Plant Molecular Biology **15**: 879–893, 1990.

Biosynthesis, cDNA and amino acid sequences of a precursor of conglutin δ, a sulphur-rich protein from *Lupinus angustifolius*

Kenwyn R. Gayler, Sotirios Kolivas, Alison J. Macfarlane, Glenn G. Lilley[1], Mauro Baldi, Robert J. Blagrove[1] and Elizabeth D. Johnson
Russell Grimwade School of Biochemistry, Parkville, VIC 3052, Australia; [1]*CSIRO, Division of Biomolecular Engineering, Parkville, VIC 3052, Australia*

Received 2 April 1990; Accepted 3 September 1990

Key words: Biosynthesis, DNA sequence, *Lupinus angustifolius*, protein sequence, plant cDNA, 2S storage protein

Abstract

The biosynthesis of conglutin δ has been studied in developing cotyledons of *Lupinus angustifolius* L. Precursors of conglutin δ formed the major sink for [^{35}S]-cysteine incorporated by developing lupin cotyledons, and these precursors were rapidly sequestered into the endoplasmic reticulum. The sequence of a cDNA clone coding for one such precursor of conglutin δ was determined. The structure of the precursor polypeptide for conglutin δ predicted from the cDNA sequence contained an N-terminal leader peptide of 22 amino acids directly preceding a subunit polypeptide of M_r 4520, together with a linking region of 13 amino acids and a subunit polypeptide of M_r 9558 at the C-terminus. The amino acid sequence predicted from the cDNA sequence showed minor variations from that established by sequencing of the protein purified from mature dried seeds (Lilley and Inglis, 1986). These were consistent with the existence of a multi-gene family coding for conglutin δ. Comparison of the sequences of conglutin δ with those of other 2S storage proteins showed that the cysteines involved in internal disulphide bridges between the mature subunits of conglutin δ, were maintained throughout this family of proteins but that little else was conserved either at the protein or DNA level.

Introduction

Conglutin δ is a protein present in seeds of *Lupinus angustifolius* L. in sufficient abundance to make a major contribution to the nutritional value of these seeds. Although originally described in 1955 as a sulphur-rich 2S globulin from *L. luteus* [14, 21, 22], it has only recently been purified and characterized [29, 30]. Conglutin δ is related to the 2S super-family of storage proteins [25], which is widespread in seeds of dicotyledonous plants including *Brassica* [33], *Ricinus* [28], *Momordica* [27], *Pisum* [12], *Arabidopsis* [24] and *Helianthus* [31]. The family is characterized by low molecular weight proteins which contain relatively high levels of cyst(e)ine and glutamine.

In this paper we report the results of a study of the biosynthesis and secretion of radiolabelled conglutin δ in intact cotyledons, using a chemically based method to overcome the difficulties inherent in attempts to detect 2S storage proteins with conventional immunological procedures [13,

19]. The results show that conglutin δ is synthesized initially as a family of closely related precursors which form a major sink for [^{35}S]-cysteine in developing cotyledons and are sequestered into the endoplasmic reticulum.

The isolation and sequencing of a cDNA corresponding to conglutin δ is also reported, and consequently the amino acid sequence for one precursor of conglutin δ. We have used the information available for the amino acid sequence of the mature form of conglutin δ [32] to deduce the extent of the post-translational modifications needed to process this protein into its functional storage form and to highlight those structural domains which are conserved amongst related 2S storage proteins.

Materials and methods

Plant material

Lupinus angustifolius L. cv. Unicrop was grown as described previously [13]. Pods to be used in labelling studies were removed from the plants approximately 33 days after flowering (DAF).

Isolation of lupin mRNA and preparation of cDNA clones

Membrane-bound polysomal RNA from cotyledons aged from 25 to 30 DAF was purified by passage through an oligo (dT) cellulose column as described by Johnson *et al.* [20]. Double-stranded complementary DNA copies were prepared from 5 μg of this poly(A)$^+$ RNA with a cDNA synthesis system (Amersham International) essentially by the method of Gubler and Hoffman [16]. The double-stranded cDNA was cloned into the *Eco*RI site in λgt11 phage after attachment of *Eco*RI linkers, the phage was packaged and used to infect *E. coli* K12 strain Y1090.

Plaques with recombinant DNAs were identified by hybridization to ^{32}P-labelled probe DNA according to the method of Grunstein and Hogness [15]. Synthetic oligonucleotide mix was end-labelled using [γ-^{32}P]-dATP and T4 polynucleotide kinase. Plaques bound to Hybond-N (Amersham Inc.) were incubated for 16 h at 30 °C in 6× SSC, 0.5% N-lauryl sarcosine and 20 mg/ml salmon sperm DNA and hybridized overnight with 0.1 μg of ^{32}P-labelled oligonucleotide (6×10^8 cpm/mg) in the same buffer. Following hybridization, filters were washed twice for 10 min in 2× SSC at room temperature, at 37 °C and at 42 °C and autoradiographed.

Sequencing

The DNA sequence was determined by the dideoxy method [35] using M13 mp18 and mp19 as vectors and T7 DNA polymerase (Sequenase; United States Biochemicals) [39].

Extraction and purification of conglutin δ

Conventional immunological procedures could not be used for specific identification of the radiolabelled precursors of conglutin δ as this protein failed to produce antibodies when injected into mice, chickens or rabbits. Instead, reversed-phase chromatography on Sep-pak cartridges was used to separate conglutin δ from other storage proteins in extracts of developing seeds. Proteins were extracted from cotyledons by homogenization in 5 vol of 0.1 M Tris-HCl buffer, pH 7.5, containing 10% NaCl and 0.5 mM N-ethylmaleimide (NEM) to block accessible sulphydryl groups. The suspension was stirred for 2 h at room temperature and insoluble material removed by centrifugation. Chromatography on C18 reversed-phase Sep-pak cartridges (Waters-Millipore, USA) was carried out as previously described [13]. NEM-blocked conglutin δ was selectively bound to the Sep-pak cartridge which had been activated with acetonitrile and equilibrated with water. Unbound material was washed through with distilled water, and the conglutin δ eluted with a solution of 42% (v/v) acetonitrile in 0.1% trifluoroacetic acid.

S-Carboxymethylation

NEM-conglutin δ was reduced and denatured in 7.5 M guanidine hydrochloride in 1.25 M Tris-HCl buffer, pH 8.2 in the presence of 2 mM EDTA and 30 mM DTT for 3 h at 37 °C, alkylated for 20 min in 40 mM iodoacetic acid in the same buffer, and then reacted with 1.6 mM 2-mercaptoethanol and dialysed.

The amino acid composition of S-carboxymethylated conglutin δ was determined by hydrolysis in 6M HCl at 110 °C in a sealed evacuated tube for 24 h, followed by separation on a Beckman 121MB Amino Acid Analyzer.

Gel filtration chromatography

Samples containing 75 mg NEM-blocked and S-carboxymethylated conglutin δ were passed through a 10 × 300 mm Superose 12 Pharmacia FPLC column in 50 mM sodium phosphate buffer, pH 7.5, containing 1% (w/v) NaCl and 0.02% (w/v) sodium azide. The flow rate was 0.1 ml min^{-1} and 0.25 ml fractions were collected. Molecular weights of radioactively labelled polypeptides were estimated from their elution volumes after the column was calibrated using authentic samples of NEM-blocked native conglutin δ and its S-carboxymethylated derivatives. These molecular weights had been previously determined from their respective amino acid sequences [32].

