Gene Structure, Protein Structure, and Regulation of the Synthesis of a Sulfur-rich Protein in Pea Seeds*

(Received for publication, February 10, 1986)

Thomas J. V. Higgins, Peter M. Chandler, Peter J. Randall, Donald Spencer, Larry R. Beach, Robert J. Blagrove, Alexander A. Kortt, and Adam S. Inglis

From the Commonwealth Scientific and Industrial Research Organization, Division of Plant Industry, Canberra, 2601 and Division of Protein Chemistry, Parkville, Victoria 3052, Australia

Two low molecular weight pea seed albumins (M, ~6000 and ~4000) have been characterized by protein, cDNA, and gene sequencing. Both proteins are encoded by separate regions of the same mRNA species. The initial translation product is a preproprotein from which a signal sequence is removed co-translationally. The resultant proprotein (PA1) is then cleaved post-translationally to yield the mature form of the two albumins (PA1a and PA1b). Comparison of cDNA and protein sequences suggests that at least four different PA1 genes are expressed in the pea genome. Both PA1a and PA1b have an unusually high cysteine content (7.5 and 16.2%, respectively). Pea seeds developing under suboptimal levels of sulfur nutrient supply contain reduced levels of PA1 mRNA and accumulate greatly reduced levels of PA1a and PA1b in the mature seed. In vitro transcription studies showed that this reduced level of PA1 mRNA resulted from reduced post-transcriptional stability rather than an altered rate of transcription of the PA1 gene. In contrast, during normal seed development, the level of PA1 mRNA seems to be under transcriptional control. Sequence comparisons reveal some homology between PA1 and a number of low molecular weight proteins from seeds of a wide range of mono- and dicotyledonous plants.

The seeds of legumes and most other dicotyledonous plants contain two major protein classes, globulins and albumins. The globulin fraction, which accounts for ~70% of the total seed protein, is commonly composed of two major protein families, namely, the “11 S” or legumin-like and the “7 S” or vicilin-like proteins described by Osborne (1924). During germination, the globulins are rapidly broken down to provide the major source of nitrogen utilized by the developing seedling (see Derbyshire et al., 1976 and Higgins, 1984 for reviews). The seed albumins of dicotyledonous plants are a far more diverse group (Derbyshire et al., 1976) and are usually classified as “2 S” proteins. Included in this fraction are the many proteins required for cellular metabolism, together with a limited number of major proteins that may have storage or other functions (Derbyshire et al., 1976; Youle and Huang, 1981; Gatehouse et al., 1985). The nutritional value of seeds in animal diets is often highly correlated with albumin content (Bajaj et al., 1971), probably because of their relatively high levels of the essential amino acids. However, in legumes, this desirable property is often offset by the presence of antinutritional factors, such as protease inhibitors and allergens, also present in the albumin fraction. Pea (Pisum sativum L.) seeds contain 25-30% albumins (Schroeder, 1984) and are low in antinutritional factors (Valdebouze et al., 1986). The presence of a major, sulfur-rich component of the pea albumin fraction has been reported earlier (Jakubek and Przybylska, 1979). This albumin (M, ~6000) contains only 4.5% of the total pea seed protein, but contributes ~23% of the seed’s sulfur amino acids (Schroeder, 1984). It is not found in pods or other plant tissues and was a major component of the seed albumin fraction in all 45 lines of Pisum examined (Schroeder, 1984). Gatehouse et al. (1985) have determined the amino acid sequence of this polypeptide.

This paper reports the gene and cDNA sequences of a sulfur-rich pea seed albumin (PA1) which proved to be the precursor of both the M, ~6000 albumin (PA1a) and a second, even more sulfur-rich polypeptide of M, ~4000 (PA1b). The genes encoding PA1 are developmentally regulated at the level of transcription. However, under stress conditions induced by sulfur nutrient deficiency, post-transcriptional controls appear to modulate the level of PA1 mRNA. PA1a and PA1b collectively contribute ~50% of the total sulfur amino acids of the pea seed. PA1b shows limited sequence homology with the superfamily of seed proteins which include protease inhibitors, amylase inhibitors, and a diverse range of other low molecular weight albumins (Shewry et al., 1984).

