Characterization and Sequence Analysis of a Developmentally Regulated Putative Cell Wall Protein Gene Isolated from Soybean*

Jong Chan Hong, Ron T. Nagao, and Joe L. Key‡
From the Department of Botany, University of Georgia, Athens, Georgia 30602

(Received for publication, November 17, 1986)

A cDNA clone, pTU04, which hybridizes to two different sizes of mRNA on Northern blots was isolated from soybean suspension culture cell poly(A) RNA. Northern analysis reveals that meristematic tissue produces a 1050-nucleotide mRNA while quiescent mature cells produce primarily a 1220-nucleotide mRNA homologous to pTU04. The cDNA and its corresponding genomic clone have been partially characterized. The nucleotide sequence of the gene predicts a proline-rich protein, designated SbPRP1, which contains a signal peptide sequence and 43 repeats of a sequence similar to previously isolated cell wall proteins: 1) it is very basic with a high content of Pro, Tyr, and Lys; 2) it has similar hydropathic properties; and 3) its repeating unit shares sequence homology with that of more highly characterized cell wall proteins, generally termed extensin (Chen, J., and Varner, J. E. (1985) EMBO J. 4, 2145–2151; Smith, J. J., Muldoon, E. P., Willard, J. J., and Lamport, D. T. A. (1986) Phytochemistry 25, 1021–1030).

Auxins are a class of naturally occurring plant growth regulators that are required for normal growth and development of higher plants, including cell differentiation, cell extension, and cell division (Thimann, 1969; Meins, 1977). Although the mechanism underlying auxin action is not fully defined, one early manifestation of auxin treatment is the regulation of gene expression at the transcriptional level (see review by Guilfoyle, 1986). The effects of auxin on the expression of specific poly(A) RNA sequences have been studied using cloned cDNAs isolated from the excised elongating soybean hypocotyl (Walker and Key, 1982), intact soybean hypocotyl (Baulcombe and Key, 1986; Hagen et al., 1984), and pea epicotyl (Theologis et al., 1985). RNA blot hybridization analyses using these cloned cDNAs have provided a direct demonstration that auxin rapidly mediates mRNA accumulation and alters the concentration of a few poly(A) RNA sequences selectively and dramatically in tissues undergoing both cell elongation and cell division (Baulcombe and Key, 1980; Walker and Key, 1982; Hagen et al., 1984; Theologis et al., 1985). Transcription studies using nuclei isolated from control and auxin-treated tissue indicate that altered poly(A) RNA levels relate to altered rates of transcription (Hagen and Guilfoyle, 1985).

Despite the extensive studies toward the elucidation of the mechanism of auxin action on plant growth and development, there is little information on the identities and the roles of the polypeptides encoded by auxin-regulated genes. Since auxin is involved in the enhancement of cell elongation and enlargement, possible candidates responding to auxins would be genes encoding polypeptides involved in cell wall metabolism. Auxin also activates the proton pump in the plasma membrane resulting in wall acidification (Rayle and Cleland, 1977). The lowered pH of the wall is thought to result in wall loosening by reversible breakage of some acid-labile, cell wall cross-links which permits a limited amount of wall extension, driven by turgor pressure (Cleland, 1971).

The long term auxin effects in promoting growth may require a continuous supply of wall materials. Thus, cell wall proteins which comprise about 10% of cell wall mass (McNeil et al., 1984) may be one class of polypeptides that respond to auxin.

Previously an auxin-responsive clone, pTU04, was isolated from suspension cultured soybean cells. The cDNA library was made from poly(A) RNA isolated from soybean suspension cells grown in the presence of auxin. Colonies were screened with 32P-labeled cDNAs made from RNA of + and − auxin-treated soybean suspension cells. The resulting colonies which showed differential hybridization were isolated. Northern blot analysis of poly(A) RNA extracted from cell cultures at low cell density or cultures grown with auxin showed accumulation of a sequence estimated to be 1050 nucleotides in length while cell cultures grown at high density or auxin-depleted cultures produced a longer mRNA estimated to be 1220 nucleotides when hybridized to this cDNA. These same auxin-responsive sequences were also expressed in the soybean hypocotyl. A switch in mRNA size occurred during cell growth and maturation in the intact seedling. This switch in mRNA size during normal and hormone-induced growth transitions and the possible function of encoded proteins have been our major interests in the study of changes in pTU04-related sequences.

Here we report that the pTU04 clone which is differentially expressed during growth of the soybean hypocotyl probably

*This work was supported in part by the National Institutes of Health Grant GM30317 (to J. L. K.) and the McKnight Foundation Award for interdisciplinary research in plant biology. 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.

The nucleotide sequence(s) reported in this paper have been submitted to the GenBank™/EMBL Data Bank with accession number(s) J02746.

‡To whom reprint requests should be addressed.

1 T. H. Ulrich, P. K. Kroner, and J. L. Key, manuscript in preparation.

8367
encodes a cell wall protein which may be important to cell maturation. The nucleotide sequence predicts a proline-rich protein made up primarily of 43 repeating units of Pro-Pro-Val-Tyr-Lys with substitutions for Val(Ile) and Tyr(Val) in a small number of the repeats. From the DNA sequence analyses and predicted amino acid sequence, this soybean proline-rich protein gene (SbPRP1) shares considerable homology at the nucleotide sequence level with the known cell wall protein, extensin, with lesser homology being noted at the amino acid level.

