The EMBO Journal vol.4 no.9 pp.2145–2151, 1985

An extracellular matrix protein in plants: characterization of a genomic clone for carrot extensin

Jychian Chen[1] and Joseph E.Varner

Department of Biology, Washington University, St. Louis, MO 63130, USA

[1]Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724,USA

Communicated by D.Northcote

Extensins are hydroxyproline-rich glycoproteins found in many plant cell walls as a major protein component. The peptide Ser-Hyp-Hyp-Hyp-Hyp is abundant in the extensins. Using extensin cDNA clones as probes, we isolated six different clones from carrot genomic libraries. One of the genomic clones, pDC5A1, was characterized and found to contain an open reading frame encoding extensin and a single intron in the 3′-non-coding region. The derived amino acid sequence contains a signal peptide sequence and 25 Ser-Pro-Pro-Pro-Pro repetitive sequences. Two extensin transcripts were found corresponding to pDC5A1 with different 5′ start sites. These transcripts increase in abundance after wounding. This is consistent with the reported extensin accumulation in the cell wall upon wounding.

Key words: hydroxyproline-rich wall protein/wounding

Introduction

At least three classes of hydroxyproline-rich glycoproteins exist in higher plants (McNeil *et al.*, 1984). These are (i) the extensins, (ii) the arabinogalactans and (iii) the solanaceae lectins. The extensins were discovered, named, and partially characterized by Lamport. Extensins are found most abundantly in the cell walls of dicotyledonous plants (Lamport, 1965). Extensins are synthesized on the endoplasmic reticulum, glycosylated in the Golgi apparatus, and secreted to the cell walls (Chrispeels and Sadava, 1974). After the protein is secreted into the cell wall space it becomes increasingly insolubilized (Cooper and Varner, 1983). Such insolubilization appears to be due, at least in part, to the formation of isodityrosine residues (IDT, two tyrosine residues covalently cross-linked by a diphenyl ether linkage) (Fry, 1982). Extensins are assumed to play a role in the structure of plant cell walls and may therefore be important in controlling growth. Extensin accumulates in plant cell walls upon wounding (Chrispeels *et al.*, 1974) and pathogen attack (Esquerre-Tugaye and Lamport, 1979), indicating that it is involved in defence.

The soluble form of extensin produced in wounded carrot root tissue has a mol. wt. of ~86 000 daltons and consists of one-third protein and two-thirds carbohydrate (Stuart and Varner, 1980). The estimated mol. wt. of the polypeptide is 30 000 daltons. Six amino acids (hydroxyproline, serine, histidine, lysine, tyrosine and valine) comprise 95% of the polypeptide backbone (Stuart and Varner, 1980). The carbohydrate moiety is composed largely of arabinose which is attached to hydroxyproline, and a small amount of galactose which is thought to be attached to serine (Van Holst and Varner, 1984; Lamport *et al.*, 1972). The carbohydrates serve to stabilize the protein into a rigid rod-like structure (Van Holst and Varner, 1984).

The determination of the amino acid sequence of extensin is difficult because of the presence of many post-translational modifications and repetitive amino acid sequences, especially of imino acids. However, partial amino acid sequences of proteolytic fragments of cell wall-bound tomato extensins have been reported (Lamport, 1977). The most significant finding from this work has been the identification of several peptides containing the pentapeptide Ser-Hyp-Hyp-Hyp-Hyp, and an intra-molecular linkage involving isodityrosine (IDT) which has the sequence 1/2 IDT-Lys-1/2 IDT-Lys (Lamport, 1977; Epstein and Lamport, 1984).

Two extensin cDNA clones (Chen and Varner, 1985), isolated from a cDNA library constructed from wounded carrot root mRNA, were used as probes to isolate extensin genomic clones. Here we describe (i) the isolation of six extensin genomic clones from carrot, (ii) the characterization of one of the genomic clones, pDC5A1 and (iii) the expression of pDC5A1 in wounded tissue.

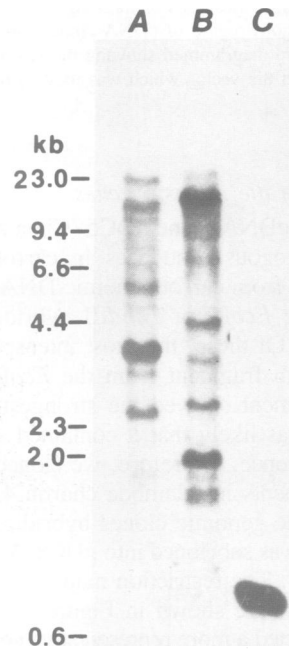

Fig. 1. Southern blot of carrot genomic DNA digested with restriction enzymes *Eco*RI or *Hind*III and hybridized to extensin cDNA clone pDC5. In **lane A** and **lane B**, 5 μg carrot DNA was digested with *Eco*RI and *Hind*III, respectively; in **lane C**, 0.1 μg pDC5 plasmid DNA was digested with *Hind*III. The insert of pDC5, which is 750 bp in size, was nick-translated and used as a probe. After hybridization, the blot was washed at 55°C with 2 × SSC. Under these conditions 30% mismatch was allowed. Markers indicate the length of DNA in kilobase pairs (kb).