MATERIALS AND METHODS

Plant Material—Peas (P. sativum L.) line PI/G 086 from cv. Greenfrost were grown in artificially lit cabinets as described (Miller and Spencer, 1974) except that the temperature was at 20 °C with a 16-h photoperiod. Control plants were grown in Perlite/Vermiculite (1:1, v/v) and supplied with complete nutrient (Randall et al., 1979). Sulfur-deficient plants were grown in sand/Perlite (1:1, v/v) and supplied with a sulfur-deficient nutrient as described earlier (Randall et al., 1979; Chandler et al., 1983).

Isolation of Total Pea Albumins and Pea Albumins 1a and 1b—Total pea albumins were prepared by the extensive dialysis against water of a total salt extract of mature pea seeds (Schroeder, 1982). PA1 was prepared by treatment of either a total extract of pea meal or a total pea albumin extract with methanol at 60%, v/v (aqueous methanol), at which concentration PA1 was soluble while all other pea albumins and globulins were insoluble. PA1a (M, ~6000) was prepared by applying either a total albumin fraction or a PA1 preparation to a DEAE-Sephalcel column in 0.05 M Tris-HCl, pH 8.0, and eluting with a linear gradient of 0-0.25 M NaCl. Individual peaks were analyzed by SDS-PAGE; the major component eluting at 0.16–0.18 M NaCl was PA1a. PA1b (M, ~4000) was the main protein component of that fraction of the total albumin or PA1 preparation which did not bind to DEAE-Sephalcel. SDS-PAGES was carried out

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 The abbreviations used are: PA1, pea albumin 1; SDS-PAGE, sodium dodecyl sulfate-polycrylamide gel electrophoresis; HPLC, high performance liquid chromatography; DAF, days after flowering.
with acrylamide gradients of 12.5–25% and 12.5–30% (Spencer et al., 1980). On SDS-PAGE, PA1a was always more weakly stained by Coomassie Blue than PA1a to the point where it was sometimes difficult to detect. The reason for this is not known.

In Vitro and in Vivo Synthesis of PA1, PA1a, and PA1b—Total RNA from developing pea seeds 20 days after flowering (20 DAF) was translated in a rabbit reticulocyte lysate system as described earlier (Higgins and Spencer, 1977) and the PA1a-related products were immunoprecipitated with antiserum from rabbits to which had been injected with PA1a conjugated to keyhole limpet hemocyanin (Ghetie and Sjoquist, 1984). The PA1a for this purpose was prepared by chromatography on DEAE-Sephacel.

Pulse and pulse-chase labeling experiments using isolated cotyledons and a mixture of 3H-labeled amino acids were carried out as described earlier (Spencer et al., 1980).

Amino Acid Sequencing—PA1a and PA1b from the DEAE-Sepharose fractionation were reduced and alkylated with iodoacetic acid as described by Crestfield et al. (1963). The carboxymethylated proteins were dissolved in 0.1% (v/v) trifluoroacetic acid and fractionated by HPLC in 0.1% (v/v) trifluoroacetic acid on a Waters μBondapak C18 analytical column using a linear gradient of acetonitrile (20–50%, v/v, CH₃CN in 15 min; flow rate, 1 ml/min). PA1a yielded two major protein peaks, whereas PA1b yielded a single major peak. The amino-terminal sequences of these components were determined using a modified protein sequenator equipped with a stationary stainless steel reaction module (Woods and Inglis, 1984). The proteins were digested with trypsin (Worthington), chymotrypsin (Worthington), and Staphylococcus aureus V8 protease (Pierce), and the peptides were isolated by reversed-phase HPLC as described by Kortt et al. (1982).

Peptides were sequenced manually by a modified Edman procedure (Peterson et al., 1972) using 50% pyridine as the couple buffer and extracting with n-heptane/ethyl acetate (2:1, v/v) instead of benzene. Phenylthiohydantoin amino acid derivatives were identified by HPLC as described by Woods and Inglis (1984).