MATERIALS AND METHODS

Plant Material—Soybeans (Glycine max, cv Wayne) were planted in moist vermiculite and grown for 4 days at 28-30 °C in the dark. The apical section and mature sections of the hypocotyl (0.5-cm long apical hook and from 1.5-cm below the cotyledon to the root, respectively) were harvested and used for the extraction of RNA.

Isolation of Poly(A) RNA—150 g of mature sections of soybean hypocotyl and the corresponding amount of apical sections were homogenized with 3 volumes of detergent buffer containing 10 mM Tris-HCl (pH 8.5), 50 mM NaCl, 6% (w/v) p-aminosalicylic acid, 2% (w/v) trisopropylphenylarsenic acid, and 6% (v/v) 1-butanol using a Waring® blender. The homogenates were centrifuged at high speed for 50-60 s. The resulting homogenates were extracted twice with an equal volume of phenol/chloroform/isooamyl alcohol (25:24:1, v/v/v) by vigorous shaking. The final aqueous phase was ethanol precipitated twice and salt (to 3 M NaCl) precipitated on ice overnight (Aviv and Leder, 1972). Total RNA was passed through a phosphocellulose column at high speed to remove proteins before ethanol precipitation.

The resulting homogenates were transferred with an equal volume of phenol/chloroform/isooamyl alcohol (25:24:1, v/v/v) by vigorous shaking. The final aqueous phase was ethanol precipitated twice and salt (to 3 M NaCl) precipitated on ice overnight (Aviv and Leder, 1972). Total RNA was passed through a phosphocellulose column at high speed to remove proteins before ethanol precipitation.

Hybrid-selected Translation Assay—The 10 µg of pTU04 plasmid DNA was denatured in 0.1 N NaOH, neutralized with 1 M Tris-HCl (pH 8.5), and mixed with 20-30 µL of chloroform in a small glass vial. The mixture was incubated for 1 h at 50 °C in 65% (v/v) deionized water containing DNA were excised using a single-hole paper punch. SDS, and 100 µg/ml wheat germ tRNA (Sigma). Hybridization was carried out overnight at room temperature, and once at 55 °C with 0.3 M salt (to 3 M NaCl) precipitated on ice overnight (Aviv and Leder, 1972). The eluates were adjusted to 0.1 M sodium phosphate, 1.5 µg of creatine phosphokinase, 80 µg of bovine serum albumin, 20 µg of RNase inhibitor, 0.4 µg of creatine phosphate, 0.4 mM Tris, 0.4 mM magnesium acetate, 150 mM KCl, 2 mM dithiothreitol, 20 µM of 19 unlabeled amino acids (methionine being excluded), 2 µCi/ml [3H]methionine (1200 Ci/mmol), 11.2 µM of hybrid-selected poly(A) RNA, and 15.5 µM of micrococcal nuclease treated wheat germ extract S30 system prepared as described by Roberta and Paterson (1973). The translation assay contained, in a final volume of 26.3 µl, 20 mM HEPES (pH 7.5), 1.2 mM ATP, 80 µM GTP, 9.6 mM creatine phosphate, 1.5 mM creatine phosphate, 0.15 mM NaCl, 0.02% (w/v) polyvinylpyrrolidone 300, 100 µg/ml tRNA, 100 µg/ml sheared salmon sperm DNA, and 0.2% (w/v) SDS at 42 °C for 18-24 h as previously described (Baulcombe and Key, 1980). After hybridization, filters were washed three times at room temperature for 10 min in 2 X SSC, 0.1% SDS, followed by two washes in 2 X SSC, 0.1% SDS at 60 °C and one wash in 0.2 X SSC, 0.1% SDS at 60 °C for 15 min each.

Southern Blot and Copy Number Reconstruction Analysis—Soybean nuclear DNA was isolated from crude nuclei prepared from etiolated soybean (G. max cv Wayne) hypocotyl (Nagao et al., 1981). 10 µg of soybean genomic DNA was digested with the appropriate restriction endonuclease and separated by electrophoresis on 6% agarose gels in Tris borate buffer (pH 8.3) (Maniatis et al., 1982). The restriction fragments were transferred to nitrocellulose paper according to Thomas (1980) and baked for 2 h at 80 °C. The blots were prehybridized and hybridized in solution of 0.5% (v/v) formamide, 5 X SSC, 50 mM Na phosphate, 0.1% (w/v) gelatin, 0.1% (w/v) Ficoll 400, 0.1% (w/v) polyvinylpyrrolidone 300, and 200 µg/ml tRNA (Sigma) at room temperature for 24 h. The blots were hybridized in the presence of 2 X SSC, 0.1% SDS, followed by two washes in 2 X SSC, 0.1% SDS at 60 °C and one wash in 0.2 X SSC, 0.1% SDS at 60 °C for 15 min each.

Library Screening and DNA Subcloning—A soybean (G. max cv Wayne) genomic library (provided by J. Slightom, Agrigenetics Advanced Science Co.) constructed by cloning Sau3AI partially digested soybean DNA into Charon 35 (Loenen and Blattner, 1983) was screened as described by Hanahan et al. (1981). Nick-translated pTU04 insert probe. The size of the soybean genome is approximately 1.3 X 10^6 base pairs/haploid genome equivalent (Gurley et al., 1979), and the size of the plasmid DNA is approximately 6.6 X 10^6 base pairs. Thus, 50 pg of pSAx421 DNA is approximately 1 X 10^6 base pairs. This condition was the same as for the Northern and Southern blotting described above.