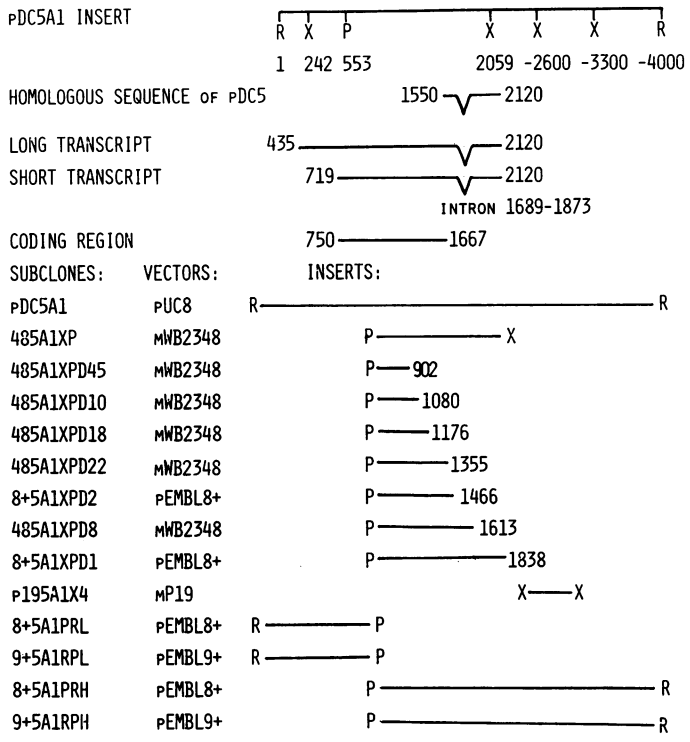

Fig. 2. Restriction map of pDC5A1 insert, containing a carrot extensin gene. Restriction endonucleases are abbreviated as follows: R: *Eco*RI, P: *Pst*I and X: *Xba*I. Relative positions of the long and short transcripts of pDC5A1, and homologous region of cDNA clone pDC5 to pDC5A1 are indicated. Subclones are diagrammed showing the region of pDC5A1 insert they contain along with the vector which was used in the construct.

Results

Genomic clones of the extensin genes

We used extensin cDNA clone pDC5 (Chen and Varner, 1985) to identify homologous sequences in carrot genomic DNA. Several fragments from carrot genomic DNA digested with the restriction enzyme *Eco*RI or *Hind*III hybridized to the pDC5 probe (Figure 1). Of these, the most intense hybridization occurred with a 4-kb fragment from the *Eco*RI-digested DNA. Because this fragment showed the strongest homology to the pDC5 probe, it was likely that it contained a large proportion of the cDNA sequence. Therefore, we cloned size-fractionated 4-kb *Eco*RI fragments into lambda charon 4A. A 4-kb *Eco*RI insert in one of the genomic clones hybridized to the extensin cDNA probe and was subcloned into pUC8 (Vieira and Messing, 1982) as pDC5A1. The restriction map and serial subclones of the pDC5A1 insert are shown in Figure 2.

We also constructed a more representative genomic library with lambda vector EMBL 3 (Frischauf *et al.*, 1983). Partially digested *Sau*3A1 DNA fragments were cloned in lambda EMBL 3. Five of the genomic clones were found to contain sequences homologous to the extensin cDNA probe by plaque hybridization. Restriction maps of these cloned DNAs and Southern blot hybridization to the pDC5A1 insert probe indicated that these clones are different from one another (data not shown).

The sequence of genomic clone pDC5A1 insert

We sequenced a 2.2-kb segment of the insert of pDC5A1, which has the homologous sequence to pDC5 (Figure 3). In this sequence 399 nucleotides (at position 1537 – 1688 and 1874 – 2120 in Figure 3) are identical to the insert of extensin cDNA clone pDC11, and 386 nucleotides (at position 1550 – 1688 and 1874 –

