

The *Ricinus communis* 2S albumin precursor: A single preproprotein may be processed into two different heterodimeric storage proteins

Stephen D. Irwin¹, Jeffrey N. Keen², John B.C. Findlay², and J. Michael Lord¹

¹ Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, UK

² Protein Sequencing Unit, Department of Biochemistry, University of Leeds, Leeds, LS2 9JT, UK

Received March 8, 1990

Summary. The *Ricinus communis* (castor bean) 2S albumin is a heterodimer of glutamine-rich, disulphide-linked 4 and 7 kDa polypeptides. A cDNA library was constructed using mRNA from maturing castor bean endosperm as template. Clones containing sequences complementary to albumin mRNA were isolated by hybridization using as a probe a mixture of synthetic oligonucleotides representing sequences predicted for a peptide present in the 2S albumin large subunit. The nucleotide sequence contained an open reading frame encoding a preproprotein of 258 amino acid residues. The preproprotein included both polypeptides of the previously sequenced 2S albumin. In addition, this precursor included two further glutamine-rich sequences which, in terms of their size and conserved cysteine residues typically found in seed proteins of the 2S albumin superfamily, possibly represent the small and large polypeptide subunits of a second heterodimeric storage protein. A post-translational processing scheme is proposed which would result in a single preproprotein generating two distinct heterodimeric 2S albumins. The generation of a second heterodimer seems likely since polypeptide candidates for its small and large subunits were found in the *Ricinus* 2S albumin fraction, and N-terminal protein sequencing confirmed the existence of the putative small subunit.

Key words: *Ricinus* – 2S albumin – Preproprotein processing

Introduction

The low molecular weight, glutamine-rich storage protein (2S albumin) from *Ricinus communis* seeds consists of two different polypeptides joined together by one or more disulphide bonds. The primary sequences of both subunit polypeptides have been determined by Sharief and Li (1982), who noted that the smaller subunit shared

significant homology with the sweet plant protein thaumatin (Iyengar et al. 1979), and the larger subunit showed homology with the Bowman-Birk protease inhibitor from lima beans (Tan and Stevens 1971). As additional seed protein sequences have been determined, it has become clear that sequence homology occurs amongst a growing list of polypeptides. The list includes napin from rape seed (Josefsson et al. 1987; Scofield and Crouch 1987), the methionine-rich Brazil nut 2S albumin (Altenbach et al. 1987), sunflower seed 2S albumin (Allen et al. 1987), and trypsin and/or α -amylase inhibitors from barley (Odani et al. 1983), maize (Mahoney et al. 1984), wheat (Maeda et al. 1983) and millet (Campos and Richardson 1983). Homology is also found between these small, water-soluble seed proteins and regions of the larger, water-insoluble prolamin storage proteins of cereals such as rye γ -secalin (Kreis et al. 1985) and wheat high molecular weight prolamin (Forde et al. 1983).

The castor bean (*Ricinus*) 2S albumin fraction is a mixture of polypeptides. Polyclonal rabbit antibodies raised against the total 2S albumin fraction immunoprecipitated a 34 kDa polypeptide when poly(A)⁺ RNA from castor beans (harvested at the stage of extensive storage protein synthesis) was translated in vitro (Butterworth and Lord 1983; Halpin et al. 1989). Several studies on the biosynthesis of 2S albumins have shown that the final storage protein product is often produced only after extensive processing of considerably larger precursors. For example, the precursor for *Brassica napin* (which is composed of 4 and 9 kDa subunits) is 20 kDa (Crouch et al. 1983; Ericson et al. 1986), that for the Brazil nut methionine-rich protein (3 and 9 kDa subunits) is 18 kDa (Altenbach et al. 1987), and that for a sunflower seed albumin (a 19 kDa monomer) is 38 kDa in size (Allen et al. 1987). In the case of the castor bean 2S albumin, processing would appear to be particularly wasteful if a precursor whose apparent molecular weight is 34 kDa is reduced to a heterodimer consisting of 4 and 7 kDa subunits.

In the present paper we confirm that the large and small polypeptides of the 2S albumin purified and se-

quenced by Sharief and Li (1982) are indeed derived from a considerably larger precursor. Processing to release these polypeptides may not be extravagantly wasteful, however, since the precursor contains two further glutamine-rich regions which retain the conserved cysteine residues of the large and small polypeptides respectively. In this case it seems likely that two distinct glutamine-rich heterodimers are post-translationally released from a single proprotein precursor.