The isolated plasmid DNA was digested with the appropriate restriction enzymes, ligated into phage vectors, and transformed into Escherichia coli strain JM101 by a competent cell technique (Hanahan, 1983). The transformants were selected by a colony hybridization technique (Maniatis et al., 1982).

DNA Sequencing—Fragments for sequence analysis were generated by digestion with restriction endonucleases, treating with alkali, alkaline phosphatase, labeling with 32P-UTP, separation of DNA by agarose gel electrophoresis, and cutting and strand-separating (Maxam and Gilbert, 1980). Fluorescent labeling was achieved by nick-translated inserts of pTU04 probe. Approximately 3 X 10^6 recombinant phage (about 3 genomic equivalents) from the soybean library were plated on Escherichia coli strain KB 9022 rec- cells at a density of 1.5 X 10^6 phage/100-mm square Petri dish. After three cycles of plaque purification, the purified A phage clone was amplified to isolate the phage DNA as described by Madjar and Traut (1980). The hybridization condition was the same as for the Northern and Southern blotting described above.

The isolated phage DNA was digested with the appropriate restriction enzymes, ligated into phage vectors, and transformed into E. coli strain JM101 by a competent cell technique (Hanahan, 1983). The transformants were selected by a colony hybridization technique (Maniatis et al., 1982).

DNA Sequencing—Fragments for sequence analysis were generated by digestion with restriction endonucleases, treating with alkali, alkaline phosphatase, labeling with 32P-UTP, separation of DNA by agarose gel electrophoresis, and cutting and strand-separating (Maxam and Gilbert, 1980). Fluorescent labeling was achieved by nick-translated inserts of pTU04 probe. Approximately 3 X 10^6 recombinant phage (about 3 genomic equivalents) from the soybean library were plated on Escherichia coli strain KB 9022 rec- cells at a density of 1.5 X 10^6 phage/100-mm square Petri dish. After three cycles of plaque purification, the purified A phage clone was amplified to isolate the phage DNA as described by Madjar and Traut (1980). The hybridization condition was the same as for the Northern and Southern blotting described above.

The isolated phage DNA was digested with the appropriate restriction enzymes, ligated into phage vectors, and transformed into E. coli strain JM101 by a competent cell technique (Hanahan, 1983). The transformants were selected by a colony hybridization technique (Maniatis et al., 1982).
According to Maniatis et al. (1982), the 5'-end of the transcript for the pTU04 gene was determined using a polynucleotide kinase-labeled RsaI restriction fragment followed by secondary AccI digestion. The fragment labeled only at the RsaI 5'-end is 396 bp long and designated RsaI*/AccI. A second mapping fragment was constructed by digestion with HpaII, 5'-end labeled, and redigested with BglII to give the 684-bp HpaII*/BglII fragment. These two fragments covered the sequences encoding the amino-terminal 53 and 91 amino acids, respectively. The 3'-end of the transcript was determined using the 475-bp pTU04 gene was determined using a polynucleotide kinase-labeled RsaI restriction fragment followed by secondary AccI digestion. The nucleic acid pellets were ethanol precipitated, and the nucleic acid pellets were dissolved in 30 μl of hybridization buffer (80% formamide, 0.4 M NaCl, 40 mM PIPES (pH 6.4), and 1 mM EDTA). The samples were incubated at 75 °C for 15 min then quickly transferred to the hybridization temperature (45 °C) for 3 h. While keeping the samples at hybridization temperature, 300 μl of ice cold nuclease S1 buffer (0.28 M NaCl, 4.5 mM ZnSO4, 0.05 M NaOAc (pH 4.6), and 20 μg/ml denatured salmon sperm DNA) containing 300 units (BRL units) of nuclease S1 was added and the tubes were quickly vortexed and incubated 30 min at 37 °C. The reaction was stopped by the addition of 50 μl of 4.0 M ammonium acetate and 0.1 M EDTA. The RNA/DNA hybrids were recovered by ethanol precipitation using 20 μg/ml tRNA as carrier. The precipitated nucleic acids were then prepared, as for sequencing, and loaded adjacent to the appropriate fragment's sequencing sample on a 6% sequencing gel.

Analysis of DNA Sequences—Computer analyses of the DNA sequences were performed using the programs made available by Pustell and Kafatos (1982) which were adapted for use on a Hewlett-Packard HP 1000 computer by M. Clegg and J. McLendon (Dept. of Botany, University of Georgia). Hydropathic values were calculated according to Kyte and Doolittle (1982).