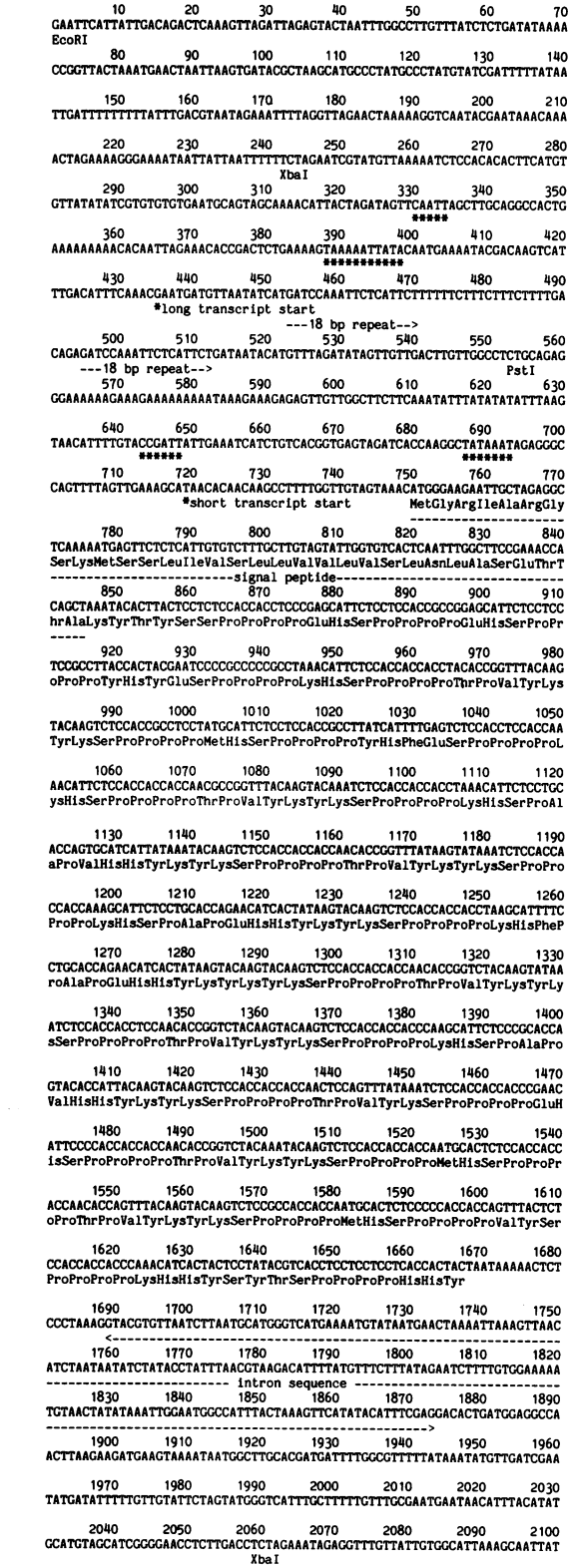

Fig. 3. DNA sequence of pDC5A1. The inserted DNA of pDC5A1 was subcloned into M13 phage vectors and sequenced according to methods described by Barnes *et al.* (1983). Only the first 2230 bp of pDC5A1 insert (4 kb) was sequenced. The start sites of the two RNA transcripts, an 18-bp direct repeat, the intron splicing sites and the polyadenylation sites are indicated. Sequences similar to the CAAT box and TATA box and polyadenylation signal sequences were underlined with **. The predicted amino acid sequence along with the putative signal peptide cleavage site is also shown.

Table I. Comparison of the amino acid composition (expressed as mol. %) between purified tomato extensin P1 (Smith *et al.*, 1984), carrot extensin (Van Holst and Varner, 1984), and that derived from the DNA sequence of the pDC5A1 insert.

Amino acid	Predicted from DNA sequence	Analysis of purified extensin	
		Carrot	Tomato P1
Proline	42.3	45.7[a]	41.8[a]
Tyrosine	12.0	11.0	8.9
Lysine	11.7	6.7	10.1
Serine	10.9	14.2	9.5
Histidine	9.5	11.8	7.1
Valine	4.0	5.9	5.0
Threonine	3.6	1.2	7.2
Glutamate	2.6	0.4[b]	1.9[b]
Glutamine	0.0	–	–
Alanine	1.4	0.4	2.0
Methionine	1.1	0.0	0.0
Phenylalanine	0.7	0.0	0.6
Tryptophan	0.0	1.2	–
Isoleucine	0.0	0.4	0.9
Aspartate	0.0	0.4[c]	1.8[c]
Asparagine	0.0	–	–
Leucine	0.0	0.4	0.8
Glycine	0.0	0.4	1.6
Cysteine	0.0	0.0	0.0
Arginine	0.0	0.0	0.7

[a]As the total amount of proline and hydroxyproline.
[b]As the total amount of glutamate and glutamine.
[c]As the total amount of aspartate and asparagine.
Because the first 32 amino acids in the derived extensin peptide sequence is likely to be signal peptide, we did not include it in the amino acid composition.

2120) are identical to the pDC5 insert. In addition, 185 nucleotides are present in the genomic clone pDC5A1 at position 1689–1873 but are not present in the cDNA clone pDC11, and may represent intron sequences. The sequence also indicates that pDC5A1 is the genomic clone corresponding to the cDNA clone pDC11.