Materials and methods

Plant material. *R. communis* plants were grown from seed in a greenhouse under a 16 h/8 h light/dark cycle. Endosperm tissue was excised from the ripening seeds during testa formation, when storage proteins are being rapidly synthesized (Roberts and Lord 1981a).

Preparation of mRNA and in vitro translation. Total RNA was isolated from 30 g maturing seeds. These were frozen and ground to a fine powder in liquid nitrogen, and RNA was extracted by the phenol-sodium dodecyl sulphate method as previously described (Roberts and Lord 1981b). Extensive washing of the RNA pellet with 3 M sodium acetate pH 5.5 was necessary to remove all traces of ricin. Poly(A)⁺ RNA, prepared by oligo(dT) cellulose chromatography (Aviv and Leder 1972) was translated in either rabbit reticulocyte lysate systems (Amersham International, Amersham, UK) or wheat germ systems prepared as previously described (Anderson et al. 1983). Products were radiolabelled by using 370–740 kBq of [³⁵S]methionine (Amersham; specific activity 37.5 Tbq/mmol) per microgram RNA. Where appropriate 0.2 A₂₆₀ unit of canine pancreatic microsomes was included. The 2S albumin precursor was identified where appropriate by immunoprecipitation using rabbit antibodies raised against purified 2S albumin (a kind gift from Dr. Steven Li) as described previously (Roberts and Lord 1981b). Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and were visualized by fluorography.

cDNA synthesis. 10 µg of poly(A)⁺ RNA was used as a template for oligo(dT)-primed synthesis of cDNA using avian myoblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, Fla) followed by second strand synthesis with DNA polymerase (Amersham) according to the method of Lehrach et al. (1979). After treatment of the double-stranded cDNA with S1 nuclease (Sigma Chemical Co., Poole, UK), it was blunt-ended with the Klenow fragment of *Escherichia coli* DNA polymerase I (Amersham). *Eco*RI restriction sites were protected by treatment with *Eco*RI methylase (New England Biolabs., Beverly, Mass) in the presence of S-adenosyl methionine. Excess *Eco*RI linkers (Amersham) were added to the ends of the cDNA strands using T4 DNA ligase (Amersham), followed by digestion with excess *Eco*RI to generate cohesive ends. cDNA was separated from unbound linkers on a Biogel P-60 column, and run on a low melting point agarose gel where fractions were excised at

100 bp intervals. cDNA was extracted from the gel and lyophilised. Fractions ranging from 0.8 to 1.2 kb were ligated into pUC19 that had previously been linearised with *Eco*RI and treated with calf intestinal alkaline phosphatase (Boehringer Mannheim, FRG). The recombinant plasmids were used to transform competent *E. coli* DH5α cells.

Isolation and screening of transformants. Transformants were selected by their ability to grow on medium containing ampicillin. Recombinants were selected by their inability to utilise as substrate the chromogenic analogue of lactose, 5-bromo-4-chloro-3-indolyl-D-galactopyranoside in the presence of isopropylthiogalactoside. Recombinant colonies were picked into microtitre wells and grown in selective medium overnight. Colonies were inoculated onto nylon filters (Amersham) overlaid on selective medium and were allowed to grow overnight whereupon they were denatured and nucleic acids fixed by ultraviolet irradiation as described by the supplier of the filters.

Filters were screened with a synthetic 23-mer oligonucleotide mixture. Oligonucleotides were synthesized using an Applied Biosystems 380 B DNA synthesizer and purified by HPLC. The probe represented 32 DNA sequences predicted from part of the amino acid sequence of the large subunit of the 2S albumin (Sharief and Li 1982).

Amino acid sequence	Cys	Asp	His	Leu	Lys	Gln	Met	Gln
Nucleotide sequence	TGT	GAT	CAT	CTA	AAA	CAA	ATG	CA
	C	C	C		G	G		

To restrict the mixture to a maximum of 32 different oligonucleotides, leucine was represented by just one codon (CTA), thus the probe may contain two mismatches. Screening and washing temperatures were adjusted accordingly. The probe was 5'-labelled with γ-[³²P]ATP using T4 polynucleotide kinase (Amersham). Filters were prehybridized, hybridized at 37° C, and successively washed and monitored at 37°, 45° and 55° C. Plasmid DNA was prepared (Holmes and Quigley 1981) from those colonies that hybridized to the probe. Southern blots of inserts excised with *Eco*RI were probed with the oligonucleotide mixture. Of 46 inserts that hybridized to the probe, 3 were chosen for sequencing.