RESULTS

The Soybean SbPRP1 Gene Is Developmentally Regulated—Based upon physiological activity, the soybean hypocotyl can be subdivided roughly into three zones representative of different stages of cell development: apical, elongating, and mature. When poly(A) RNA from various zones of the soybean hypocotyl were compared by Northern blot hybridization analysis, dramatic changes in the apparent size of poly(A) RNA hybridizing to the cDNA probe were observed. These analyses showed that a smaller RNA (1050 nucleotides) hybridized to pTU04 when poly(A) RNA from the apical and elongating regions of soybean hypocotyl were analyzed (Fig. 1, A and E) while a larger mRNA (1220 nucleotides) was present primarily in the mature region of the hypocotyl (Fig. 1M). In addition, a minor band of shorter length (approximately 850 nucleotides) that hybridized weakly to pTU04 was also present primarily in the mature region of the hypocotyl (see “Discussion”). Treatment of the mature region with auxin leads to the switching in the size of the hybridizing RNA from the larger 1220-nucleotide form to the smaller 1050-nucleotide message. The start of cell division in the normally quiescent mature section after auxin treatment coincides with the appearance of the shorter mRNA in this mature region. Together these phenomena suggest that the presence of the shorter mRNA is associated with tissues engaged in active cell growth, whereas the larger mRNA is associated with more mature, fully enlarged relatively quiescent cells.

Isolation of λ Clones Carrying Genomic Fragments Homologous to pTU04—A soybean DNA genomic library was screened for genes homologous to the cDNA insert of pTU04. Approximately 3 × 10⁶ recombinant phage were screened, and three different clones that hybridized strongly to the pTU04 cDNA insert were isolated. One positive clone, λSAx42, which carries a 12.8-kb insert, showed similar Southern blot hybridization patterns to genomic DNA (Fig. 3) and was selected for further characterization. Southern blot hybridization analyses showed that sequences homologous to pTU04 were located within a 3.8-kb EcoRI fragment and a 3.0-kb SstI fragment (Fig. 2). The 3.8-kb EcoRI fragment and 3.0-kb SstI fragment were subcloned into pUC18 giving rise to the plasmids pSAx421 and pSAx422, respectively. A 1.7-kb BglII fragment from these subclones was the only region that hybridized to this cDNA. The entire 3.0-kb SstI fragment of pSAx422 was sequenced.

Determination of Gene Number—As a preliminary approach to estimate the gene number involved in the production of mRNA homologous to pTU04, Southern analyses of genomic DNA were performed using pTU04 cDNA insert as a probe. Digestion of soybean nuclear DNA with EcoRI or BglII yielded primarily one fragment of 4.0 and 1.7 kb, respectively, which hybridized with pTU04 (Fig. 3). The 1.7-kb BglII restriction fragment from the λ genomic clone λSAx42 matched well with the corresponding restriction digestion for the genomic DNA. The difference in size between the 3.8-kb EcoRI fragment of the λ clone and the 4.0-kb fragment from the genomic digest is probably due to Sau3AI cloning truncation of the genomic fragment which is excised at the EcoRI linker of the right arm of Charon 35 (see Fig. 2). Two additional minor fragments of approximately 6.2 and 10.5 kb on BglII digests and about a 10-kb fragment on EcoRI digests with lesser intensity of hybridization were observed. This suggests the possibility that cDNA pTU04 is a member of a small multigene family with the larger BglII and EcoRI hybridizing fragments representing cross-hybridization to other less homologous family members (see “Discussion”). For an estimation of gene copy number, copy number reconstruction experiments were performed using the BglII restriction fragment. 10 μg of soybean genomic DNA, which corresponds to 7 × 10⁶ copies of the soybean haploid genome (1.3 × 10⁶ nucleotides/ haploid genome; Gurlay et al., 1979) and 50 pg of pSAx422 plasmid DNA, equivalent in number of copies, were digested with BglII and loaded in adjacent lanes as a 1-copy unit. These experiments (Fig. 3) demonstrate that soybean proline-rich protein gene (SbPRP1) which is responsible for pTU04 is present at 1 or at most 2 copies/haploid genome equivalent.

SbPRP1 Encodes Only 1220-Nucleotide mRNA—The nucleotide sequences of cDNA pTU04 and its corresponding gene SbPRP1 were determined by the Maxam and Gilbert DNA sequencing method (1980). As shown in Fig. 4, the
Gene for a Soybean Proline-rich Protein

Fig. 2. Restriction endonuclease maps of the λ clone λSAx42, cDNA, and corresponding subclones. The solid boxes illustrate regions hybridizing to 32P-labeled cDNA. The coding region, transcription unit, and sequencing strategies are denoted. Since the cDNA sequence was identical to the genomic DNA, the sequence of Accl genomic fragment within the coding region was confirmed by cDNA sequence. Numbering refers to the position of nucleotides with regard to the CAP site as +1.

Fig. 3. Southern hybridization analysis and copy number reconstruction. Soybean genomic DNA (10 μg) was digested with EcoRI (lane 1) or BglII (lane 2) and separated on a 0.8% agarose gel next to 1, 2, and 4 copy number equivalents of BglII-digested pSAx421 subclone (lanes 3, 4 and 5, respectively). The hybridization probe was 32P-labeled PTU04 cDNA insert.

cDNA sequence corresponds to nucleotide positions 103–1019 of the genomic clone. The cDNA sequence is identical to and co-linear with the nucleotide sequence of the BglII genomic fragment (Fig. 2) and appears to contain 43 15-bp direct repeats in the coding region. The homology between the 3′-ends of the genomic sequence and the cDNA stops within a region of 13 adenosines in the genomic sequence with the cDNA having an additional 58 adenosines as described in Fig. 2.