In the DNA sequence (Figure 3) there are three open reading frames, two being on the same strand and one (270 amino acid residues) on the complementary strand. However, the latter strand is not transcribed (as described below). One open reading frame, encoding 292 amino acid residues, does not contain an ATG at the 5′ end, and is therefore unlikely to be translated. The remaining open reading frame encodes a proline-rich polypeptide (Figure 3) which contains 306 amino acids and has a mol. wt. of 34 225 daltons. In the predicted amino acid sequence, Ser-Pro-Pro-Pro-Pro, Tyr-Lys-Tyr-Lys and Thr-Pro-Val were found as major repeat units. Due to analogous amino acid sequences in purified extensin, we believe that the genomic clone pDC5A1 represents one of the genomic sequences encoding for extensin in carrot. There are, however, some differences in the amino acid composition between purified carrot extensin (Van Holst and Varner, 1984) and the composition of extensin predicted from the DNA sequence of pDC5A1 (Table I).

A putative signal peptide sequence of 32 amino acids at the amino terminus of the predicted protein encoded by the genomic clone pDC5A1 was identified (Figure 3) using the rules of Von Heijne (1983). The analysis of hydrophobicity (Kyte and Doolittle, 1982) of the amino acid sequence indicates that the whole protein is hydrophilic except for the first 32 amino acids.

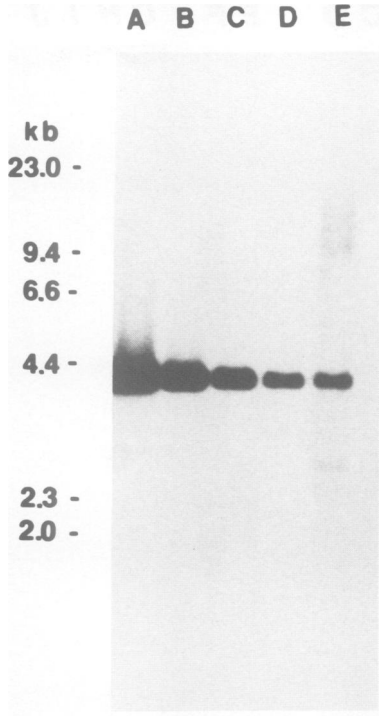

Fig. 4. DNA hybridization analysis to determine the copy number of the 4-kb fragment, the extensin DNA sequence encoded by pDC5A1 insert, in carrot genome. 10 μg of carrot DNA **(lane E)** along with 500, 250, 100 and 50 pg of pDC5A1 DNA **(lanes A–D)** were digested with *Eco*RI. The size of pDC5A1 is ~7000 bp. The nick-translated pDC5A1 insert DNA was used as a probe. After hybridization the blot was washed at 68°C with 2 × SSC. Under these conditions 20% mismatch was allowed. Taking carrot haploid genome size as 1.4×10^9 bp, a single copy equivalent of pDC5A1 DNA will be 50 pg.

Characterization of the genomic clone pDC5A1

We determined the copy number of the 4-kb *Eco*RI fragment in the carrot genome corresponding to the DNA insert of the pDC5A1 by Southern blot hybridization. Taking the carrot haploid genome size as 1.4×10^9 bp (Bennett and Smith, 1976), multiples of single copy equivalents of pDC5A1 were used to determine the number of copies of this sequence in the carrot genome by DNA blot hybridization (Figure 4). It was determined that there is only a single copy of the 4-kb *Eco*RI fragment corresponding to DNA insert of pDC5A1 per haploid genome of carrot.

To identify extensin transcripts in unwounded or wounded carrot root slices, we isolated total RNA and poly(A)$^+$ RNA from both types of tissue. RNA samples were separated in 1% formaldehyde gels, blotted to nitrocellulose filters, and hybridized to the DNA insert of pDC5A1 (Figures 5 and 8). An RNA of ~1500 nucleotides was the major species recognized by the probe pDC5A1 in RNA from both samples. A longer exposure of the autoradiograph revealed several larger transcripts as well. There was a higher level of extensin RNA transcripts detected in wounded carrot than in unwounded carrot (Figure 5). These Northern blot hybridization results indicate that extensin transcripts accumulate upon wounding. This is consistent with the result that hydroxyproline-rich glycoproteins accumulate in the cell wall upon wounding (Chrispeels *et al.*, 1974).

Analysis of the possible coding sequence from the DNA sequence data indicated that, in addition to the coding sequence