Isolation of genomic clones. A *Ricinus* genomic library (kindly provided by J. Tregear of this laboratory), cloned in the λ vector Charon 35 after partial digestion with *Sau*3A, was screened using cDNA clone 14g4 as a probe. Of 5 × 10⁵ plaques probed, 40 gave strong positive hybridization signals. Restriction analysis indicated that three different 2S albumin genes had been cloned, one of which was sequenced (Irwin and Lord 1990).

DNA sequencing. cDNA clones 8g8, 10a12 and 14g4 showed the same restriction digest maps with the enzymes *Sac*I, *Pvu*II and *Sph*I. They were subcloned into M13mp19 and various restriction fragments further subcloned into M13mp18 and M13mp19 (see Fig. 1). These

were sequenced using the dideoxynucleotide chain termination technique (Sanger et al. 1977). The complete sequences of clones 8g8 and 14g4 were obtained but only the 5' end of 10a12 was investigated.

In vitro transcription. cDNAs were subcloned into the vector pGEM3blue (Promega Biotech, Madison, Wis) and transcribed with T7 RNA polymerase (Amersham) (Melton et al. 1984). Transcripts were translated in vitro as described above.

Protein sequencing. Protein was subjected to N-terminal amino acid sequence analysis according to the procedure of Findlay et al. (1989). Lyophilized protein was dissolved in 35 μ l of 0.1 M NaHCO₃, 0.1 M Na₂CO₃, 0.25% SDS, pH 10 and added to 10 mg of DITC-glass (17 mm pore size, 200–400 mesh). The glass was incubated for 60 min at 56° C under N₂, then washed with water and methanol to remove non-covalently bound material. The glass-coupled peptide was then sequenced by automated solid-phase Edman degradation (Laursen 1971).

Results

Characterization of 2S albumin precursor cDNA and genomic clones

A cDNA library was constructed using mRNA from maturing castor bean endosperm as template. The library was screened by hybridization with a mixture of synthetic oligonucleotides (23-mers) representing a DNA sequence encoding an 8 amino acid peptide present in the large subunit of the 2S albumin sequenced by Sharief and Li (1982). Several positive clones were identified and the nucleotide sequence of one of these, designated 14g4, was determined by the dideoxy method. The nucleotide sequence (Fig. 1) contained an open reading frame which defined a 258 residue polypeptide of molecular weight 29 329 daltons. The precursor contained the large and small polypeptide sequences of the previously sequenced 2S albumin. Residues 136–169 defined the small subunit (molecular weight 3957 daltons), matching exactly the published primary sequence. This was separated by a 3 residue linker from the large subunit (residues 173–237, molecular weight 7349 daltons), which differed slightly from the published sequence. In the sequence derived here, residue 201 is Glu (rather than Gln in the published sequence), residue 213 is Asp (rather than Asn) and residue 234 is Glu (rather than Gln). More significantly, the sequence derived here is 4 residues longer than the published large subunit sequence. This discrepancy is readily explained. In the published sequence, Sharief and Li (1982) found both Ser and Leu at position 33 (residue 205 in Fig. 1 here) followed by Gln-Gly-Gln at the carboxy-terminus of a fragment generated during sequencing. As shown here the sequence over this region is actually Ser-Gln-Gly-Gln-Leu-Gln-Gly-Gln (residues 205–212, Fig. 1).

We also sequenced a second insert, 8g8. Apart from

the putative signal sequence region, this clone differed in only a single nucleotide within the proprotein coding region (which resulted in the conversion of Thr-53 into Pro), but it had a shorter 3' untranslated region than 14g4 (Fig. 1). We tentatively identified the first 21 residues of 14g4 as the signal sequence. This identification is based on the general features, including size (2246 daltons) and hydrophobicity, of this sequence and the Ala residues at positions –1 and –3 which constitute a typical signal peptidase cleavage site (von Heijne 1984). As shown in Fig. 2 the incomplete signal sequence of 8g8 shows several residue differences when compared with 14g4. We sequenced the 5' region of a further insert (designated 10a12), and compared its putative signal sequence with the other castor bean sequences. This sequence is slightly shorter than that of 14g4, with which it has considerable homology, including the putative signal peptidase site (Fig. 2). The three castor bean 2S albumin signal sequences derived here share limited homology but have the same general features as other 2S albumin signal sequences (Fig. 2).