Construction and Sequencing of cDNA and Genomic Clones and Primer Extension—cDNA clones were constructed using poly(A) RNA from midstage (22 DAF) developing pea seeds as a template as described earlier (Chandler et al., 1983, 1984). The plasmid pPS15-91 was identified as one of nine abundant hybridization families in the total plasmid population; by hybrid release translation, it selected mRNA for a protein of M₆ = 13,000. For the preparation of a pea genomic DNA library, a partial Sau3A1 digest of pea shoot DNA was separated on low melting point agarose, and the 12–20-kilobase fraction was isolated and ligated into the BamH1 site of λ1059 (Karn et al., 1980). The library was screened by plaque hybridization (Maniatis et al., 1982) using the insert of the PA1 cDNA clone pPS15-91. Recombinant DNA work was carried out under C1 conditions as recommended by the Australian Recombinant DNA Monitoring Committee.

Sequencing of both the cDNA clone (pPS15-91) and the genomic clone (PS125.1) was carried out using the dideoxy chain termination method (Sanger et al., 1977). The sequence of the genomic clone was obtained from a series of random clones generated by sonication of the DNA followed by subcloning into the Smal1 site of M13mp8 (Deininger, 1983). Some additional sequencing of the cDNA clone was carried out by the chemical method (Maxam and Gilbert, 1980).

The DNA sequence of both strands of the genomic and cDNA clones were determined several times.

A synthetic oligonucleotide corresponding to positions 218–237 was annealed to total RNA from cotyledons 27 DAF. Primer extension was carried out as previously described (Higgins et al., 1983a). The length of the extended primers were determined by electrophoresis on 12% polyacrylamide gel containing 7 M urea (Maniatis et al., 1982).

In Vitro Transcription and PA1 mRNA Levels—The extent of transcription of PA1 genes in isolated nuclei from cotyledons at different stages of development and at different stages of recovery from sulfur stress was estimated as described elsewhere (Beach et al., 1983a) using an excess of the PA1 cDNA clone, pPS15-91, bound to nitrocellulose to hybridize to the PA1-related transcription products. PA1a mRNA levels in total cotyledon RNA were determined by the dot-blot procedure (White and Bancroft, 1982) using nick-translated pPS15-91 as the probe.

RESULTS

Isolation, Tissue Location, and Function of Pea Albumins 1a and 1b—The total albumin fraction of mature pea seeds, prepared as described under "Materials and Methods," contained two major and several minor polypeptides when fractionated by SDS-PAGE (Fig. 1a, lane 2). This is in marked contrast to the total extract of pea seeds (Fig. 1a, lane 1) whose greater complexity is due to the globulin storage proteins. The major albumin components revealed by Coomassie Blue staining are the Mr ~ 25,000 and ~ 6,000, respectively. The Mr ~ 6,000 albumin, which will be referred to as pea albumin 1a (PA1a), could be separated from the other major albumins by either one of two methods. PA1a, together with a minor albumin of Mr ~ 4,000 (pea albumin 1b, PA1b), remained in solution when an aqueous solution of the total albumin fraction was treated with methanol at 60% final concentration. Both proteins were precipitated from this solution with 4 volumes of acetone (Fig. 1a, lane 2). Alternatively, PA1a could be purified by chromatography of the total albumin fraction on DEAE-Sephaloc (Fig. 1a, lane 4). PA1b did not bind to DEAE-Sephaloc and was isolated from the nonbound fraction by HPLC (Fig. 1a, lane 5).

PA1a is a major component of both the cotyledons and embryonic axes of mature pea seeds (Fig. 1b). It is broken down during germination, suggesting that it may have a
Protein and Gene Sequences of a Sulfur-rich Pea Seed Albumin

Pea cotyledons at 22 DAF was used as the template in a wheat germ cell-free translation system and the total ¹⁴C-labeled products were treated with rabbit antiserum against PA1a. Electrophoresis of the resultant immunoprecipitated product on SDS-PAGE followed by fluorography showed the presence of a single major radioactive product of Mₓ ~ 13,000 (Fig. 2, lane 1).

Pea cotyledons at 22 DAF were pulse-labeled with a mixture of ¹⁴C-amino-acids for 2 h, after which some were incubated for a further 23 h with a nutrient solution containing nonradioactive amino acids (pulse-chase). Total salt-soluble extracts of these cotyledons were treated with 60% methanol and the soluble components were fractionated by SDS-PAGE. The only aqueous methanol-soluble, radioactive product detected in the pulse-labeled cotyledons had an Mₓ ~ 11,000 (Fig. 2, lane 2). After a 23-h chase period, this product had largely disappeared and the only two methanol-soluble products were of Mₓ ~ 6,000 and 4,000, respectively (Fig. 2, lane 3).