Labelling of proteins in detached cotyledons

Mature conglutin δ contains 8.5 residues percent of cyst(e)ine in contrast to the other lupin storage proteins which contain a maximum of 2.9 residues percent [3]. Cotyledons were therefore incubated with a short pulse of [^{35}S]-cysteine to preferentially label conglutin δ-related proteins. Cotyledons were placed on 20 μl drops containing 0.37 MBq of L-[^{35}S]cysteine and incubated as previously described [13, 38]. In pulse-chase experiments, cotyledons were rinsed with sterile water and the incubation continued on 20 μl of sterile water or 1 mM unlabelled cysteine for the duration of the chase. Following the incubation, a 1 mm slice was taken from the face of the cotyledon and either used directly for extraction of proteins or subjected immediately to subcellular fractionation.

Preparation of endoplasmic reticulum

Slices from 8 cotyledons pre-labelled with ^{35}S-cysteine were rinsed in distilled water and the endoplasmic reticulum fraction isolated according to the method of Chrispeels *et al.* [6]. Fractions were localized by NADH: cytochrome c reductase activity [5]. The membrane fraction prepared in this way from *L. angustifolius* was shown previously to contain endoplasmic reticulum and golgi but to be free from protein bodies and soluble proteins [20].

Proteins were extracted from this fraction by the addition of 2 volumes of 10 mM Tris-HCl buffer, pH 7.5 containing 10% (w/v) NaCl and 0.5 mM N-ethylmaleimide and incubation for 2 h at room temperature.

Electrophoresis and fluorography

SDS gel electrophoresis of the radiolabelled proteins on 8 to 20% gradient polyacrylamide gels (SDS-PAGE) and visualization by fluorography was performed as previously described [20].

Reversed-phase HPLC

Purified polypeptides of NEM-blocked native and S-carboxymethylated conglutin δ were analysed by reversed-phase high-performance liquid chromatography (RP-HPLC) on a Vydac 218TP54 column using a gradient of 0% to 42% acetonitrile in 0.1% trifluoroacetic acid. The effluent was monitored at 214 nm and the radioactivity in the collected fractions measured directly by liquid scintillation counting. Authentic

samples of conglutin δ purified from mature seeds of *L. angustifolius* cv. Uniwhite were used as standards.

Results

Detection of precursors of conglutin δ

The separation of conglutin δ from other storage proteins by reversed-phase chromatography was first verified. Proteins were extracted from juvenile cotyledons and fractionated by chromatography on C18 Sep-pak cartridges as described in the methods. The material retained by the Sep-paks and eluted with 42% acetonitrile in 0.1% trifluoroacetic acid showed the following properties. Firstly, the majority of protein in this bound fraction had the same mobility as authentic samples of mature NEM-conglutin δ, when analysed by SDS-PAGE under reducing conditions (Fig. 1). Secondly, this bound material had the distinctive amino acid composition characteristic of conglutin δ; that is, high levels of cyst(e)ine, glutamine and glutamate and a low level of methionine (Table 1). Thirdly, the bound material behaved similarly to authentic conglutin δ during

Table 1. Amino acid composition of fraction isolated on Sep-pak cartridges.

Amino acid	Residues percent	
	Bound fraction	Conglutin δ[1]
Lysine	2.9	2.6
Histidine	2.5	2.4
Arginine	9.3	9.1
Aspartate	7.5	9.0
Threonine	1.6	0.2
Serine	8.2	8.0
Glutamate	35.1	38.0
Proline	4.0	3.9
Glycine	6.0	3.7
Alanine	2.7	1.2
SCM-Cysteine	5.3	7.8
Valine	1.9	1.3
Methionine	0.2	0.0
Isoleucine	3.0	4.3
Leucine	7.4	9.2
Tyrosine	1.1	1.2
Phenylalanine	1.4	1.4
Tryptophan	n.d.	n.d.

[1] Lilley and Inglis [32]; r^2 (bound, δ) = 0.982.

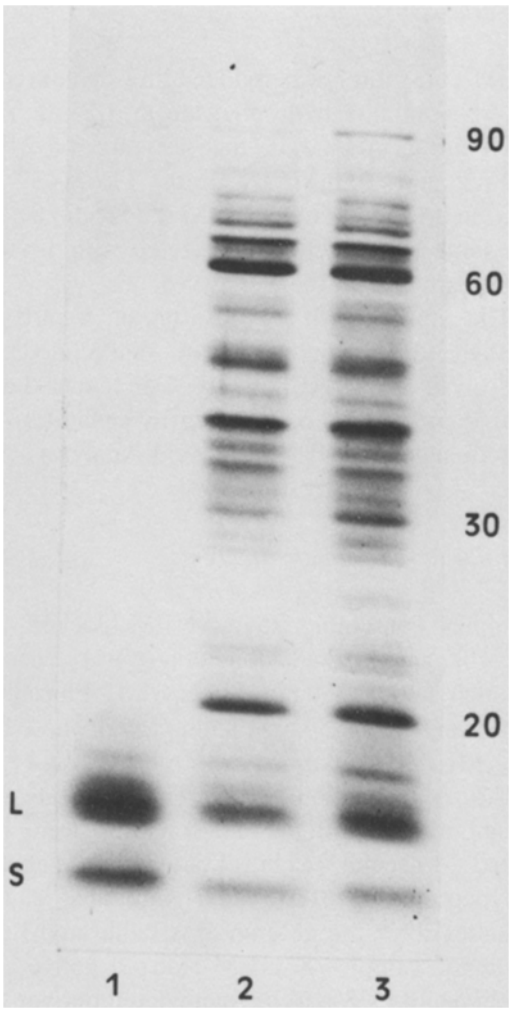

Fig. 1. Fractionation of globulins from developing cotyledons on C18 Sep-pak cartridges. Globulins were extracted from cotyledons aged 33 DAF in the presence of N-ethylmaleimide and fractionated on a reversed-phase Sep-pak cartridge as described in Materials and methods. Track 3 shows the original homogenate, track 2, unbound polypeptides eluted with water and track 1, bound polypeptides eluted with 42% (w/v) acetonitrile in 1% trifluoroacetic acid. Polypeptides from each fraction were separated on SDS-PAGE and stained with Coomassie Blue. L & S indicate the mobilities of the large and small polypeptides of authentic conglutin δ.