The clones 8g8, 14g4 and 10a12 were subcloned into pGEM3blue and transcribed with T7 RNA polymerase. All produced transcripts of the expected size. The 8g8 transcript, which did not encode the initiator Met (Fig. 2), did not generate any product in a wheat germ system. In contrast, the 10a12 and 14g4 transcripts, which did encode the initiator Met, each produced a single polypeptide of the apparent molecular weight previously observed (Fig. 3, lanes 1 and 3) which was immunoprecipitated by antibodies raised against the castor bean 2S albumin (Fig. 3, lanes 2 and 4). Furthermore, the translation products were N-terminally processed and segregated when the 14g4 and 10a12 transcripts were translated in reticulocyte lysates in the presence of dog pancreatic microsomes (data not shown).

Genomic Southern blots probed with 14g4 indicated that the castor bean 2S albumins are encoded by a small gene family comprising at least four homologous sequences (Fig. 4a). A developmental Northern RNA blot showed that the 2S albumin precursor gene was not transcribed in the ripening seeds at developmental stages (Roberts and Lord 1981a) prior to testa formation (stages A to C, Fig. 4b), but significant transcriptional activity was observed during testa formation (stage D), after testa formation (stage E) and during desiccation (stage F) (Fig. 4b).

The genomic sequence corresponding to the region of DNA encoding the preproalbumin gene was obtained by screening a *Ricinus* genomic library with the cDNA clone 14g4. The precursor coding sequence with 5' and 3' flanking regions has been deposited in the database (accession no. Irwin and Lord 1990) and is not reproduced here. A comparison of the region immediately upstream of the putative TATA box of the *Ricinus* albumin gene with the corresponding regions upstream of the TATA boxes of the *Brassica* napin (Josefsson et al. 1987) and *Arabidopsis* 2S albumin (Krebbers et al. 1988) genes is shown in Fig. 5. These regions of homology, which include the CATGCAT box (Dickinson et al. 1988), which is not entirely conserved in all three se-

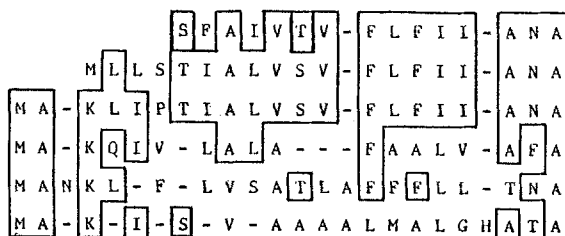


Fig. 2. Comparison of the putative *Ricinus* albumin precursor signal sequence with those of other plant albumins. The incomplete signal sequence of 8g8 and the complete sequences of 10a12 and 14g4 are compared. Sequences are aligned to maximise homology, shown in *boxed areas*. Sunflower albumin sequenced by Allen et al. (1987), *Brassica napin* by Crouch et al. (1983), and Brazil nut albumin by Altenbach et al. (1987).

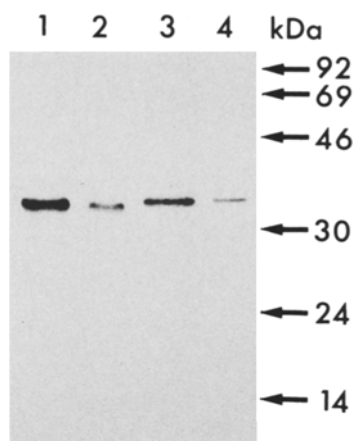


Fig. 3. Transcription and translation of *Ricinus* albumin precursor clones. Clones 10a12 and 14g4 were subcloned into a transcription vector, transcribed in vitro and the transcripts were translated in a wheat germ cell-free system. Translational products were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) before (lanes 1 and 3) or after (lanes 2 and 4) immunoprecipitation with anti-2S albumin antibodies. Lanes 1 and 2, 10a12; lanes 3 and 4, 14g4

quences, and an incomplete form of this motif (CATGCA), which is, may confer temporal and tissue-specific expression on these genes. The *Ricinus* gene encoding ricin, which shows developmental and tissue specific expression identical to the 2 S albumin gene, also contains the CATGCAT sequence in its 5' flanking region (J. Tregear, personal communication).

Processing the 2S albumin precursor could generate two glutamine-rich heterodimers

While the data presented here clearly confirm that the glutamine-rich cysteine-containing small and large subunits of the castor bean 2S albumin (Sharief and Li 1982) are derived from a preproprotein precursor, processing of a 30 kDa precursor to yield an 11 kDa heterodimer seems particularly wasteful. Examination of the precursor sequence revealed two further glutamine-rich sequences having the most conspicuous conserved features of typical small and large 2S albumin subunits respectively (indicated by solid underline, Fig. 1). The putative small subunit (residues 15–51, molecular weight 4349 daltons, Fig. 2) shares homology with the other castor bean 2S albumin small subunit and other seed