The respective sizes of the PA1-related in vitro product (Mₓ ~ 13,000) and the pulse (Mₓ ~ 11,000) and pulse-chase (Mₓ ~ 6,000 and ~ 4,000) in vivo products are consistent with the proposal that PA1 is initially synthesized as a large precursor molecule which undergoes both co- and post-translational modification. This model is confirmed by the sequence data presented in the next section.

Unexpectedly, antiserum to PA1a proved to be suitable only for immunocytochemistry on tissue sections and for immunoprecipitation of the in vitro precursor product. It could not be used to immunoselect the related products in cotyledon extracts. For this reason, their unusual, but characteristic, solubility in aqueous methanol was used to identify PA1-related proteins in the pulse-chase experiments.

Genomic, cDNA, and protein Sequences of PA1—A cDNA library was prepared using poly(A) RNA from pea cotyledons at 22 DAF and a cDNA clone (pPS15-91) was isolated which, in hybrid release translation experiments, selected an mRNA coding for a polypeptide of Mₓ ~ 13,000 (data not shown). This cDNA was sequenced and found to consist of 699 bases 5' to the poly(A) tail with an open reading frame of 390 bases extending from an ATG located 18 bases from the 5' end (Fig. 3b).

FIG. 3. The gene, cDNA, and amino acid sequences of pea albumin 1. a, gene sequence; b, cDNA sequence; c, amino acid sequences deduced from gene and cDNA, respectively; e, chemically determined amino acid sequences of PA1a (Mₓ ~ 6000) and PA1b (Mₓ ~ 4000). There is an intron (shown in lower case letters) in the gene sequence between positions 269 and 351 and a putative signal sequence between positions 220 and 590. — indicates homology between nucleotide or amino acid sequences; * indicates where a space has been inserted in the sequence of either the gene or the cDNA to maximize homology.

PA1a was isolated from a total albumin preparation by fractionation on DEAR-Sephadex, followed by reversed-phase HPLC to yield two major protein peaks (see "Materials and Methods"), both of which were fully sequenced. The sequence of one peak is shown in Fig. 3e. The sequence of the other peak was identical with that shown except for an Arg/Lys substitution at position 549 and a Phe/Val substitution at position 642. When compared with the deduced amino acid sequence of the insert cDNA in pPS15-91 (Fig. 3d), the chemically determined sequences for PA1a were highly homologous to a region of the deduced amino acid sequence beginning at position number 510 in Fig. 3 and extending for the equivalent of 159 bases.

Fig. 2. The synthesis of pea albumin 1-related proteins in vivo and in vitro. Lane 1, total RNA was translated in a cell-free system from wheat germ and the radioactive products were immunoselected with antiserum against PA1a; lanes 2 and 3, pea cotyledons (22 DAF) were pulse-labeled with a mixture of ¹⁴C-amino-acids for 2.5 h (lane 2) and then transferred to nonradioactive medium for a further 23 h (lane 3). Cotyledons were then extracted and the proteins soluble in aqueous methanol were fractionated by SDS-PAGE and detected by fluorography. Approximate Mₓ values (×10⁶) of polypeptides are indicated on the vertical axes.
PA1b, the other albumin component which was soluble in aqueous methanol, did not bind to DEAE-Sephaloc and yielded a single major component on reversed-phase HPLC. However, in contrast to PA1a, the major PA1b component was found to be composed of three related sequence variants. Substitutions were found at 10 positions in the amino acid sequence: 381 (Ile/Ala/Val), 411 (Asp/Glu), 414 (Ile/Met), 429 (Ser/Thr), 432 (Pro/Ser), 435 (Leu/Ala), 453 (Ala/Tyr/Val), 462 (Val/Phe), 471 (Asn/Tyr/Lys), and 486 (Tyr/Ser). The complete sequence of the major variant of PA1b (Fig. 3e) showed 73% homology with a different region of the deduced amino acid sequence from pPS15-91. Its chemically determined NH$_2$ terminus corresponds to position 381 in Fig. 3 and it extends to position 491 with a further 6 amino acid residues in the deduced sequence prior to the NH$_2$ terminus of PA1a. The protein sequence data indicate that the mature PA1b protein has a molecular weight of 3742. The mRNA represented in pPS15-91 thus codes for proteins homologous to both PA1b and PA1a in that order from the 5' end.

