), Vol. XXV, pp. 143–154, Academic Press, New York.
19. Beavis, R. C., and Chait, B. T. (1990) *Anal. Chem.* **62,** 1836–1840.
20. Beavis, R. C., Chaudhury, T., and Chait, B. T. (1992) *Mass Spectrom.* **27,** 156–158.
21. Hara-Nishimura, I., Takeuchi, Y., Inoue, K., and Nishimura, M. (1993) *Plant J.* **4,** 793–800.
22. Nong, V. H., Schlesier, B., Bassuner, R., Repik, A., Horstmann, C., and Muntz, K. (1995) *Plant Mol. Biol.* **28,** 61–72.
23. Muren, E., Ek, B., and Rask, L. (1995) *Eur. J. Biochem.* **227,** 316–321.
24. Svendsen, I. B., Nicolova, D., Goshev, I., and Genor, N. (1994) *Int. J. Pept. Protein Res.* **43,** 425–430.
25. Nirasawa, S., Nishino, T., Katahira, M., Uesugi, S., Hu, Z., and Kurihara, Y. (1994) *Eur. J. Biochem.* **223,** 989–995.
26. Devi, L. (1991) *FEBS Lett.* **280,** 189–194.
27. Devi, L. (1991) *in* The Biosynthesis of Peptide Hormones and Neurotransmitters (Fricker, L. D., Ed.), pp. 175–198, CRC Press, Boca Raton, FL.
28. Lindberg, I., and Hutton, J. C. (1991) *in* The Biosynthesis of Peptide Hormones and Neurotransmitters (Fricker, L. D., Ed.), pp. 141–172, CRC Press, Boca Raton, FL.
29. Lord, J. M., and Robinson, C. (1986) *in* Plant Proteolytic Enzymes (Dalling, M. J., Ed.), pp. 69–80, CRC Press, Boca Raton, FL.
30. Roberts, L. M., Lamb, F. I., Pappin, D. J. C., and Lord, J. M. (1985) *J. Biol. Chem.* **260,** 15682–15686.
31. Lord, J. M., Roberts, L. M., and Robertus, J. C. (1994) *FASEB J.* **8,** 201–208.
32. Scott, M. P., Jung, R., Muntz, L., and Nielsen, N. C. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 658–662.