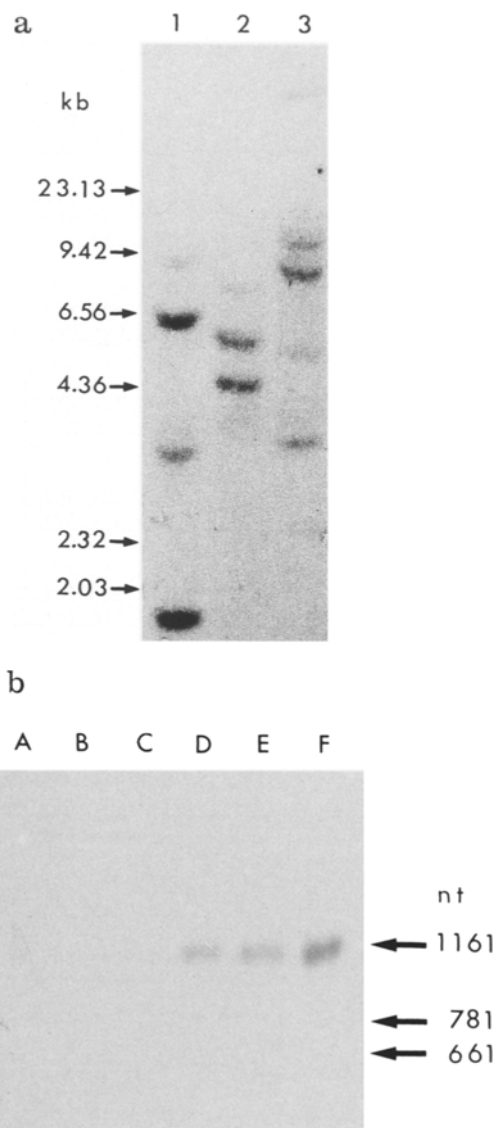


Fig. 4. a Genomic restriction fragments hybridizing to 2S albumin precursor cDNA. A genomic Southern blot of *Ricinus* DNA digested with 1, *Bam*HI; 2, *Eco*RI; 3, *Hind*III, probed with nick-translated 14g4. **b** Northern blotting and hybridization of *Ricinus* RNA to 14g4 cDNA. RNA was recovered from *Ricinus* endosperm tissue at various developmental stages [stages A to F, Roberts and Lord (1981a)], and separated on a denaturing agarose gel (10 µg RNA/lane). After transfer to nylon filters the immobilized RNA was hybridized to nick-translated 14g4. *Arrows* indicate positions of molecular weight markers

1 ATGCCATGCCAGAA CATTAgcTACAc gTTACACaTag CATGCAG CGcC
 |||||
 2 ATGCCAGAA ATgAcTgGCCATTcCATACAT CTTACACGTtctCGaCaTGCAc TCCCTTaaATtctCCTATAA
 |||||
 3 cacaCATACAATcACAtGC gTGCATGCATTaTTACACGTGATCGCCATGCCAAaTCTCCTTTATagCCTATAA
 |||||
 4 TACAA ACaAaCataTGCATGtATTCTTACACGTGAaCtCCATGCCAAgTCTCtTTTcTcaCCTATAA

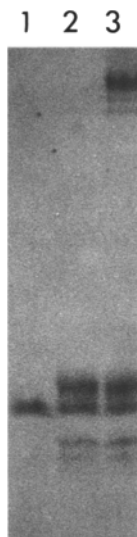


Fig. 7. Water-soluble components of *Ricinus* protein bodies, lane 1, purified 2S albumin (Sharief and Li 1982); lane 2, total matrix proteins after removing the 7S lectins; lane 3, total matrix proteins. Proteins were separated by SDS-PAGE and stained with Coomassie blue

Known small subunit

Deduced: PSQGGCRGQIQEQQLRQC

Determined: ?SQQG?RGQIQEQQLRQ?

Putative small subunit

Deduced: ESKGEREGSS

Determined: ?S?GEREGSS

Fig. 8. N-terminal sequence analysis identifying the 2S albumin small subunits: the previously characterized small subunit (Sharief and Li 1982) and the second small subunit

of the latter two bands was faint, possibly due to losses during the staining procedures. The previously characterized *Ricinus* 2S albumin, purified by the method of Sharief and Li (1982) as modified by McGurl (1987), was also run on the gel. The lower band of the upper doublet corresponded to the previously characterized large subunit and the lower band of the lower doublet corresponded to the characterized small subunit (Fig. 7). The upper band of each doublet might therefore represent the large and small subunits of the second putative 2S albumin. All four bands were eluted from the gel and subjected to N-terminal microsequencing. N-terminal sequence could not be obtained for either the known or the putative large subunits. It had been found earlier that the large subunit was N-terminally blocked as a result of cyclization of the N-terminal glutamine residue (Sharief and Li 1982). Since the putative second large subunit is also believed to have an N-terminal glutamine (Gln-66, Fig. 1) it should also be blocked. However, N-terminal sequence was obtained which could be attributed to the smaller subunits. The identity of the previously characterized small subunit was confirmed (Fig. 8). More interestingly, an N-terminal sequence was

also determined in which eight out of the first ten residues identified fitted exactly the proposed N-terminal sequence of the putative small subunit (Fig. 8).