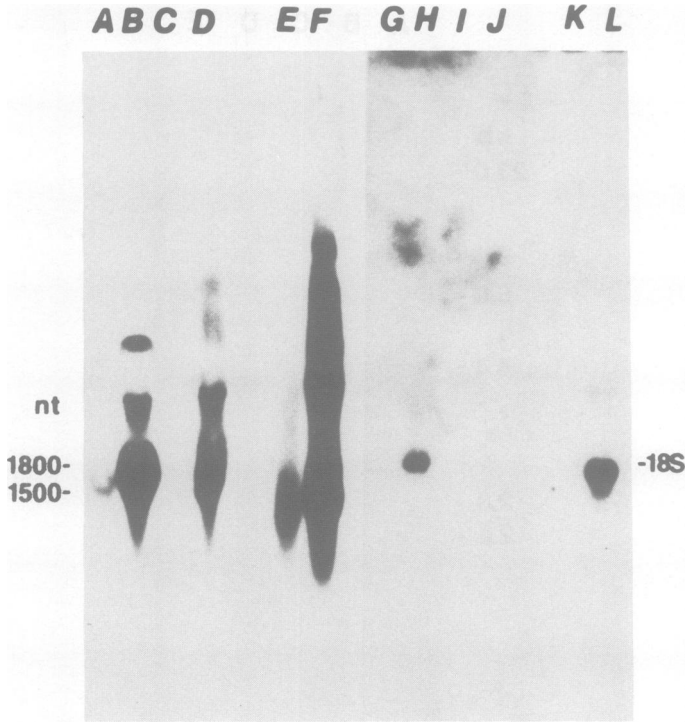

Fig. 5. Upon wounding carrot extensin RNA is accumulated as shown by hybridization of RNA isolated from unwounded **(lanes A,C,E,G,I,K)** and from wounded carrot **(lanes B,D,F,H,J,L)** with probes of pDC5A1 insert DNA **(lanes A – F)** and 8+5A1PRL **(lanes G – L)**. **Lanes A, B, G** and **H** contain 20 μg of total RNA; **lanes C, D, I** and **J** contain 20 μg of poly(A)$^-$ RNA, and **lanes E, F, K** and **L** contain 4 μg of poly(A)$^+$ RNA. The probe of pDC5A1 insert DNA was labeled by nick-translation and the probe of 8+5A1PRL was synthesized from single-stranded DNA with a sequencing primer (GTAAAACGACGGCCAGT). The RNA hybridized to the probe 8+5A1PRL has the size of 1800 nucleotides which is the long transcript of pDC5A1.

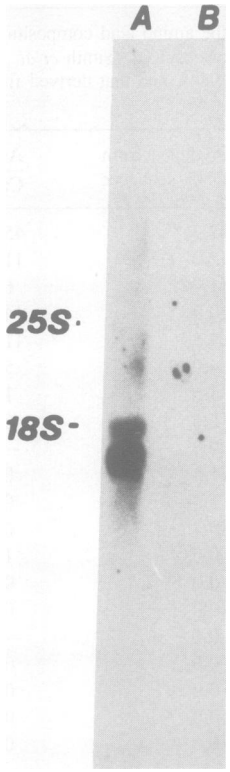

Fig. 6. Determination of which of the two strands of pDC5A1 is transcribed by RNA hybridization analysis with strand-specific probe. Strand-specific probes were synthesized from 8+5A1PRH **(lane A)** and 9+5A1RPH **(lane B)** single-stranded DNA with a hybridization primer (CACAATTCCA-CACAAC). 5 μg poly(A)$^+$ RNA isolated from wounded carrot was loaded in each sample.

for extensin, there is another possible coding sequence on the complementary strand. We determined whether both strands of pDC5A1 are transcribed by first producing strand-specific probes generated from clones 8+5A1PRH or 9+5A1RPH using a 16-mer (CACAATTCCACACAAC) as primer. Single-stranded [^{32}P]DNA was hybridized to the total RNA from the wounded tissue. Only the strand which coded for extensin was observed to be transcribed upon Northern hybridization analysis (Figure 6).

We examined the gene structure of extensin encoded by pDC-5A1 using S1 nuclease mapping. From the comparison of the cDNA sequence to the genomic DNA sequence, it appeared that there is an intron in the 3′-untranslated region. This was confirmed by S1 mapping using the DNA fragment (558 – 2064) of pDC5A1 and poly(A)$^+$ RNA from wounded tissue. The probe was synthesized from 485A1XP with the sequencing primer (GTAAAACGACGGCCAGT from New England Biolabs) as described in Materials and methods. A 191-bp fragment, which is resistant to S1 nuclease and corresponds to 1874 – 2064 of pDC5A1 sequence, was identified (Figure 7); the other S1 nuclease-resistant fragment (558 – 1688) which would contain 1131 nucleotides could not be resolved by this gel. We also used S1 mapping to identify the 5′ end of extensin transcripts by using DNA fragments (1 – 557 or 558 – 902) of pDC5A1 (synthesized with sequencing primers GTAAAACGACGGCCAGT from 8+5A1PRL and 485A1XPD45) and poly(A)$^+$ RNA from wounded carrot. Data obtained from these experiments indicate that there are two transcripts with different 5′ start sites at positions 435 and 719 (Figure 7). This result was further confirmed by using the DNA fragment (1 – 557) of pDC5A1 as a probe (synthesized with a sequencing primer GTAAAACGACGGC-CAGT from 8+5A1PRL) on Northern blots of total RNA and poly(A)$^+$ RNA from wounded tissue. Only the long transcripts (~1800 nucleotides in size) were found to hybridize to this probe (Figures 5 and 8). However, when the whole insert of genomic clone pDC5A1 or the DNA fragment (553 – 4000) of pDC5A1 (synthesized from 9+5A1RPH with a sequencing primer GTA-AAACGACGGCCAGT) were used as probes, we found both the long (1800 nucleotides) and short transcripts (1500 nucleotides) (Figure 8). When the DNA sequence of pDC5A1 was analyzed, an 18-bp direct repeat in the 5′-non-coding sequence (453 – 470 and 495 – 512) was detected. This repeat would be present in the 1800-nucleotide transcript but would not be in the 1500-nucleotide transcripts (Figure 3).