To define the transcription unit of the SbPRP1 gene, nuclease S1 protection analysis was performed using two different DNA fragments that span up to 295 bp downstream into the coding sequence. End-labeled fragments Rsal*/Accl (396 bp) and HpaII*/BglII (696 bp) (see Fig. 2) were generated as described under “Materials and Methods” and each was hybridized to 10 μg of poly(A) RNA isolated from apical or mature hypocotyl tissue. Although various conditions were used, only RNA isolated from mature sections yielded a nuclease S1 protected band to both fragments (Fig. 5). The size of the protected fragment indicated the same RNA terminator for both fragments and predicts that the 5′-end of the mature section mRNA is located 20 bp upstream from the first ATG initiation codon (nucleotide +1, Fig. 3). An identical result was obtained with primer extension experiments which showed that only RNA unique to the mature section resulted in elongation of the cDNA primer (data not included).

Based upon comparison between the cDNA and the genomic sequences, it was originally assumed that the 3′-end of the transcription unit was near a region of 13 adenosines in the genomic sequence (Fig. 4) since the cDNA homology ended within these adenosines with an additional 58 adenosines presumably occurring as a result of polyadenylation of the RNA transcript (Fig. 2). However, a 3′-end labeled HpaII*/BglII fragment protected nuclease S1 further downstream into the genomic DNA beyond the region of 13 adenosines. Three major protected sites were observed at 1076, 1078, and 1080 only when RNA of mature hypocotyl was used (Fig. 6). From the 5′- and 3′-end nuclease S1 protection analyses, we conclude that the SbPRP1 gene gives rise only to the mature 1220-nucleotide mRNA.

The 5′-flanking region contains a TATAAAAA motif (-21 to -28) that resembles the Goldberg-Hogness TATA box (Efstratiadis et al., 1980). At 86 bp upstream from the putative CAP site, a GTCAAAT sequence is present that may be related to the CAAT box sequence GGC/tCAATCT (Shenk, 1981). The 3′-end nontranslated region contains a typical eukaryotic polyadenylation recognition signal TATAAA at positions 1048–1053 (Efstratiadis et al., 1980). Additional sequences of note include six sequences that are homologous (8–10 out of 12 bp) to the core activator enhancer sequence of eukaryotic viruses (TGGAAG) (see for review, Lamins et al., 1983) observed in the 5′ upstream and 3′ downstream regions of the gene (Fig. 4) and an ATT sequence repeated 19 times about 1-kb upstream of the CAP site.

The SbPRP1 Gene Encodes a Proline-rich (or Hydroxyproline-rich) Protein—The deduced amino acid sequence of the translated SbPRP1 gene is shown in Fig. 4. Two translation initiation codons corresponding to two open reading frames (ORF) were deduced from the genomic DNA sequence. In ORF1, translation is assumed to start at the first ATG (nucleotides 21–23) and terminate after 768 nucleotides at position 8370.
The nucleotide sequence and the predicted amino acid sequence of the ShbPRP1 gene. The cDNA sequence starts at position 103 and ends at 1019 as depicted by Fig. 2. Two possible open reading frames (ORF1 and ORF2) of the ShbPRP1 gene are shown. The first 26 amino acids, representing a putative signal sequence are bracketed. Short single lines indicated homologous regions to viral enhancer-like sequence (see for review, Lamins et al., 1983). Long single underline denotes an ATT sequence repeated 19 times. The CAT box, TATA box, and polyadenylation signal are double underlined. Transcription termination sites are indicated with bold arrowheads.
Gene for a Soybean Proline-rich Protein

Fig. 5. Nuclease S1 mapping of the 5' terminus of the soybean SbPRP1 gene. A. 32P-labeled RsaI or BglII fragments were hybridized without poly(A) RNA (C), with poly(A) RNA isolated from apical sections (A), or with poly(A) RNA isolated from mature sections (M). The RsaI*/AccI protected fragment is indicated by the arrow. B, the RsaI*/AccI protected fragment from A was electrophoresed next to a sequencing ladder of the corresponding DNA. The protected fragment is indicated by the arrow.

DISCUSSION

Northern analysis using pTU04 cDNA as a probe demonstrated that the expression pattern of two major messages homologous to this probe in soybean hypocotyl changed as the development of the hypocotyl progressed from regions of active growth to the region of relatively quiescent, mature cells. The presence of a smaller 1050-nucleotide mRNA in the apical and elongation regions while the longer 1220-nucleotide mRNA appeared in the mature region of hypocotyl (Fig. 1) is consistent with the concept of an auxin gradient existing within the hypocotyl; that is, a high auxin concentration associated with the rapidly growing tissues to lower or depleted levels of auxin in the mature tissue. The SbPRP1 gene is responsive to auxin or growth changes induced by auxin; treatment of the quiescent mature region with auxin causes resumption of cell division and a corresponding switch to the shorter message. Likewise, inhibition of growth in the apical hypocotyl by auxin (Key et al., 1966) results in accumulation of the larger form mRNA. That the larger form message is associated with the cell maturation process also is suggested by data which shows the accumulation of the larger form mRNA as excised elongating hypocotyl tissue approaches the fully elongated state during auxin-induced cell elongation (data not shown).