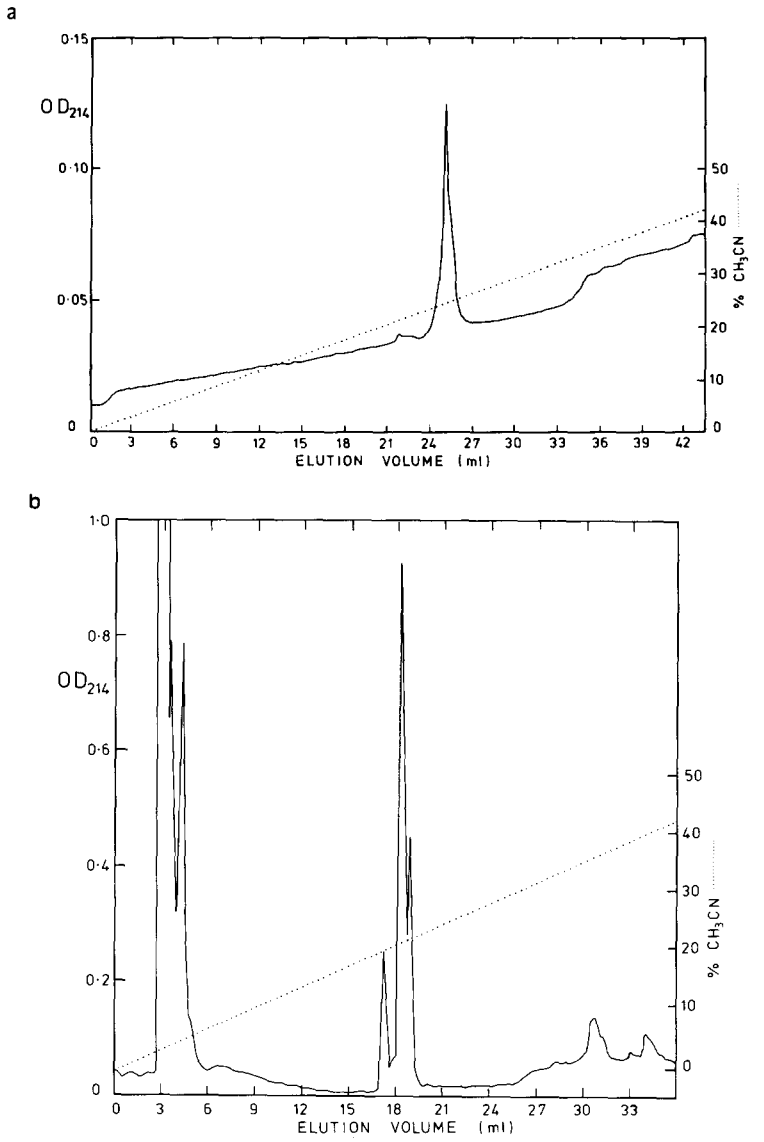

Fig. 2. Chromatography of conglutin δ from (a) mature and (b) juvenile lupin seeds on RP-HPLC. a) Authentic conglutin δ purified by the methods of Lilley [30] from mature seeds of *L. angustifolius* and b) the bound fraction from extracts of cotyledons aged 33 DAF homogenized and fractionated by chromatography on Sep-pak cartridges as described in Fig. 1, were chromatographed by RP-HPLC on a Vydac 218TP54 column using a gradient of 0% to 42% acetonitrile in 0.1% trifluoroacetic acid. Peaks at the breakthrough volume of the column (b) did not contain protein.

analytical C18 reversed-phase HPLC chromatography (Figs. 2 & 3). Figure 2 shows the material isolated from juvenile lupin seeds by binding to Sep-pak columns and run on RP-HPLC without denaturation. Figure 3 shows RP-HPLC when the samples were denatured and S-carboxymethylated before chromatography. In both cases small families of polypeptides which showed similar retention times on RP-HPLC to samples of authentic conglutin δ were detected in the Sep-pak-bound material. RP-HPLC was therefore used to assist in the identification of radiolabelled polypeptides related to conglutin δ in the subsequent studies of its biosynthesis.

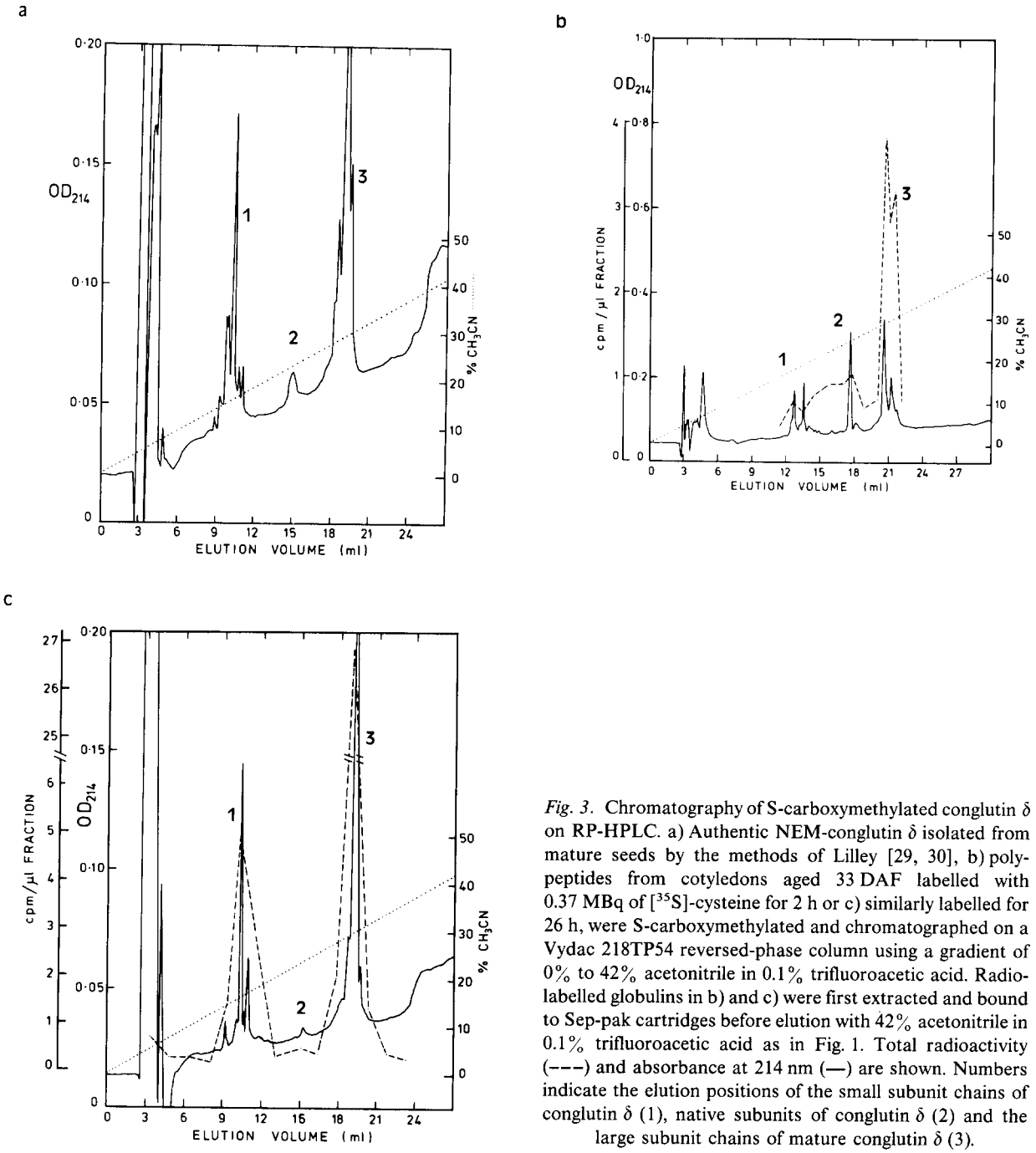

Fig. 3. Chromatography of S-carboxymethylated conglutin δ on RP-HPLC. a) Authentic NEM-conglutin δ isolated from mature seeds by the methods of Lilley [29, 30], b) polypeptides from cotyledons aged 33 DAF labelled with 0.37 MBq of [^{35}S]-cysteine for 2 h or c) similarly labelled for 26 h, were S-carboxymethylated and chromatographed on a Vydac 218TP54 reversed-phase column using a gradient of 0% to 42% acetonitrile in 0.1% trifluoroacetic acid. Radiolabelled globulins in b) and c) were first extracted and bound to Sep-pak cartridges before elution with 42% acetonitrile in 0.1% trifluoroacetic acid as in Fig. 1. Total radioactivity (---) and absorbance at 214 nm (—) are shown. Numbers indicate the elution positions of the small subunit chains of conglutin δ (1), native subunits of conglutin δ (2) and the large subunit chains of mature conglutin δ (3).