Discussion

The cell walls of higher plants are made up of large and complex macromolecules, mainly polysaccharides. Extensin, which comprises 5 – 10% of the cell wall, is the major protein component of the cell wall. There is evidence that the level of extensin is developmentally regulated (reviewed in Varner and Cooper, 1983) and wounding leads to a dramatic increase in the rate of synthesis of extensin (Chrispeels and Sadava, 1974).

We have taken a molecular approach towards analyzing the structure and wound-related expression of extensin. We have isolated and characterized a genomic clone pDC5A1 from a carrot genomic DNA library, which contains a 4-kb *Eco*RI insert and

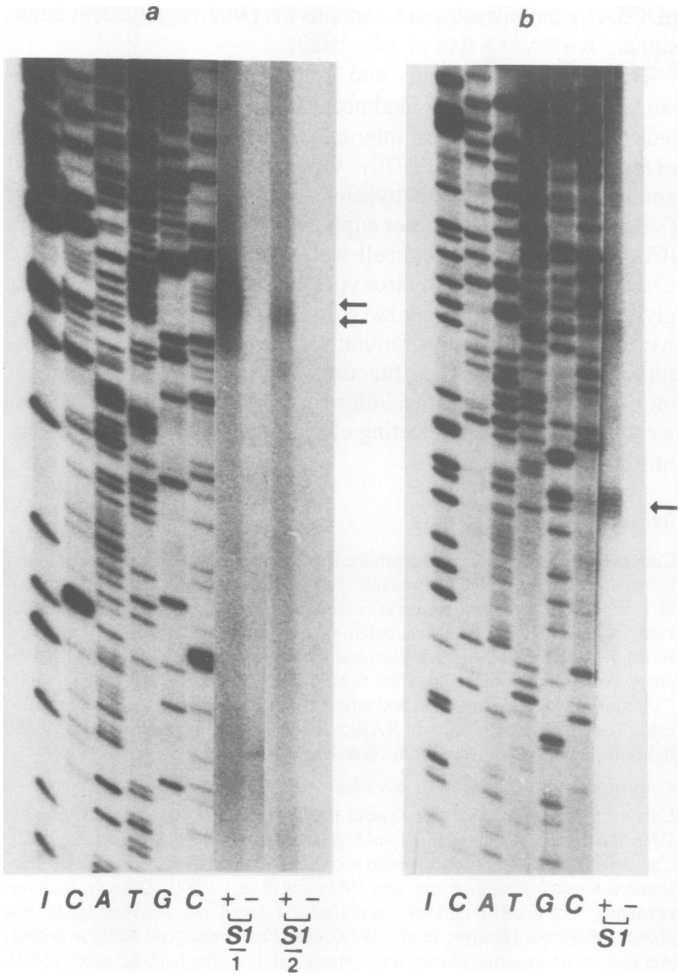

Fig. 7. S1 mapping of extensin RNA transcripts. In **panel a**, DNA fragments synthesized from single-stranded DNA of 485A1XP (sample 1) and 485A1XPD45 (sample 2) with a sequencing primer (GTAAAACG-ACGGCCAGT) were annealed with 10 μg poly(A)$^+$ RNA isolated from wounded carrot and then treated with S1 nuclease. The S1 nuclease-resistant fragments (indicated by arrows) were separated by a sequencing gel along with the sequence ladder of 485A1XP as a size marker. The sizes of resistant fragments are 191 and 184 nucleotides, respectively. The label of + or − represents with or without S1 nuclease treatment. In **panel b**, DNA fragments synthesized from single-stranded DNA of 8+5A1PRL with a sequencing primer (GTAAAACGACGGCCAGT) were used for the S1 mapping. The sequence ladder of 8+5A1PRL was used as a size marker. The S1 nuclease-resistant fragment is 123 nucleotides in size.

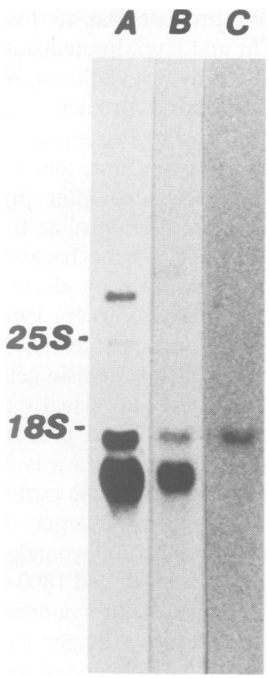

Fig. 8. To demonstrate that there are two transcripts from pDC5A1, 5 μg poly(A)$^+$ RNA isolated from wounded carrot was hybridized to the following probes: pDC5A1 insert **(lane A)**, 9+5A1RPH **(lane B)** and 8+5A1PRL **(lane C)**. Probes of 9+5A1RPH and 8+5A1PRL are synthesized from single-stranded DNA with a sequencing primer (GTAAAACGACGGCCAGT) and pDC5A1 insert DNA is labeled by nick-translation. The two transcripts are 1500 and 1800 nucleotides in size, respectively. Other larger transcripts hybridized to the pDC5A1 insert probe may represent RNA transcribed from other genes which have sequences homologous to extensin gene.