In the pPS15-91 cDNA sequence, there are three possible methionine codons 5' to the chemically determined NH$_2$ terminus of PA1b. The ATG codon at position 220 is preceded by an A at -3 and followed by a G at +4, consistent with the rules for an initiator codon (Kozak, 1984) and is followed by a deduced sequence of 26 amino acids which has the characteristics of a signal sequence typical of proteins destined for transmembrane transport (Perlman and Halvorsen, 1985). Seventeen of the 26 amino acids in the signal sequence are hydrophobic, an alanine residue (position 378) immediately precedes the NH$_2$ terminus of the mature protein and positive lysine residues occur before (position 232) and after (position 366) a hydrophobic core region. Assuming that the methionine at position 220 in Fig. 3 is the initiator, the total length of the polypeptide encoded by the open reading frame of pPS15-91 is 130 amino acid residues with a total molecular weight of 13,970. This is consistent with the size of the in vitro translation product (apparent $M_,$ ~13,000) which is selectively precipitated by antiserum to PA1a (Fig. 2, lane 1). Cleavage of the putative preproprotein at position 380 to remove the leader sequence would yield a proprotein of $M_,$ = 11,274, in agreement with the size of the polypeptide observed in vivo after pulse labeling ($M_,$ ~11,000, Fig. 2, lane 2).

A genomic library of pea DNA was constructed in X phage and the cDNA of pPS15-91 was used to identify fragments containing the gene for PA1. Part of one such fragment of approximately 2 kilobases was sequenced and found to be highly homologous with, but not identical with, the sequence of pPS15-91 cDNA (Fig. 3a). Within the coding regions of the gene, including the putative signal sequence, there are only six differences in nucleotide sequence and these result in four changes in the predicted amino acid sequence.

The genomic DNA sequence includes a region of 83 bases (beginning at residue 269), not represented in the cDNA and having the characteristics of an intron. The intron occurs in the region coding for the putative signal sequence and it divides the codon for a glycine residue (Fig. 3). The sequences at the 5' (TAG/GTACGT) and 3' (TGGCAG/GT) splice sites are closely homologous to the respective consensuses, C_{AG}/GTAGT and T_{C/T}NC,TAG/G, for these two sites (Padgett et al., 1985).

In maximizing homology between the cDNA and the gene sequences, we have noted the presence of a 20-base insertion in the cDNA at nucleotide positions 727–746 and single base deletions at positions 716 and 842. In addition, there are 25 base changes in the 3' untranslated region of the cDNA with respect to the gene.

Primer extension experiments designed to determine the 5' end of the PA1 mRNA showed that the 5' untranslated region was heterogeneous in length. Four discrete 5' termini for PA1 mRNA were detected, indicating that four genes are expressed, although mRNA degradation to discrete lengths cannot be ruled out (data not shown). Although the start site of transcription on the particular PA1 gene described here is not known, there is a TATAAA sequence beginning at position 146.

There are putative poly(A) addition signals (AATAAA) at positions corresponding to base numbers 756 and 816 in both the cDNA and the gene and an additional sequence at position 875 in the cDNA, the latter being 20 bases from the poly(A) tail.

**Transcriptional and Post-transcriptional Control of PA1 Gene Expression**—The regulation of PA1 gene expression was studied under two physiologically distinct situations, namely, under nutrient stress and as a function of normal seed development.

It has been shown previously that in pea plants growing at suboptimal levels of sulfur supply there is a selective reduction in the level of accumulation of several major seed proteins (Chandler et al., 1983; Schroeder, 1984; Chandler et al., 1984) including PA1a (Fig. 4a). Pulse-labeling with $^{14}$C-amino-acids at intervals throughout development showed that reduced accumulation of PA1 appears to be a consequence of reduced synthesis (Fig. 4b). Restoration of an adequate sulfur supply (as inorganic sulfate) to the whole plant as late as midway
through seed development results in a rapid increase in the level of PA1 synthesis. Fig. 4c shows the results of an experiment in which sulfur-deficient plants at 20 DAF were given adequate nutrient sulfur and cotyledons were harvested at daily intervals thereafter and pulse-labeled with a $^{14}$C-amino-acid mixture. A marked increase in the level of synthesis of PA1 ($M_1 = 11,000$) occurred within 24 h. This increase in the level of PA1 synthesis on restoration of normal sulfur supply was accompanied by an equal rapid increase in the level of PA1 mRNA (Fig. 5). Dot-blot analysis of equal amounts of total RNA from pea cotyledons before and during recovery from sulfur deficiency showed a marked increase in PA1 mRNA 1 day after restoration of sulfur supply. After 3 days, the level was as high or higher than in the cotyledons of a normal, nondeficient plant of similar age.