In vivo *synthesis*

The total globulins labelled with a pulse of [^{35}S]-cysteine in developing cotyledons aged 30 to 35 DAF were NEM-treated, bound to C18 Sep-pak cartridges to select for conglutin δ-related peptides and separated on gradient SDS-PAGE under reducing conditions as shown in Fig. 4

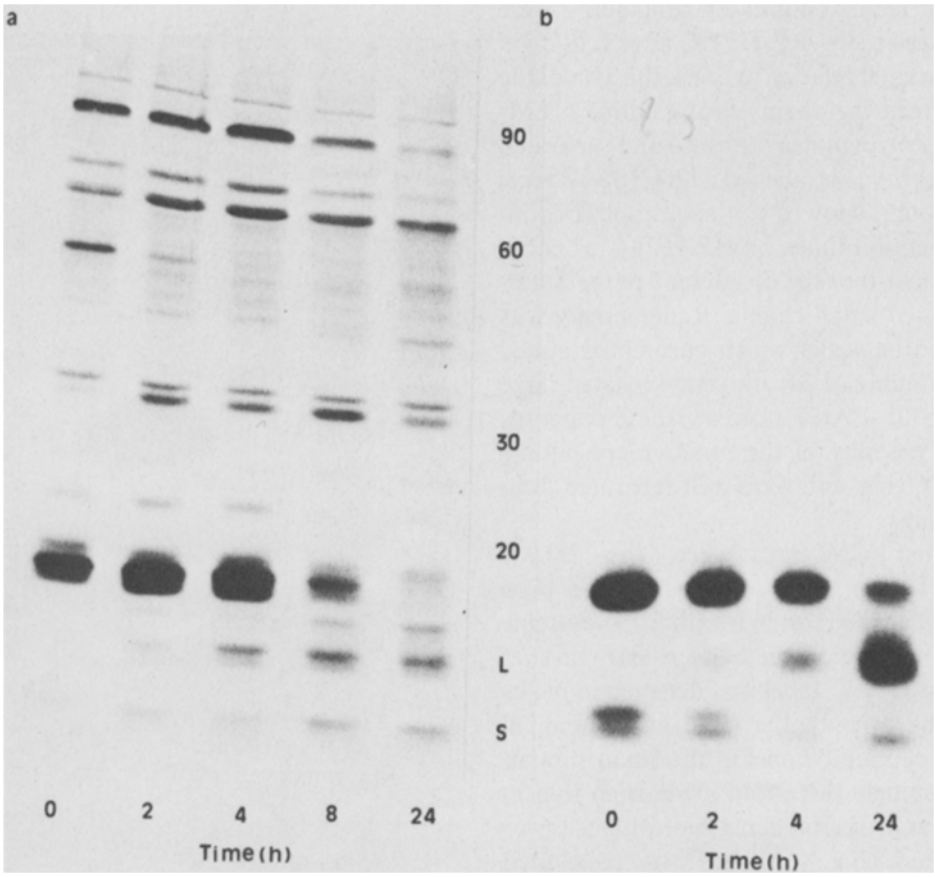

Fig. 4. Pulse-chase labelling of globulins in developing cotyledons using [^{35}S]-cysteine. Cotyledons aged 33 DAF were incubated with 0.37 MBq of [^{35}S]-cysteine for 2 h and then with unlabelled cysteine for the times shown. Protein was extracted in the presence of N-ethylmaleimide and labelled polypeptides detected by SDS-PAGE and fluorography, before and after binding to C-18 Sep-pak cartridges as in Fig. 1. a) Whole extracts; b) the bound fractions from Sep-paks. L & S indicate the expected mobilities of the large and small subunit chains from authentic mature conglutin δ.

(a: before fractionation; b: after fractionation). Protein which was the most intensely labelled with [^{35}S] and was selected by Sep-pak chromatography, had a mobility lower than either of the authentic subunit polypeptides of conglutin δ. The relatively diffuse band observed suggests a family of polypeptides rather than a single product. In subsequent sections of this paper we show that this labelled protein exhibits *in vivo* the properties expected of precursors of conglutin δ.

The behaviour of the radiolabelled protein when short-term pulse labelling was followed by an unlabelled chase was consistent with that of a precursor. Label gradually disappeared from the region identified as precursor and began to accumulate instead in those regions of the gel corresponding to the dissociated subunit polypeptides of conglutin δ (Fig. 4a & b, tracks 2 to 4). The accumulation of ^{35}S in the large subunits is clearly shown in Fig. 4b. Once again the relatively diffuse band suggests a family of large subunits rather than a single product. On SDS gels, however, the other product of cleavage of the precursors, the less heavily labelled small subunits, ran too close to contaminating radioactivity at the electrophoretic front for positive identification.

Small and large subunits of conglutin δ were also characterized by RP-HPLC after reduction and S-carboxymethylation to break the disulphide bridges between the chains in the native NEM-conglutin δ. Polypeptides prepared and separated in this way from seeds labelled with [^{35}S]-cysteine for only 2 hours, showed no label in peaks with the same retention times on RP-HPLC as either the native NEM-blocked conglutin δ or the S-carboxymethylated small chains. Radioactivity was only detected in peaks which chromatographed close to the family of S-carboxymethylated large chains. On SDS-PAGE however these materials showed the mobility of the putative conglutin δ precursor proteins, and were well separated from the large chains.

In contrast (Fig. 3c), when the NEM-conglutin δ fraction from seeds which had been continuously labelled for 26 h with [^{35}S]-cysteine was similarly S-carboxymethylated and chromatographed on HPLC, label was detected in peaks corresponding to both the large and the small subunit polypeptides. Label in the small subunit was approximately three-fold lower than that in the large subunit, a result consistent with its lower cysteine content (Fig. 8; [32]). It was concluded that [^{35}S]-cysteine initially labelled a small family of precursors of conglutin δ each of which required subsequent internal cleavage before the mature subunits could be detected. The behaviour of these precursors on RP-HPLC was dominated by the properties of that part of the chain destined to become large subunit polypeptides.

Subcellular location of conglutin δ precursors

The proteins putatively identified as precursors of conglutin δ were shown to pass through the endoplasmic reticulum fraction of cells. Cotyledons were incubated with [^{35}S]-cysteine for several hours. Tissue was then homogenized and subjected to fractionation on sucrose gradients which separated the membrane fraction containing the endoplasmic reticulum and golgi bodies from the contents of other subcellular compartments, including the protein bodies. As shown in Fig. 5, the