hybridized to a previously described cloned extensin cDNA (Chen and Varner, 1985). Part of the DNA insert from genomic clone pDC5A1 was sequenced. The clone pDC5A1 is thought to represent an actively transcribed gene in carrot root tissue based on the following evidence: (i) the sequence of the cDNA clones pDC5 and pDC11 are contained within the sequence of the genomic clone pDC5A1; (ii) the S1 mapping experiments showed that poly(A)$^+$ transcripts corresponding to this gene were protected by the pDC5A1 genomic DNA. The DNA sequence of pDC5A1 contains an open reading frame encoding a proline-rich polypeptide. We found a signal peptide-like sequence in this predicted amino acid sequence. Because the amino acid sequence of carrot extensin is unknown, the cleavage site can only be predicted based on the rules of Von Heijne (1983). We expected to find such a signal sequence in extensin, because newly formed extensins must be transported through the cell membrane in order to reach the cell wall.

Ser-Pro-Pro-Pro-Pro, Thr-Pro-Val and Tyr-Lys-Tyr-Lys are the major repetitive units in the amino acid sequence derived from the sequence of pDC5A1. Lamport reported that Ser-Hyp-Hyp-Hyp-Hyp, Thr-Hyp-Val and 1/2 IDT-Lys-1/2 IDT-Lys exist in the tomato cell wall protein as repetitive units (Lamport, 1977; Epstein and Lamport, 1984). Both hydroxyproline and isodityrosine (IDT) are formed by a post-translational modification of peptide residues (Chrispeels, 1970; Cooper and Varner, 1983). Therefore, Ser-Pro-Pro-Pro-Pro, Thr-Pro-Val and Tyr-Lys-Tyr-Lys would be the repetitive sequences found in the extensin precursor. The similarity between the extensin peptide sequences of tomato and the predicted amino acid sequences from carrot genomic clone pDC5A1, indicate that pDC5A1 represents one of the genomic sequences encoding extensin in carrot. The similarities also indicate that the peptide sequences of extensins may be conserved in other plants. There are differences in the amino acid composition between purified carrot extensin and the composition of extensin predicted from the DNA sequence (Table I). The differences may be due to deviations in the amino acid composition analysis, minor impurities or a population of extensins in the preparation of extensin from carrot. Two different extensin proteins have been isolated from tomato cell walls (Smith *et al.*, 1984). Peptide mapping with cyanogen bromide or proteases and sequence of peptides of the carrot extensin would distinguish among these possibilities.

Most of the proline residues in carrot extensin are hydroxylated. The distance from the hydroxyl group of serine, of threonine, or of hydroxyproline to the #4 carbon of the adjacent proline in the peptide chain is similar as estimated from molecular models

of these peptides. We propose that the hydroxyl group of the amino acids, Ser, Thr and Hyp, immediately in front of proline residues is the signal for hydroxylation by a specific hydroxylation enzyme. This proposal predicts that the first proline residue in the Ser-Pro-Pro-Pro-Pro repeat will be hydroxylated, and the second only after the first, and so on. The Ser-Hyp-Hyp-Hyp-Hyp repeats produced after proline hydroxylation would be expected to have a polyproline II helical structure according to c.d. spectra and be the backbone for the extensin molecule which when visualized by electron microscopy is an extended rod (Van Holst and Varner, 1984).

In many genes there are codon usage biases (Lycett *et al.*, 1983) and this is the case in the carrot extensin gene. The codons CCA (73 times) or CCT (22 times) are usually used for proline and TCT (25 times out of 30) is used predominantly for serine. Because of these codon usage biases it is possible that there is a translational control of extensin gene expression which depends upon the availability of specific charged tRNAs.

RNA from either unwounded or wounded carrots hybridized to the pDC5A1 probe. Both 1500- and 1800-nucleotide transcripts were found in RNA isolated from wounded and actively growing unwounded carrot. With a longer exposure we detected several large mol. wt. transcripts homologous to pDC5A1. We do not know whether these transcripts are transcribed from the gene corresponding to pDC5A1 or from the related genes. However, only a low level of 1500-nucleotide transcript and an even lower level of the 1800-nucleotide transcript were present in RNA isolated from unwounded carrot which had been stored for prolonged periods of time at 4°C. The mRNA levels of the extensin increased remarkably with wounding (100-fold difference) in carrots that had been stored at 4°C (Figure 5). However, the increase in the extensin RNA of wounded carrots which are active in growth is only several fold (data not shown). During wounding extensin transcripts accumulate in the cytoplasm and extensin accumulates in the cell wall (Chrispeels *et al.*, 1974). It also has been observed that extensins accumulate in cell walls following ethylene treatment (Toppan *et al.*, 1982) and fungal infection (Esquerre-Tugaye *et al.*, 1979). However it is not understood what primary signal activates the expression of extensin genes.