This rapid change in PA1 mRNA level in response to sulfur status could be due to either an altered rate of transcription of the PA1 genes, an altered stability of the gene transcripts, or to a combination of these two possibilities. To resolve this question, the rate of transcription of PA1 genes in isolated nuclei and the level of accumulated PA1 mRNA were measured in cotyledons taken from the same batch of recovering plants (Fig. 6a). The results are expressed as a ratio of the activity in recovering plants (+S) to that in untreated, sulfur-deficient plants (−S). Forty-eight h after restoration of sulfur supply, there was a 9-fold increase in the level of accumulated PA1 mRNA (Fig. 6a). However, in the same period, there was little change in the relative rate of transcription of PA1 genes (Fig. 6b, histogram), as measured by in vitro assays with isolated nuclei. This result indicates that the control of PA1 gene expression in response to altered sulfur status mainly takes place at the post-transcriptional level rather than at the level of gene transcription.

In contrast, regulation of PA1 gene expression during development of normal, nondeficient seeds appears to be under direct transcriptional control. It was reported earlier (Chandler et al., 1984) that PA1 mRNA levels showed a characteristic developmental pattern with maximum levels at 20–22 DAF followed by a sharp decline. Fig. 6b shows the results of an experiment in which both PA1 gene transcription in isolated nuclei and total cellular levels of PA1 mRNA were measured at intervals during seed development. In contrast to the situation under sulfur stress, a marked parallel was found between the level of PA1 transcription and the level of PA1 mRNA (Fig. 6b).

**DISCUSSION**

Legume seed proteins in general contain very low levels of sulfur amino acids. Here, we report the existence of a low molecular weight protein in peas (PA1) which contributes less than 10% of the total seed protein, yet provides over 50% of the seed protein sulfur amino acids. We speculate that it may function as a source of sulfur for the germinating seedling. A similar role in the storage of sulfur has previously been proposed for other low molecular weight proteins in plants (Youle and Huang, 1981).

The nucleotide and protein sequence data in this report, together with in vitro and in vivo biosynthesis results, all show that PA1 is a precursor molecule which is post-translationally cleaved to yield two polypeptides, PA1a of $M_1 = 6,000$ and PA1b of $M_1 = 4,000$. PA1 is first synthesized in a prepro form with a signal sequence which is presumably removed co-translationally to yield the earliest, stable product detected in
vivo, a proprotein (PA1, M, ~11,000) containing the sequences of both PA1a and PA1b. This proprotein is then cleaved endoproteolytically to yield two polypeptides which, after removal of some carboxyl-terminal amino acids, represent the mature forms of PA1a and PA1b. The COOH-terminal amino acid of PA1b was identified as glycine from peptide sequence data and from digestion with carboxypeptidase Y indicating the loss of the terminal 6 amino acid residues during post-translational processing. Attempts to confirm the COOH-terminal residue of PA1a by carboxypeptidase Y digestion yielded an equivocal result, but sequence and amino acid composition data for the COOH-terminal peptide indicate a terminal aspartic acid residue (Fig. 3) implying the terminal residue seems unlikely since the gene and cDNA sequences predict that Leu follows Asp at this position. Furthermore, sedimentation equilibrium measurements (data not presented) show that PA1a is monomeric; Gatehouse et al. (1985) suggest that this component is dimeric on the basis of an empirical molecular weight measured by gel filtration.