The two major transcripts homologous to the pTU04 cDNA...
may be accounted for in two ways. One mechanism is that two structurally different but homologous genes are differentially expressed (Fyrberg et al., 1983) depending upon the stage of the soybean hypocotyl development. Another mechanism is that a single gene makes two RNAs either by alternate splicing (Medford et al., 1984) or by alternative transcription initiation or termination (Nabeshima et al., 1984; Young et al., 1981). The latter possibilities are not consistent with the structural analyses of the gene performed by nuclease S1 protection experiments. A third smaller hybridizing RNA was also observed at lower levels on Northern blots of poly(A) RNA isolated from the mature region (Fig. 1); this RNA appears to be the product of a related but different gene which also has been sequenced (data not shown).

Genomic Southern blot analyses revealed a major single band hybridizing to pTU04 cDNA of 1.7 kb on BgII digests and 3.8 kb on EcoRI digests and additional minor larger bands. The isolated λ clone λSAx42 showed a similar restriction enzyme digestion pattern to genomic Southern blots which is suggestive that this λ clone contains the gene responsible for at least one of two major messages. Sequence data of cDNA and genomic DNA λSAx42 further confirmed that the SbPRP1 gene is located within the EcoRI and BgII

**Fig. 6.** S1 mapping of transcription termination site. 3′-end labeled HpaII*/BglII fragment representing the 3′ portion of gene was hybridized to mature poly(A) RNA. Three major protected bands are indicated by arrowheads.

**Fig. 7.** Hydropathic profiles of deduced amino acid sequences of soybean SbPRP1 ORF1 (A), carrot 33-kDa PRP (B) (Chen and Varner, 1985a), and carrot extensin cell wall protein (C) (Chen and Varner, 1985b). Plots were constructed by the method of Kyte and Doolittle (1982) by progressively moving along the amino acid sequence and averaging the hydropathy index for 9 amino acids. Points above the horizontal line correspond to hydrophobic region, and points below this line are hydrophilic. Data for B plotted from a partial cDNA clone.

**Fig. 8.** In vitro translation and analysis of SbPRP1 hybrid-selected RNA. pTU04 cDNA plasmid was immobilized on nitrocellulose filters and hybridized to poly(A) RNA isolated from apical (A) or mature (B) sections of soybean hypocotyl. Hybridized mRNA was eluted and translated in vitro. 35S-Labeled translation products were analyzed by two-dimensional electrophoresis using acid-urea/SDS polyacrylamide gel electrophoresis. The hybrid-selected translation product is circled.
sites of the map (Fig. 2). However, 5'- and 3'-end S1 protection experiments demonstrated that only mature 1220-nucleotide RNA is derived from SbPRP1 gene; the apical mRNA gave no protection in these experiments. Although the 650-nucleotide mRNA homologous to pTU04 is observed in RNA isolated from mature tissue, this message is not homologous enough to compete with 1220-nucleotide RNA during nuclease S1 protection experiments.3

As described before, cDNA pTU04 contains a highly repeated structure in the coding region. The repeated nature of the nucleotide sequence may allow hybridization of other messages that share homology with this repeating structure. Partial sequence data of a gene encoding the 650-nucleotide RNA which weakly hybridizes to pTU04, showed a basic 15-bp repeating structure over a short portion of the coding region and more variation in the sequence than in the SbPRP1 gene (data not shown). Based upon the strong hybridization of cDNA pTU04 to the apical 1050-nucleotide message on Northern blots, the 1050-nucleotide RNA is assumed to have a repeating structure similar to pTU04 at least over a part of the sequence. Although an apical gene for the 1050-nucleotide message has not been identified, minor larger bands observed on Southern blot may constitute additional related genes; one or more of these might correspond to the transcription unit for the apical message poly(A) RNA of 1050-nucleotides in length which hybridizes to pTU04 on Northern blots but which does not give a S1 protected fragment upon S1 mapping. These data indicate that the 5'-end of the transcripts for the 1050- and 650-nucleotide long RNA are not sufficiently homologous to the isolated genomic sequence to provide protection from S1 nuclease digestion. This observation coupled with the Northern analyses indicates that there are at least three different genes encoding these three poly(A) RNAs.

The SbPRP1 gene is similar to other eukaryotic genes in that the 5'-flanking regions contain a Goldberg-Hogness TATA box and a CAAT box, and the 3'-flanking region contains consensus AAATAA polyadenylation signal.

The coding region of the gene is basically composed of a tandemly repeated 15-bp unit which suggests a 5-amino acid repeat upon translation. When this repeat is read in the three possible frames, one reading frame (ORF1) showed a high degree of codon degeneracy within the third base resulting in an extremely conserved amino acid repeating pattern predicting a basic peptide of approximately 29 kDa (Fig. 4). A second possible reading frame (ORF2) is also indicated in Fig. 4. This ORF has a more variable amino acid repeat unit, with a termination codon occurring before the end of the repeating structure; the resulting peptide has a predicted molecular weight of approximately 22,000. Two additional open reading frames are present on the complementary strand; however, both reading frames on the complementary strand do not contain an ATG translation initiation codon and are judged to reflect a coincidence of the repeat nature of the coding strand. A similar result is observed in the carrot extensin gene which has two ORFs on the coding strand and one on the complementary strand (Chen and Varner, 1985b). The multiple reading frames of SbPRP1 gene are likely the consequence of the highly regular repeat nature of the DNA sequence. ORF1 of the SbPRP1 gene predicts a proline-rich (likely becomes hydroxyproline-rich in the mature protein), highly basic protein composed primarily of 43 repeating units of Pro-Pro-Val-Tyr-Lys and a putative signal sequence of 26 amino acids. The hybrid select translation analyses produced a basic protein of 28,000–29,000 on an acid-urea/SDS-urea two-dimensional gel system, consistent with ORF1 being the appropriate reading frame.