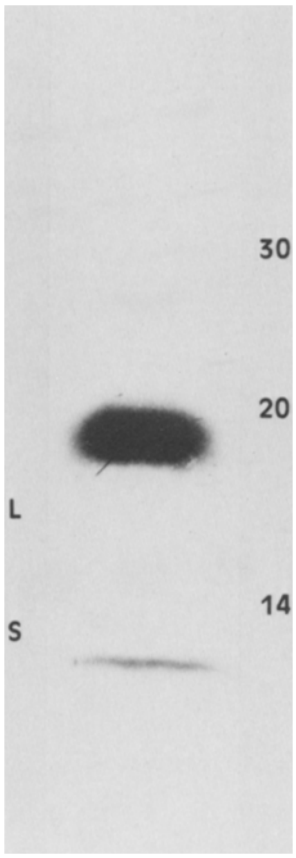

Fig. 5. [^{35}S]-Cysteine-labelled protein isolated from the endoplasmic reticulum/golgi membrane fraction of developing cotyledons. Cotyledons were labelled with [^{35}S]-cysteine for 2 h and the endoplasmic reticulum/golgi membrane fraction isolated on discontinuous sucrose gradients [20]. Protein was extracted in the presence of N-ethylmaleimide and the fraction bound to Sep-pak cartridges (as in Fig. 1) was analysed by SDS-PAGE and fluorography. Mobilities are shown relative to a scale of standard molecular weights but do not infer values for the apparent M_rs of the conglutin δ polypeptides. L & S mark the expected mobilities of the large and small subunit chains, respectively, of authentic NEM-conglutin δ.

putative NEM-treated precursors of conglutin δ were sequestered into this membrane fraction. The mean molecular weight of the ^{35}S-labelled precursors of conglutin δ which were sequestered into the endoplasmic reticulum was estimated by gel filtration chromatography after S-carboxymethylation to be 15000.

Negligible amounts of label were detected in the

ER-golgi membrane fraction in bands with mobilities equivalent to the processed conglutin δ subunits. This indicated that internal processing to generate the small and large chains of conglutin δ did not occur until after the proteins had left the endoplasmic reticulum. Such behaviour is a widespread characteristic of precursors of storage proteins and has been demonstrated specifically for the 11S proteins both in lupins [20] and in other legumes [6, 11].

Isolation and sequencing of a cDNA clone encoding conglutin δ

A cDNA library consisting of 5×10^6 recombinants was constructed from lupin mRNA. Clones containing the coding region for conglutin δ were selected by plaque hybridization, using as the probe a synthetic 17 base oligonucleotide mixture, containing the 32 possible sequences predicted from the amino acid sequence of residues 15 to 20 (Fig. 6) of the large subunit polypeptide of conglutin δ [32]. The DNA sequence of a conglutin δ clone selected in this way was established by completely sequencing both strands, making use of the restriction sites shown in Fig. 7.

As shown in Fig. 8, clone pLD14 was 684 bases long and contained an open reading frame coding for a pre-proprotein precursor of conglutin δ, M_r 17765. The predicted structure of this precursor contains an amino-terminal leader peptide of 22 amino acids, a small chain of M_r 4520 and 13 amino acids linking this to a large chain of M_r 9558. Processing of this molecule to produce the mature protein is predicted to require cleavage of the leader peptide from the N-terminus of the initial gene product during secretion into the endoplasmic reticulum, and subsequent removal of the 13 amino acid link. The cDNA sequence predicts that the size of pro-conglutin δ produced by removal of the N-terminus and sequestered into the endoplasmic reticulum would be M_r 15538. This closely matches the size of pro-conglutin δ detected in the endoplasmic reticulum by radiolabelling.

```
                OLIGONUCLEOTIDE PROBE SEQUENCES:

              Conglutin δ-large subunit polypeptide

Amino acid Sequence:

 (15 to 20):      ASP     GLN     CYS     CYS     GLU     GLN

Predicted DNA Sequence:

                5'-GA(T/C) CA(A/G) TG(T/C) TG(T/C) GA(A/G) CA-3'

Synthesized Complementary sequences:

                3'-CT(A/G) GT(C/T) AC(A/G) AC(A/G) CT(C/T) GT-5'
```

Fig. 6. Oligonucleotide probe sequences.

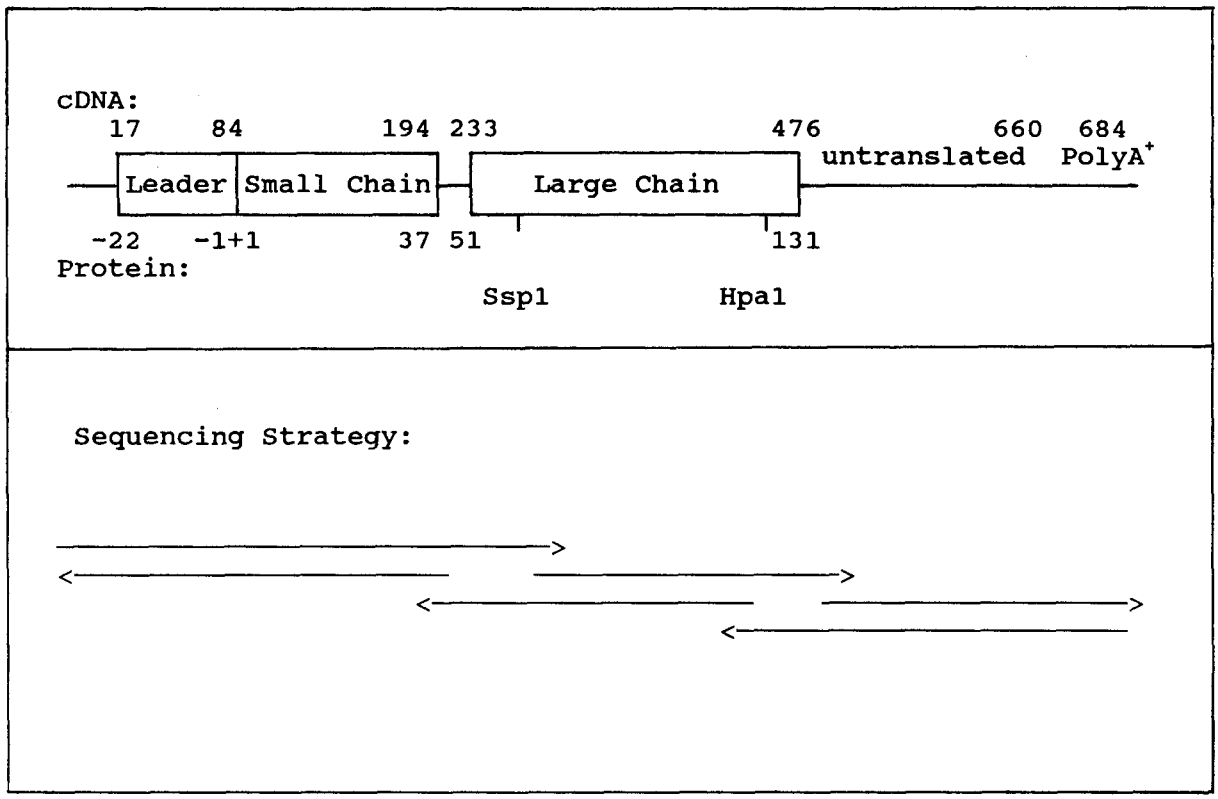

Fig. 7. Sequencing strategy for pLD14. pLD14 was sequenced in both directions as shown making use of *Ssp* I and *Hpa* I sites as shown. Amino acid + 1 corresponds to the amino terminus of conglutin δ established by amino acid sequencing [32].