Discussion

The present study has confirmed that the previously characterized *R. communis* 2S albumin, a heterodimer of disulphide-linked 4 and 7 kDa polypeptides (Sharief and Li 1982) is derived by co- and post-translational processing of a 30 kDa preproprotein precursor. Even after cleavage of the N-terminal signal sequence (21 amino acids) the glutamine-rich 2S albumin (99 amino acids) accounts for little more than 40% of the precursor (237 amino acids). If post-translational processing resulted in the degradative loss of approximately 60% of a newly synthesized protein, it would not simply be inefficient, but would be extremely wasteful because of the abundance of the 2S albumin [which accounts for around 20% of the total *Ricinus* seed protein (Youle and Huang 1978). Examination of the precursor sequence revealed two further peptide sequences which might represent the small and large subunits of a second, glutamine-rich 2S albumin (underlined in Fig. 1). The putative small and large subunits contain the conserved cysteine residues also found in the known small and large subunits and which characterize this family of proteins (Fig. 6). We propose that post-translational processing of the *Ricinus* 2S albumin precursor generates two distinct heterodimeric 2S albumins. The total *Ricinus* 2S albumin fraction contains polypeptides which are candidates for the small and large subunits of a second heterodimer (Fig. 7). These polypeptides are distinct from the small and large subunits of the characterized 2S albumin (Sharief and Li 1982), but appear to occur in equimolar amounts. While both large subunits are apparently N-terminally blocked, microsequencing provided clear evidence for the presence of the known and the putative small subunits (Fig. 8).

Our definition of the putative small and large subunits is based on a speculative assignment of processing sites. The subunits of both the putative and the known heterodimers are separated by a linking peptide whose first residue is serine and whose last residue is asparagine. Previous work in our laboratory has shown that the A and B chains of the castor bean lectins ricin and *R. communis* agglutinin are cleaved from a proprotein precursor (Lamb et al. 1985; Roberts et al. 1985). In this case the lectin A and B subunits are separated from each other by a 12 amino acid linker peptide whose first residue is serine and whose last residue is asparagine. It is clear that many seed proteins are released from precursors by a protein body endoprotease that cleaves on the carboxyl side of asparagine residues (Roberts et al. 1985; Lord and Robinson 1986). In some instances a single cleavage occurs after an asparagine residue (Boulter 1984; Higgins 1984) and in others a linking peptide is removed (Higgins et al. 1983; Bowles et al. 1986).

During the biosynthesis of 2 S albumins in develop-

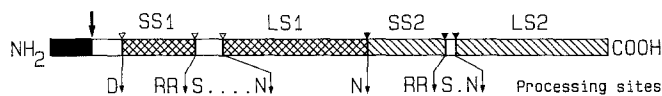


Fig. 9. Schematic representation of the proposed *Ricinus* prealbumin structure. The arrow indicates the proposed signal peptidase cleavage site, the closed triangles indicate known post-translational processing sites and the open triangles indicate further putative post-translational processing sites. SS1 and LS1 refer to the small and large subunits of the putative albumin. SS2 and LS2 refer to the albumin protein sequenced by Sharief and Li (1982)

ing *Ricinus* seeds, the N-terminal signal sequence is removed as the proprotein is segregated into the lumen of the endoplasmic reticulum (ER) (Lord 1985). From the ER the preproprotein is transported, via the Golgi and a population of Golgi-derived vesicles, to the protein bodies where the heterodimeric storage proteins are finally cleaved from the proprotein (Lord 1985). In our proposed processing scheme, the asparagine endoprotease, and, possibly, other protein body processing enzymes, would remove the -Ser-Asp-Asn- (residues 170–172, Fig. 1) linker which joins the small and large subunits of the 2S albumin sequenced by Sharief and Li (1982), and the proposed linker (residues 52–65, Fig. 1) which joins the small and large subunits of the putative second 2S albumin. Cleavage after Asn-135 would separate the large subunit of one 2S albumin from the small subunit of the other. The processing site for the post-translational removal of the N-terminal peptide from the proprotein is proposed to occur after Asp-14 (Fig. 1).