While the structure and function of the gene product is unknown, the proline-rich protein (SbPRP1) predicted from the DNA sequence may be converted into hydroxyproline-rich glycoprotein (HRGP) by post-translational modification. Several types of HRGPs are known to occur in higher plants: e.g. cell wall protein extensins (Smith et al., 1986), arabinogalactan proteins (Fincher et al., 1983), solanaceous lectins and agglutinin (Lamport and Catt, 1981). Although these HRGPs are rich in hydroxyproline and serine, each HRGP is clearly distinguishable due to a great difference in the content of cysteine and alanine and different solubility characteristics. Potato lectin is rich in cysteine while arabinogalactan-containing proteins are highly acidic and alanine-rich. Extensins are very basic and highly insoluble proteins rich in histidine, valine, tyrosine, and lysine in addition to hydroxyproline and serine (Smith et al., 1984; Stuart and Varner, 1980; Lamport, 1966). Extensins are presumed to have a highly periodic peptide backbone structure, and the characteristic repeating pentapeptide Ser-Hyp-Hyp-Hyp-Hyp has been identified so far as the major constituent of extensins (Lamport, 1977; Smith et al., 1986; Chen and Varner, 1985b; Showalter et al., 1985).

The presence of a signal sequence on this Pro-, Tyr-, and Lys-rich SbPRP1 protein is suggestive that it may be a secreted extracellular, cell wall protein. A comparison of the hydropathy profiles of this soybean gene (Fig. 7A) with carrot 33-kDa proline-rich protein (PRP) derived from partial cDNA sequence analyses (Fig. 7B) and carrot extensin gene pDC5A1 (Fig. 7C) shows that each protein is very hydrophilic, has an NH2-terminal hydrophilic region, and has a repeat structure (Chen and Varner, 1985a, 1985b). It should be emphasized that hydropathic comparisons such as these are intended to illustrate the similarities of primary amino acid sequence only and not intended to imply functional similarities since it is known that proline-rich proteins undergo extensive modification in vivo (see below).

A comparison of the amino acid composition for the predicted SbPRP1 (ORF1) with the predicted extensin and 33-kDa PRP from carrot and the analyses of the purified extensin protein from carrot and tomato are presented in Table I. Some major similarities include a very high proline content and relatively high lysine and tyrosine contents. A major difference in the amino acid composition of SbPRP1 with known extensin proteins is the extremely low content of serine and the absence of histidine in SbPRP1. While the serine content of the 33-kDa PRP of carrot is also low, the comparatively high histidine content and moderately high threonine content demonstrate that these proteins are different. Thus while major similarities are noted, major differences also indicate that SbPRP1 represents a type of proline-rich protein distinct from those previously described.

A comparison of the major unmodified amino acid repeat units of various proline-rich proteins illustrates the differences observed from the amino acid composition data (Table II). Inspection of the nucleotide sequence of the various repeats shows that the sequences are more closely related than revealed at the amino acid level. The nucleotide sequence for SbPRP1 repeat, CCA-CCA-GTT-TAC-AAG, encoding Pro-Pro-Val-Tyr-Lys is almost identical to one of the repeats for the carrot 33-kDa PRP of CCA-CCA-GTT-CAC-AAG encoding Pro-Pro-Pro-Val-His-Lys. A single T/C nucleotide conversion would convert Tyr (TAC) to His (CAC), or vice versa, and would make the amino acid and nucleotide repeat units identical. In an analogous manner one or two nucleotide

3 J. C. Hong, R. T. Nagao, and J. L. Key, unpublished data.
convert Lys to Thr, or vice versa, again yielding identical Thr-Pro-Val-Tyr-Lys. A single nucleotide conversion in the other major carrot 33-kDa PRP repeat would change in the other major carrot 33-kDa PRP repeat would convert Lys to Thr, or vice versa, again yielding identical Thr-Pro-Val-Tyr-Lys. A single nucleotide conversion in the other major carrot 33-kDa PRP repeat would changes in the other major carrot 33-kDa PRP repeat would convert Lys to Thr, or vice versa, again yielding identical Thr-Pro-Val-Tyr-Lys. A single nucleotide conversion in the other major carrot 33-kDa PRP repeat would

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>SbPRP1* (ORF1)</th>
<th>Carrot*</th>
<th>Analysis of purified extensin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33-kDa PRP</td>
<td>Extensin</td>
<td>Tomato P1a</td>
</tr>
<tr>
<td>Pro*</td>
<td>39.6</td>
<td>32.6</td>
<td>42.3</td>
</tr>
<tr>
<td>Lys</td>
<td>19.6</td>
<td>10.2</td>
<td>11.7</td>
</tr>
<tr>
<td>Val</td>
<td>16.1</td>
<td>10.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Tyr</td>
<td>16.1</td>
<td>4.5</td>
<td>12.0</td>
</tr>
<tr>
<td>Ser</td>
<td>0.4</td>
<td>2.5</td>
<td>10.9</td>
</tr>
<tr>
<td>His</td>
<td>0.0</td>
<td>13.7</td>
<td>9.5</td>
</tr>
<tr>
<td>Glu</td>
<td>3.9</td>
<td>5.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Gln</td>
<td>0.0</td>
<td>1.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Ile</td>
<td>2.6</td>
<td>3.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Thr</td>
<td>0.4</td>
<td>6.0</td>
<td>3.6</td>
</tr>
<tr>
<td>Gly</td>
<td>0.4</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Asp</td>
<td>0.4</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Asn</td>
<td>0.4</td>
<td>2.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Leu</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Cys</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Arg</td>
<td>0.0</td>
<td>1.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Met</td>
<td>0.0</td>
<td>1.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Phe</td>
<td>0.0</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>Trp</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Ala</td>
<td>0.0</td>
<td>2.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