Discussion

As shown by the radiolabelling studies, the family of precursors of conglutin δ was the material most highly labelled with short-term pulses of [^{35}S]-cysteine. Such behaviour is consistent with conglutin δ providing a major sink of cysteine in the seed and in fact conglutin δ accounts for up to 70% of protein sulphur in seeds of *L. angustifolius* at maturity [32]. These radioactive precursors of conglutin δ were sequestered into the endoplasmic reticulum in the manner conventional for the 11S to 7S storage proteins. Pea albumin 1, a cysteine-rich protein from peas, behaves similarly to conglutin δ and is sequestered into the endoplasmic reticulum as a high-molecular-weight precursor [19]. Both PA1 and conglutin δ have also been shown to be further transported to the protein bodies in the respective seeds [13, 19]. Not all sulphur-rich proteins from legume seeds however behave in this way. Pea albumin 2, for example, is another protein high in cysteine, but it is neither located in the protein body [17], nor rapidly mobilized upon germination [36]. The gene coding for pea albumin 2, however, does not contain a region coding for an N-terminal leader sequence [18]. It is unlikely therefore that this sulphur-rich protein in peas is secreted via the endoplasmic reticulum in the conventional way. Conglutin δ therefore appears from its biosynthesis and secretion to resemble the conventional storage proteins rather than this less conventional cysteine-rich protein.

The isolation and sequencing of cDNA coding for one of the precursors of conglutin δ has allowed the determination of its full amino acid sequence and confirmed the close relationship between conglutin δ and other 2S storage pro-

```
[1]                    CA TAA TAT CCC ACT GAA ATG GCT AAG CTC ACT ATT CTC ATA GCC CTT GTT GCT GCT
[2]                                       Met Ala Lys Leu Thr Ile Leu Ile Ala Leu Val Ala Ala
[3]

[1] CTT GTA CTA GTG GTC CAC ACT AGC GCC TTT CAA TCT TCT AAA CAA AGC TGC AAG AGG CAA CTC CAG CAG GTG
[2] Leu Val Leu Val Val His Thr Ser Ala Phe Gln Ser Ser Lys Gln Ser Cys Lys Arg Gln Leu Gln Gln Val
[3]                                  *** Arg *** *** Glu *** *** *** *** *** *** *** *** *** ***
[4]                                      -G-         G--

[1] AAC CTG AGG CAC TGC GAG AAC CAC ATA GCT CAG AGG ATT CAG CAA CAA CAA GAG GAG GAA GAA GAT CAT GCT
[2] Asn Leu Arg His Cys Glu Asn His Ile Ala Gln Arg Ile Gln Gln Gln Gln Glu Glu Glu Glu Asp His Ala
[3] *** *** *** *** *** *** *** *** *** Asp *** *** *** *** *** *** *** *** *** *** *** ***
[4]                                    -A-

[1] CTA AAA TTG CGA GGA ATC AAG CAT GTT ATT CTA CGT CAC CGA AGT AGC CAA GAA TAT TCA GAA GAA TCA GAA
[2] Leu Lys Leu Arg Gly Ile Lys His Val Ile Leu Arg His Arg Ser Ser Gln Glu Tyr Ser Glu Glu Ser Glu
[3]                                       *** *** Lys *** *** *** ***     *** *** *** *** ***
[4]                                             AA-

[1] GAA CTG GAT CAA TGC TGT GAG CAA CTG AAT GAG CTT AAT AGC CAG AGA TGC CAG TGC CGT GCA TTG CAG CAG
[2] Glu Leu Asp Gln Cys Cys Glu Gln Leu Asn Glu Leu Asn Ser Gln Arg Cys Gln Cys Arg Ala Leu Gln Gln
[3] *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
[4]

[1] ATA TAT GAG AGT CAA AGC GAG CAA TGT GAG GGA AGT CAA CAG GAA CAG CAG TTG GAG CAA GAG CTT GAG AAA
[2] Ile Tyr Glu Ser Gln Ser Glu Gln Cys Glu Gly Ser Gln Gln Glu Gln Gln Leu Glu Gln Glu Leu Glu Lys
[3] *** *** *** *** *** *** *** *** *** *** *** Arg *** *** *** *** *** *** *** Gly *** *** *** ***
[4]                                          C-A/G                          GG-

[1] TTG CCT AGG ACT TGT GGG TTT GGA CCC CTT CGC AGG TGT GAT GTT AAC CCT GAT GAA GAG TGA AAT GTA GTA
[2] Leu Pro Arg Thr Cys Gly Phe Gly Pro Leu Arg Arg Cys Asp Val Asn Pro Asp Glu Glu ***
[3] *** *** *** Ile *** *** *** *** *** *** *** *** *** Asn Ile *** *** *** *** ***
[4]             -T-                                      A-- A--

[1] CAG GAT ATG ATA GCT TAT TCT TGA AGG TTA TAA GCT ATC TTT AAT TAT GCT TTT AAT AAA GAT CAT AAT CAC
[1] TAG TTC ATG ATC ATG TAT CTA AGG TAA CGT AAA GCA TTT AAG GTA TGC TTT ATC GTT TGC CTA TGC TTA GTT
[1] AAT GAG TTG AAA TAA TGA TGT TAA ACT GAA AAA AAA AAA AAA AAA AAA AAA A: (684 Bases)
```

Fig. 8. Sequence of cDNA clone pLD14. [1] pLD14: DNA sequence determined as in Fig. 7. [2] cPre-: Amino acid sequence derived from cDNA sequence. [3] Con δ: Amino acid sequence of conglutin δ by protein sequencing [32]. [4] Base substitutions shown are the minimum needed to account for the sequence of conglutin δ.

teins. The structure of this precursor, prepro-conglutin δ, deduced from the cDNA pLD14 sequence (Fig. 8), was compared with the sequence of conglutin δ established by amino acid sequencing of the purified protein [32]. HPLC analysis of purified native conglutin δ (Fig. 8; [30]) has consistently indicated the presence of families of polypeptides corresponding to both the small and the large subunits of conglutin δ.

The amino acid sequence [32] of the most abundant of each of these families of subunit polypeptides has been compared with the sequence of clone pLD14. The comparison (Fig. 8) predicts a series of amino acid substitutions between the sequence deduced from clone pLD14 and the corresponding sequences of both the small and large subunit polypeptides of conglutin δ. In the majority of cases, single base pair substitutions in the DNA chain would account for these differences. Nevertheless, there are several amino acids (Arg 53, Tyr 58, Ser 99 and Gln 108) which require more extensive changes in the DNA. None of the substitutions create alterations at the cleavage sites predicted for the protein nor do they create any potential Asn-X-Thr or Asn-X-Ser N-glycosylation sites. The minor scale of the heterogeneity of amino acid composition therefore observed, suggests that in fact these two copies of the conglutin δ proteins were encoded by members of a closely related multigene family.

The structure of the precursor, prepro-conglutin δ, deduced from the cDNA sequence has also been compared with the precursors of other 2S proteins. The fully processed 2S proteins from many species [2, 7] resemble conglutin δ in size and physical properties. There are however sub-

stantial differences in the structures of the precursors for these proteins, and in the processing events involved in their maturation. The precursors of the 2S proteins in *Brassica* [7, 10], *Bertholletia* [9], *Arabidopsis* [24] and *Helianthus* [31] all contain up to 22 extra amino acids between the leader peptide and the N-terminus of the small subunit peptide, which must be excised during processing to form the mature proteins. By contrast, in prepro-conglutin δ the signal sequence is directly attached to the N-terminus of the mature small subunit peptide. As shown in Fig. 8, the N-terminus of prepro-conglutin δ contains many features which are widely conserved in signal peptides [34, 40, 41]. These include the terminal Met_{-22}, Ala_{-21}, Lys_{-20}; then 13 hydrophobic amino acids, residues −17 to −5; helix breakers Thr and Ser at positions −3 and −2 and finally Ala at −1. The signal peptide is therefore attached directly to the small subunit in the precursor of conglutin δ (Fig. 7). As shown in Fig. 9 this leader peptide is in fact the only substantial hydrophobic zone in this gene product.