PA1 has several features in common with the major pea seed storage proteins, legumin and vicilin, and also with pea seed lectin, with respect to biosynthesis and post-translational processing. The latter three proteins are synthesized as preproteins on the rough endoplasmic reticulum and are transported within the endoplasmic reticulum to the protein bodies where they are endoproteolytically cleaved to form the subunits characteristic of the mature protein (Spencer and Higgins, 1980; Higgins and Spencer, 1981; Chrisspeels et al., 1982, a and b; Higgins et al., 1983, a and b; Spencer et al., 1983). Given the similarities in biosynthesis between PA1 and the major seed proteins, together with the fact that its endoproteolytic processing time (between 2.5 and 22 h) is in the same range as that of vicilin and pea seed lectin (Chrisspeels et al., 1982b; Higgins et al., 1983a), it seems likely that PA1 is also transported to the protein bodies as the proprotein (M, ~11,000) prior to cleavage. Using immunogold cytochemical labeling, it has been demonstrated that PA1a is associated with the endoplasmic reticulum and is deposited in the protein bodies but we have no direct knowledge of the subcellular site of accumulation of PA1b.

The high cysteine content (11%) and the small size of PA1a and PA1b suggest a possible relationship to the propeptide inhibitors of the Bowman-Birk trypsin inhibitor type. Some slight sequence homology was found between the NH-terminal region of PA1b and the region surrounding the active site in chick pea and lima bean propeptide inhibitors and in human a, anti-trypsin (Fig. 7a). However, the ksp/Ser substitution in PA1b at the active site would result in a lack of anti-protease activity for PA1b (Birk, 1976). This sequence homology does not extend to the remainder of PA1b nor to the sequence of PA1a. Gatehouse et al. (1985) have previously reported that PA1a lacks anti-protease activity.

In addition to the limited homology between PA1b and some of the propeptide inhibitors (Fig. 7a), there is a low level of homology between PA1b and a number of other low molecular weight seed albumins discussed recently by Shewry et al. (1984) (Fig. 7b). This latter group of albumins is thought to constitute a superfamly of proteins which include such diverse members as barley protease inhibitors soluble in chloroform and methanol, wheat α-amylase inhibitor 0.28 (Kashaian and Richardson, 1981), millet bifunctional inhibitor (Campos and Richardson, 1983), castor bean 2S small subunit (Sharief and Li, 1982), napin small subunit (Crouch et al., 1983).

Expression of the PA1 genes appear to be under both transcriptional and post-transcriptional control. During normal seed development, there is little or no transcription of these genes in early development. Transcription progressively increases during seed development and results in an increased accumulation of PA1 mRNA. This developmentally regulated pattern is also characteristic of pea vicilin and legumin genes (Beach et al., 1985) and of soybean seed protein genes (Goldberg et al., 1983). This contrasts sharply with the situation in sulfur-deficient pea seeds. Here, PA1 and legumin gene expression appears to be largely under post-transcriptional control (this paper and Beach et al., 1985).

2 S. Craig, personal communication.

3 P. M. Chandler, personal communication.
Acknowledgments—We thank Dr. W. Bottomley for his major contribution to the computer analysis of all the sequence data. We also thank M. Calderamith, E. J. Newbiggin, J. Cassells, and R. E. Guthrie for their skilled technical assistance and D. McCann for preparation of the manuscript.