* Amino acid composition excluding putative signal sequence.
† Chen and Varner, 1985a.
‡ Chen and Varner, 1985b.
As the total amount of proline and hydroxyproline.

There are several known amino acid sequences of proline-rich or hydroxyproline-rich proteins in animals and plants (for example, collagen (Campbell and Rosen, 1984), PRP isolated from salivary glands of various animals (Ann and Carlson, 1985), HRGP extensin (Chen and Varner, 1985b), and γ-zein of maize (Wang and Esen, 1986)). These proteins are known to have repeating structures at the amino acid and nucleotide level. When SbPRP1 protein is compared with tomato cell wall extensin and carrot 33-kDa PRP, a shared homology of Pro-Pro-Val-Tyr-Lys and CCX-CCX-GTX-TAX-AAX are noted at the amino acid and nucleotide level, respectively (Table II). Interestingly, the NH2-terminal half of γ-zein protein contains 8 repeats of Pro-Pro-Pro-His-Leu (CCX-CCX-CCG-GTX-CAX-CTG) which could be converted to the same repeat by a change of 1 or 2 nucleotides in His (CAX) and Leu (CTG) codons as previously mentioned. The presence of γ-zein in the endosperm protein bodies, bordering the inner part of the membrane of the organelle (Ludevid et al., 1984) suggests that this proline-rich or hydroxyproline-rich repeating structure may endow the protein

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>SbPRP1* (ORF1)</th>
<th>Carrot*</th>
<th>Analysis of purified extensin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33-kDa PRP</td>
<td>Extensin</td>
<td>Tomato P1a</td>
</tr>
<tr>
<td>Pro*</td>
<td>39.6</td>
<td>32.6</td>
<td>42.3</td>
</tr>
<tr>
<td>Lys</td>
<td>19.6</td>
<td>10.2</td>
<td>11.7</td>
</tr>
<tr>
<td>Val</td>
<td>16.1</td>
<td>10.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Tyr</td>
<td>16.1</td>
<td>4.5</td>
<td>12.0</td>
</tr>
<tr>
<td>Ser</td>
<td>0.4</td>
<td>2.5</td>
<td>10.9</td>
</tr>
<tr>
<td>His</td>
<td>0.0</td>
<td>13.7</td>
<td>9.5</td>
</tr>
<tr>
<td>Glu</td>
<td>3.9</td>
<td>5.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Gln</td>
<td>0.0</td>
<td>1.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Ile</td>
<td>2.6</td>
<td>3.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Thr</td>
<td>0.4</td>
<td>6.0</td>
<td>3.6</td>
</tr>
<tr>
<td>Gly</td>
<td>0.4</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Asp</td>
<td>0.4</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Asn</td>
<td>0.4</td>
<td>2.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Leu</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Cys</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Arg</td>
<td>0.0</td>
<td>1.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Met</td>
<td>0.0</td>
<td>1.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Phe</td>
<td>0.0</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>Trp</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Ala</td>
<td>0.0</td>
<td>2.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

* Amino acid sequences predicted from DNA sequence.
† Chen and Varner, 1985a.
‡ Chen and Varner, 1985b.
§ Smith et al., 1986.
with some structural role. In addition, the nucleotide homology of each repeating unit among different plant species is suggestive of a common origin.

Based on the foregoing results, the appearance of the longer message hybridizing to pTU04 in mature hypocotyl may have some relationship to the accumulation of a hydroxyproline-rich protein in cell wall of pea epicotyl associated with the cessation of elongation (Sadava and Chrispeels, 1973). However, the functional significance of the differential expression of two (or more) mRNAs homologous to pTU04 depending on the stage of soybean hypocotyl development is unknown. We have recently cloned the 1050-nucleotide apical mRNA homologous to pTU04, and characterization of the apical cDNA clone is in progress.

It should be emphasized that the SbPRP1 gene has not been shown to encode a cell wall protein. However, circumstantial evidence strongly suggests that this gene encodes a protein destined for the cell wall. Future research will involve the characterization of HHRP in the cell wall of soybean hypocotyl. Although further characterization is needed, data presented here are suggestive of a role of SbPRP1 protein in cell differentiation and maturation and that the alteration of cell wall protein gene expression in response to auxin may be important in these developmental transitions.

Acknowledgments—We wish to thank P. Kroner, C. Welchel, and G. Goekjian for technical assistance.

REFERENCES

Meins, F. (1977) Am. J. Pathol. 69, 687-702
Silflow, C. D., Hammett, R. J., and Key, J. L. (1979) Biochemistry 18, 2720-2731