Processing of the precursor of conglutin δ also requires the removal the 13 amino acid linker at positions 38 to 50 between the two chains which ultimately occur as disulphide-bonded peptides in the mature subunits. It is not clear, however, whether this processing of pro-conglutin δ is achieved by two single cleavage events to remove the entire linking peptide or by one cleavage followed by exopeptidase action to trim the subunits. Certainly the C-terminus of the large subunit was not subjected to such proteolysis. This second stage of proteolytic cleavage within the precursor chain is a common, but not ubiquitous, feature of protein maturation within the 2S family of proteins. The subunits of other members of the 2S superfamily of proteins including the CIII isoinhibitor of α-amylase from wheat [23] and the albumin SFA8 from sunflower [31] each contain single polypeptide chains. Therefore internal cleavage within subunits is not essential for maturation of all 2S storage proteins.

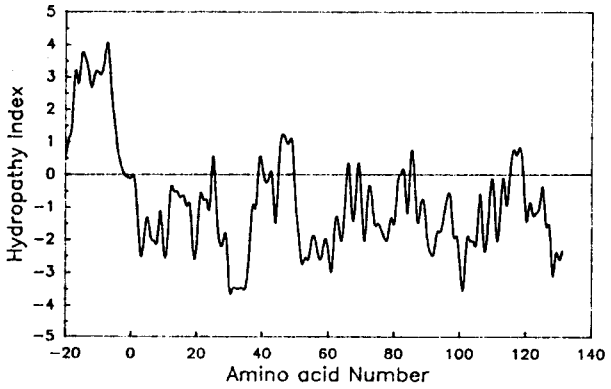

Fig. 9. Hydrophobicity profile of conglutin δ precursor. Hydropathic indices for the amino acid sequence of conglutin δ were determined by the method of Kyte and Doolittle [26]. Amino acids are numbered as in Fig. 7.

In contrast, comparisons of conglutin δ with a range of 2S storage proteins from widely divergent species (Fig. 10) show clearly that conservation of the ability to form internal disulphide bridges is important. In all but one case, that of the small S-rich protein from peas, PA1 [19], where there was no correlation between the amino acid sequences of conglutin δ and PA1, the number and relative positions of cysteines within the 2S proteins have been conserved. Figure 10 shows that all of the cysteine residues involved in complex internal disulphide bridges in conglutin δ [30] have been retained virtually in the same relative positions within the subunits of six other 2S proteins. This is so even in proteins such as the sunflower SFA8 albumin where internal cleavage is absent. In many cases the amino acids immediately surrounding the cysteines are either also conserved or at least substituted by amino acids which retain the same kind of physical properties. The one cysteine in conglutin δ (Cys_{96}) which is not conserved in any of the other 2S proteins shown in Fig. 10, participates specifically in the formation of dimeric forms of this protein [30]. It is therefore predicted that the complex network of internal disulphide bridges which has been demonstrated in lupins to be formed by the conserved cysteines [32] will also occur as major structural features of many other members of the 2S superfamily of storage proteins.

No significant areas of homology, either amino acid or DNA, beyond these regions were detected

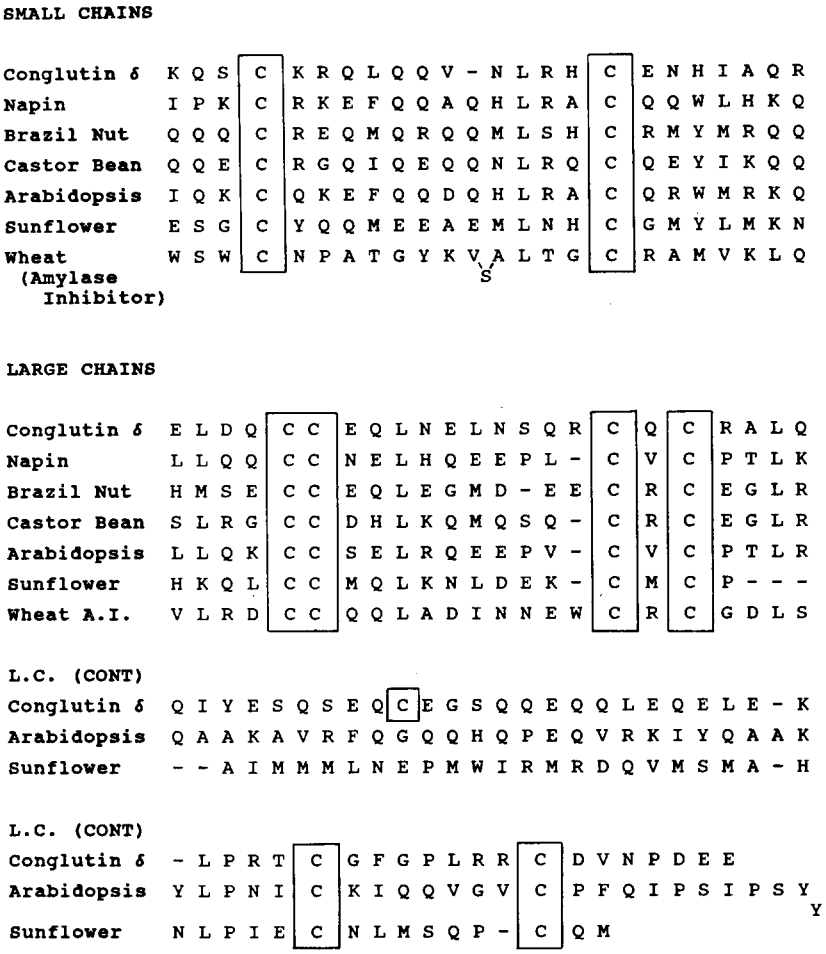

Fig. 10. Amino acid sequences of 2S seed proteins. The amino acid sequence derived from pLD14 cDNA ('Conglutin δ'), is aligned with the amino acid sequences of other 2S albumins. The segments of sequence shown as 'Conglutin δ' (amino acids +5 to +39 and +64 to +131, Fig. 7) include all cysteines in the small and large subunits respectively of conglutin δ. Blocks mark all cysteines. Segments which span the conserved cysteines are also shown for the 2S proteins from *Arabidopsis* [24] ('Arabidopsis'), *Helianthus* [31] ('Sunflower'), *Brassica* [10] ('Napin'), *Bertholletia* [1] ('Brazil Nut') and *Ricinus* [37] ('Castor Bean'), and for the CIII-α-amylase inhibitor from *Triticum* [23] ('Wheat (Amylase inhibitor)').

by the Dayhoff programmes for sequence alignment [8]. Nevertheless the presence of this complex network of cystines within the different 2S proteins shown in Fig. 10, clearly indicates that conglutin δ also shares the common ancestry of the 2S superfamily of genes [25]. It is probable that this family is derived from a third type of ancestral gene in addition to the two proposed by Borretto and Dure [4].

Acknowledgements

This work was supported by grants from the Australian Research Grants Scheme, the Rural Credits Developments Fund and from the CSIRO/University of Melbourne Collaborative Research Fund. Assistance from Mr D. Scanlon, Ludwig Institute for Cancer Research with synthesis of oligonucleotides is gratefully acknowledged.