REFERENCES
Bajaj, S., Mickelsen, O., Lillevik, H. A., Barker, L. R., Bergen, W. G., and Gill, J. L. (1971) Crop Sci. 11, 813–815
Beach, L. R., Spencer, D., Randall, P. J., and Higgins, T. J. V. (1985) Nucleic Acids Res. 13, 999–1013
Birk, V. (1976) Methods Enzymol. 45B, 707–709
Campos, F. A. P., and Richardson, M. (1983) FEBS Lett. 152, 300–304
Chandler, P. M., Higgins, T. J. V., Randall, P. J., and Spencer, D. (1983) Plant Physiol. 71, 47–54
Chandler, P. M., Spencer, D., Randall, P. J., and Higgins, T. J. V. (1984) Plant Physiol. 75, 651–657
Chrispeels, M. J., Higgins, T. J. V., Craig, S., and Spencer, D. (1982a) J. Cell Biol. 93, 5–14
Chrispeels, M. J., Higgins, T. J. V., and Spencer, D. (1982b) J. Cell Biol. 93, 396–313
Crestfield, A. M., Moore, S., and Stein, W. H. (1963) J. Biol. Chem. 238, 622–627
Crouch, M., Tenbarge, K. M., Simon, A. E., and Ferl, R. (1983) J. Mol. Appl. Genet. 2, 273–283
Deininger, P. L. (1983) Anal. Biochem. 129, 216–223
Derbyshire, E., Wright, D. J., and Boulter, D. (1976) Phytochemistry 15, 3–24
Gatehouse, J. A., Gilroy, J., Hoque, M. S., and Croy, R. R. D. (1985) Biochem. J. 225, 239–247
Ghetie, V., and Sjoquist, J. (1984) Methods Enzymol. 108, 405–413
Goldberg, R. B., Croach, M. L., and Walling, L. (1983) in Manipulation and Expression of Genes in Eucaryotes (Nagley, P., Linnane, A. W., Peacock, W. J., and Pateman, J. A., eds) pp. 193–201, Academic Press Australia, North Ryde, Australia
Higgins, T. J. V. (1984) Annu. Rev. Plant Physiol. 35, 191–221
Higgins, T. J. V., and Spencer, D. (1977) Plant Physiol. 60, 655–661
Higgins, T. J. V., and Spencer, D. (1981) Plant Physiol. 67, 205–211
Higgins, T. J. V., Chandler, P. M., Zurawski, G., Button, S. C., and Spencer, D. (1983a) J. Biol. Chem. 258, 9544–9549
Higgins, T. J. V., Chrispeels, M. J., Chandler, P. M., and Spencer, D. (1983b) J. Biol. Chem. 258, 9550–9552
Jakubeck, M., and Przybylska, J. (1979) Genet. Pol. 20, 369–380
Johnson, D., and Travis, J. (1978) J. Biol. Chem. 253, 7142–7144
Karn, J., Brenner, S., Barnett, L., and Caserini, G. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5172–5176
Kashlan, N., and Richardson, M. (1981) Phytochemistry 20, 1781–1784
Kort, A. A., Burns, J. E., Trinick, M. J., and Appleby, C. A. (1985) FEBS Lett. 180, 55–60
Kozak, M. (1984) Nucleic Acids Res. 12, 857–872
Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning—A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–500
Miller, A., and Spencer, D. (1974) Aust. J. Plant Physiol. 1, 331–341
Osborne, T. B. (1924) in The Vegetable Proteins, 2nd Ed, Longman Inc., New York
Padgett, R. A., Grabowski, P. J., Konarska, M. M., and Sharp, P. A. (1985) Trends Biochem. Sci. 10, 154–157
Perlman, D., and Halvorson, H. O. (1983) J. Mol. Biol. 167, 391–409
Petersen, J. D., Nehrlich, S., Oyer, P. E., and Steiner, D. F. (1972) J. Biol. Chem. 247, 4866–4871
Randall, P. J., Thomson, J. A., and Schroeder, H. E. (1979) Aust. J. Plant Physiol. 6, 11–24
Richardson, M. (1977) Phytochemistry 16, 159–169
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
Schroeder, H. E. (1982) J. Sci. Food Agric. 33, 623–633
Schroeder, H. E. (1984) J. Sci. Food Agric. 35, 191–198
Sharief, F. S., and Li, S. S.-L. (1982) J. Biol. Chem. 257, 14753–14759
Shewry, P. R., Lafiandra, D., Salcedo, G., Aragoncillo, C., Garcia-Olmedo, F., Lew, E. J.-L., Dietler, M. D., and Kasarda, D. D. (1984) FEBS Lett. 175, 359–363
Spencer, D., and Higgins, T. J. V. (1980) Biochem. Int. 1, 502–509
Spencer, D., Higgins, T. J. V., Button, S. C., and Davey, R. A. (1980) Plant Physiol. 66, 510–515
Spencer, D., Chandler, P. M., Higgins, T. J. V., Inglis, A. S., and Rubira, M. (1983) Plant Mol. Biol. 2, 259–267
Valdeouze, P., Bergeron, E., Gabotit, T., and Delort-Laval, J. (1980) Can. J. Plant Sci. 60, 695–701
White, B. A., and Bancroft, F. C. (1982) J. Biol. Chem. 257, 8569–8572
Wood, E. F., and Inglis, A. S. (1984) Int. J. Biol. Macromol. 6, 277–283
Yole, R. J., and Huang, A. H. C. (1981) Am. J. Bot. 68, 44–48