The structure of the preproprotein precursor and its proposed processing sites are illustrated schematically in Fig. 9. The proposed processing events would generate two heterodimeric storage proteins from the precursor. One heterodimer, previously purified and sequenced, has an overall molecular weight of 11 306 daltons and contains a total of 8 cysteine and 23 glutamine residues, while the second heterodimer has an overall molecular weight of 12 390 daltons and contains a total of 8 cysteine and 21 glutamine residues. The portion of the precursor removed during processing (a total of 52 amino acids derived from the signal sequence, the post-translationally removed N-terminal peptide and the linker peptides) contains no cysteine or glutamine residues (Fig. 1).

Other studies have shown that multiple copies of the *Euglena* nuclear-encoded polypeptides of the light harvesting complex (Houlne and Scharzt 1988) or the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Chan et al. 1990) are processed from a single large polypeptide precursor in each case. The present report extends this complex precursor processing pattern to higher plant storage proteins.

Acknowledgements. This work was supported by the United Kingdom Science and Engineering Research Council via Grant GR/D 74185 (to JML) and a research studentship (to SDI). We thank J. Tregear (University of Warwick) for providing the *Ricinus* genomic library, and Dr. Steven Li (National Institutes of Health) for providing purified *Ricinus* 2S albumin.

References

- Allen RD, Cohen EA, Vonder Haar RA, Adams CA, Ma DP, Nessler CL, Thomas TL (1987) Sequence and expression of a gene encoding an albumin storage protein in sunflower. *Mol Gen Genet* 210:211–218
- Altenbach SB, Pearson KW, Leung FW, Sun SSM (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein extremely rich in methionine. *Plant Mol Biol* 8:239–250
- Anderson CW, Straus JW, Dudock BS (1983) Preparation of a cell-free protein-synthesizing system from wheat germ. *Methods Enzymol* 101:635–644
- Aviv H, Leder P (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid-cellulose. *Proc Natl Acad Sci USA* 69:1408–1412
- Boulter D (1984) Cloning of pea storage protein genes. *Philos Trans R Soc Lond [Biol]* 304:323–332
- Bowles DJ, Marcus SE, Pappin DJC, Findlay JBC, Eliopolous E, Maycox PR, Burgess J (1986) Posttranslational processing of concanavalin A precursors in jackbean cotyledons. *J Cell Biol* 102:1284–1297
- Butterworth AG, Lord JM (1983) Ricin and *Ricinus communis* agglutinin subunits are all derived from single sized precursor polypeptides. *Eur J Biochem* 137:57–65
- Campos FAP, Richardson M (1983) The complete amino acid sequence of the bifunctional α -amylase/trypsin inhibitor from seeds of ragi (Indian finger millet, *Eleusine coracana*). *FEBS Lett* 152:300–304
- Chan RL, Keller M, Canaday J, Weil J-H, Imbault P (1990) Eight small subunits of *Euglena* ribulose-1,5-bisphosphate carboxylase/oxygenase are translated from a large mRNA as a polypeptide. *EMBO J* 9:333–338
- Crouch ML, Tenbarger KM, Simon AE, Ferl R (1983) 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* 2:273–283
- Dickinson CD, Evans RP, Nielson NC (1988) RY repeats are conserved in the 5' flanking regions of legume seed protein genes. *Nucleic Acids Res* 16:371
- Ericson ML, Rodin J, Lenman M, Glimelius K, Josefsson L-G, Rask L (1986) Structure of the rape seed 1.7S storage protein, napin, and its precursor. *J Biol Chem* 261:14576–14581
- Findlay JBC, Pappin DJC, Keen JN (1989) Automated solid-phase microsequencing. In: Findlay JBC, Geisow M (eds) *Protein sequencing – A practical approach*. IRL Press, Oxford, pp 69–84
- Forde J, Forde BG, Fry RP, Kreis M, Shewry PR, Mifflin BJ (1983) Identification of barley and wheat cDNA clones related to the high M_r polypeptides of wheat gluten. *FEBS Lett* 162:360–366
- Halpin C, Conder MJ, Lord JM (1989) Different routes for integral protein insertion into *Ricinus communis* protein body and glyoxysome membranes. *Planta* 179:331–339
- Higgins TJV (1984) Synthesis and regulation of major proteins in seeds. *Annu Rev Plant Physiol* 35:191–221
- Higgins TJV, Chandler PM, Zurawski G, Button S, Spencer D (1983) The biosynthesis and primary structure of pea seed lectin. *J Biol Chem* 258:9544–9549
- Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* 114:193–197
- Houlne G, Scharzt R (1988) Characterization of cDNA sequences for LHCl apoproteins in *Euglena gracilis*: the mRNA encodes a large precursor containing several consecutive divergent polypeptides. *Mol Gen Genet* 213:479–486
- Irwin SD, Lord JM (1990) Nucleotide sequence of a *Ricinus communis* 2S albumin precursor gene. *Nucleic Acids Res*
- Iyengar RB, Smits P, van der Ouderaa F, van der Well H, van Browsershaven J, Ravestein P, Richters G, van Wassenaar PD (1979) The complete amino acid sequence of the sweet protein thaumatin 1. *Eur J Biochem* 96:193–204
- Josefsson L-G, Lenman M, Ericson ML, Rask L (1987) Structure