References

1. Ampe C, Van Damme J, Castro LAB, Sampaio MAM, Van Montagu M, Vandekerkhove J: The amino acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* H.B.K.). Eur J Biochem 159: 597–604 (1986).
2. Bhatty RS, McKenzie SL, Finlayson AJ: The proteins of rapeseed (*Brassica napus* L.) soluble in salt solutions. Can J Biochem 46: 1191–1197 (1968).
3. Blagrove RJ, Gillespie JM: Isolation, purification and characterization of the seed globulins of *Lupinus angustifolius*. Aust J Plant Physiol 2: 13–27 (1975).
4. Borroto K, Dure III L: The globulin seed storage proteins of flowering plants are derived from two ancestral genes. Plant Mol Biol 8: 113–131 (1987).
5. Bowles DJ, Kauss H: Characterization, enzymatic and lectin properties of isolated membranes from *Phaseolus aureus*. Biochim Biophys Acta 443: 360–374 (1976).
6. Chrispeels MJ, Higgins TJV, Craig S, Spencer D: Role of the endoplasmic reticulum in the synthesis of reserve proteins and the kinetics of their transport to protein bodies in developing pea cotyledons. J Cell Biol 93: 5–14 (1982).
7. Crouch ML, Tenbarge KM, Simon AE, Ferl R: cDNA clones for *Brassica napus* seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. J Mol Appl Genet 2: 273–283 (1983).
8. Dayhoff MO, Schwarts RM, Orcutt BC. In: Dayhoff MD (ed) Atlas of Protein Sequence and Structure, Vol 5, Suppl 3, pp. 345–362. National Biomedical Research Foundation, Washington, DC (1979).
9. De Castro LAB, Lacerada Z, Aramayo RA, Sampaio MJAM, Ganders ES: Evidence for a precursor molecule of Brazil nut 2S seed proteins from biosynthesis and cDNA analysis. Mol Gen Genet 206: 338–343 (1987).
10. Ericson ML, Rodin J, Lenman M, Glimelius K, Josefsson L-G, Rask L: Structure of the rapeseed 1.7S storage protein, napin, and its precursor. J Biol Chem 261: 14576–14581 (1986).
11. Gatehouse JA, Croy RRD, Boulter D: The synthesis and structure of pea storage proteins. CRC Crit Rev Pl Sci 1: 287–314 (1984).
12. Gatehouse JA, Gilroy J, Hoque MS, Croy RRD: Purification, properties and amino acid sequence of a low Mr abundant seed protein from pea (*Pisum sativum* L.). Biochem J 225: 239–247 (1985).
13. Gayler KR, Wachsmann F, Kolivas S, Nott R, Johnson ED: Isolation and characterization of protein bodies in *Lupinus angustifolius*. Plant Physiol 91: 1425–1431 (1989).
14. Gerritsen TH: Lupin seed proteins: IV. Amino acid composition of the globulins from *Lupinus angustifolius* and *Lupinus luteus*. Biochim Biophys Acta 22: 269–273 (1956).
15. Grunstein M, Hogness DS: Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci USA 72: 3961–3965 (1975).
16. Gubler U, Hoffman BJ: A simple and very efficient method for generating cDNA libraries. Gene 25: 263–269 (1983).
17. Harris N, Croy RRD: The major albumin from pea (*Pisum sativum* L). Localization and immunocytochemistry. Planta 165: 522–526 (1985).
18. Higgins TJV, Beach LR, Spencer D, Chandler PM, Randall PJ, Blagrove RJ, Kortt AA, Guthrie RE: cDNA and protein sequence of a major pea seed albumin (PA 2: $M_r \sim 26000$). Plant Mol Biol 8: 37–45 (1987).
19. Higgins TJV, Chandler PM, Randall PJ, Spencer D, Beach LR, Blagrove RJ, Kortt AA, Inglis AS: Gene structure, protein structure, and regulation of the synthesis of a sulphur-rich protein in pea seeds. J Biol Chem 261: 11124–11130 (1986).
20. Johnson ED, Knight J, Gayler KR: Biosynthesis and processing of legumin-like storage proteins in *Lupinus angustifolius* (lupin). Biochem J 232: 673–679 (1985).
21. Joubert FJ: Lupin seed proteins: II. A physicochemical study of the proteins from yellow lupin seed (*Lupinus luteus*). Biochim Biophys Acta 17: 444–445 (1955).
22. Joubert FJ: Lupin seed proteins: III. A physicochemical study of the proteins from white lupin seed (*Lupinus albus*). Biochim Biophys Acta 19: 172–173 (1956).
23. Kashlan N, Richardson M: The complete amino acid sequence of a major wheat protein inhibitor of α-amylase. Phytochemistry 20: 1781–1784 (1981).
24. Krebbers E, Herdies L, De Clercq A, Seurinck J, Leemans J, Van Damme J, Segura M, Gheysen G, Van Montagu M, Vandekerckhove J: Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. Plant Physiol 87: 859–866 (1988).
25. Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR: Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183: 499–502 (1985).
26. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132 (1982).
27. Li SS-L: Purification and characterization of seed storage proteins from *Momordica charantia*. Experientia 33: 895–896 (1977).
28. Li SS-L, Lin TT-S, Forde MD: Isolation and characterization of a low-molecular-weight seed protein from *Ricinus communis*. Biochim Biophys Acta 492: 364–369 (1977).
29. Lilley GG: Isolation of conglutin δ, a sulphur-rich protein from the seeds of *Lupinus angustifolius* L. J Sci Food Agric 37: 20–30 (1986).
30. Lilley GG: The subunit structure and stability of conglutin δ, a sulphur-rich protein from the seeds of *Lupinus angustifolius* L. J Sci Food Agric 37: 895–907 (1986).
31. Lilley GG, Caldwell JB, Kortt AA, Higgins TJ, Spencer

D: Isolation and primary structure for a novel, methionine-rich protein from sunflower seeds (*Helianthus annuus* L). In: Applewhite TH (ed) Proceedings of the World Congress on Vegetable Protein Utilization in Human Foods and Animal Feedstuffs, pp. 497–502. American Oil Chemists' Society, Champaign, Illinois (1989).

32. Lilley GG, Inglis AS: Amino acid sequence of conglutin δ, a sulphur-rich seed protein of *Lupinus angustifolius* L. FEBS Lett 195: 235–241 (1986).
33. Lonnerdal B, Janson JC: Studies on *Brassica* seed proteins. 1. The low molecular weight proteins in rapeseed. Isolation and characterization. Biochim Biophys Acta 278: 175–183 (1972).
34. Perlman D, Halvorsen HO: A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J Mol Biol 167: 391–409 (1983).
35. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
36. Schroeder HE: Major albumins of *Pisum* cotyledons. J Sci Food Agric 35: 191–198 (1984).
37. Sharief PS, Li SSL: Amino acid sequence of small and large subunits of seed storage protein from *Ricinus communis*. J Biol Chem 257: 14753–14759 (1982).
38. Spencer D, Higgins TJV, Button SC, Davey RA: Pulse-labelling studies on protein synthesis in developing pea seeds and evidence of a precursor form of legumin small subunit. Plant Physiol 66: 510–515 (1980).
39. Tabor S, Richardson CC: DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 84: 4767–4771 (1987).
40. Von Heijne G: How signal sequences maintain cleavage specificity. J Mol Biol 173: 243–251 (1984).
41. Von Heijne G: Patterns of amino acid near signal cleavage sites. Eur J Biochem 133: 17–21 (1983).