- of a gene encoding the 1.7S storage protein, napin, from *Brassica napus*. J Biol Chem 262:12196–12201
- Krebbbers E, Herdies L, de Clerq A, Seurinck J, Leemans J, van Damme J, Segura M, Gheysen G, van Montagu M, Vanderkerckhove J (1988) Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. Plant Physiol 87:859–866
- Kreis M, Shewry PR, Forde BG, Forde J, Mifflin BJ (1985) Structure and evolution of seed storage proteins and their genes, with particular reference to wheat, barley and rye. Oxford Surv Plant Mol Cell Biol 2:253–317
- Lamb FI, Roberts LM, Lord JM (1985) Nucleotide sequence of cloned cDNA coding for preprorizin. Eur J Biochem 148:265–270
- Laursen RA (1971) Solid phase Edman degradation. An automatic peptide sequencer. Eur J Biochem 20:89–102
- Lehrach H, Frischauf AM, Hanahan D, Wozney J, Fuller F, Boedtker H (1979) Construction and characterization of pro α 1 collagen complementary deoxyribonucleic acid clones. Biochemistry 18:3146–3152
- Lord JM (1985) Synthesis and intracellular transport of lectin and storage protein precursors in endosperm from castor bean. Eur J Biochem 146:403–409
- Lord JM, Robinson C (1986) The role of proteolytic enzymes in the posttranslational modification of proteins. In: Dalling MJ (ed) Plant proteolytic enzymes. CRC Press, Boca Raton, Fla, pp 69–80
- Maeda K, Hase T, Matsubara H (1983) Complete amino acid sequence of an α -amylase inhibitor in wheat kernel. Biochim Biophys Acta 743:52–57
- Mahoney WC, Hermodson MA, Jones B, Powers DD, Corfman RS, Reech GR (1984) Amino acid sequence and secondary structure analysis of the corn inhibitor of trypsin and activated Hageman factor. J Biol Chem 259:8412–8416
- McGurl B (1987) Purification and characterization of putative protein allergens from the seeds of the castor oil plant, *Ricinus communis*. PhD Thesis, University of Warwick, UK
- Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res 12:7035–7056
- Odani S, Koide T, Ono T (1983) The complete amino acid sequence of barley trypsin inhibitor. J Biol Chem 258:7998–8003
- Roberts LM, Lord JM (1981a) Protein biosynthetic capacity in the endosperm tissue of ripening castor bean seeds. Planta 152:420–427
- Roberts LM, Lord JM (1981b) The synthesis of *Ricinus communis* agglutinin. Cotranslational and posttranslational modification of agglutinin polypeptides. Eur J Biochem 119:31–41
- Roberts LM, Lamb FI, Pappin DJC, Lord JM (1985) The primary sequence of *Ricinus communis* agglutinin. Comparison with ricin. J Biol Chem 260:15682–15686
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
- Scofield SR, Crouch ML (1987) Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*. J Biol Chem 262:12202–12208
- Sharief FS, Li SS-L (1982) Amino acid sequence of the small and large subunits of seed storage proteins from *Ricinus communis*. J Biol Chem 257:14753–14759
- Tan CGL, Stevens FC (1971) Amino acid sequence of lima bean protease inhibitor component IV. 2. Isolation, sequence determination of the chymotryptic peptides and the complete amino acid sequence. Eur J Biochem 18:515–523
- Von Heijne G (1984) How signal sequences maintain cleavage specificity. J Mol Biol 173:243–251
- Youle RJ, Huang AHC (1978) Albumin storage proteins in the protein bodies of castor bean. Plant Physiol 61:13–16

Communicated by J. Schell