In the 5′-flanking sequence of pDC5A1 we found sequences similar to the TATA boxes reported in other eukaryotic genes (Efstratiadis *et al.*, 1980) at positions ~35 nucleotides upstream of either long (1800 nucleotides) or short (1500 nucleotides) transcripts. Sequences similar to the CAAAT were also observed upstream of the TATA box, with ~40–50 nucleotides separating the two sequences (Figure 3). We do not know whether there are two promoters or if there is a processing of transcripts, which is responsible for these two transcripts. An 18-bp direct repeat is present in the 5′-non-coding region of the longer transcript but not in the short transcript. The significance of this direct repeat is unclear, but the repeat may be relevant to the existence of two transcripts from this extensin gene. Further experiments (e.g., *in vitro* transcription, mutation analysis) are required to establish the sequences required for the transcriptional control. A single intron in the extensin gene exists in the 3′-non-coding region. Sequences conserved at the splicing junctions of other systems, (exon)/GT.....AG/(exon) (Lerner *et al.*, 1980), also occur in the junctions of this gene. In mouse major urinary protein genes, one of the introns is in the 3′-non-coding-region (Clark *et al.*, 1984). The rarity of introns in the 3′-non-coding-region has been suggested to play some, as yet unknown, role in gene expression (Clark *et al.*, 1984). The 3′-flanking sequence of the pDC5A1 genomic sequence contains the putative polyadenylation signal, AATAAG (Hu *et al.*, 1982).

The fact that extensins and collagens are hydroxyproline-containing glycoproteins, and occur in extracellular matrices, has led to the suggestion that they may have common evolutionary precursors (Aaronson, 1970). Comparisons of the DNA and amino acid sequences between carrot extensin and collagens (Vogeli *et al.*, 1981) do not support this hypothesis. Agglutinin (Cooper *et al.*, 1983) and cell wall proteins (Roberts, 1979) of *Chlamydomonas* are hydroxyproline-containing extracellular glycoproteins, which may have domains rich in hydroxyproline with extended structures similar to extensin and other domains for specialized biological function. As more data are generated on the structure of extracellular matrix proteins in eukaryotes, we anticipate more interesting clues to their biological and evolutionary relationships.

Materials and methods

Carrots were either purchased from local grocery stores and stored at 4°C until use or were grown in the greenhouse. For wounded tissue, carrots were sliced to 1 mm in thickness and incubated in a moist chamber at 25°C for 48 h. Enzymes used in nucleic acid manipulations were purchased from New England Biolabs and Bethesda Research Labs and were used as recommended by the suppliers. Nitrocellulose paper was from Schleicher and Schuell. [α-^{32}P]dATP (>400 Ci/mmol) was from Amersham and other nucleotides were from P.L. Biochemicals. The X-ray film used was XAR-5 from Kodak. Sequencing primers and hybridization primers were from New England Biolabs.

Construction of genomic DNA libraries

Carrot DNA was isolated from carrot root by lysing nuclei and purifying the DNA through a CsCl-ethidium bromide gradient as described (Rivin *et al.*, 1982). Carrot DNA was completely digested with the restriction enzyme *Eco*RI and fractionated through a sucrose gradient (Maniatis *et al.*, 1982). DNA in fractions containing 4-kb *Eco*RI fragments was pooled and cloned into bacteriophage lambda charon 4A vector (Blattner *et al.*, 1977) with *Escherichia coli* K802 as a host. Another carrot genomic library was constructed using the lambda vector EMBL 3 (Frischauf *et al.*, 1983). Genomic DNA partially digested with *Sau*3A1 was size fractionated. DNA fragments of ~17 kb were cloned into lambda EMBL 3 with *E. coli* Q364 as a host. Plaque hybridization was according to the methods of Benton and Davis as described by Maniatis *et al.* (1982).

DNA sequencing and Southern hybridization

The insert DNA of recombinant lambda phages was subcloned into plasmid vectors pUC8 (Vieira and Messing, 1982), pEMBL8+, or pEMBL9+ (Dente *et al.*, 1983) or M13 phage vectors mp19 (Norrande *et al.*, 1983) or mWB2348 (Barnes *et al.*, 1983) with *E. coli* strains JM 83 and JM 103 as hosts (Vieira and Messing, 1982). For sequencing, serial deletions were made in the subcloned inserts and sequenced by the Sanger method as described by Barnes *et al.* (1983). DNA sequences were analyzed with computer programs written by W.M.Barnes. Restriction enzyme-digested DNA was separated in agarose gels containing 135 mM Tris-HCl, 45 mM boric acid, and 1 mM EDTA. DNA transfers to nitrocellulose filters were performed as described by Southern (1975).

RNA isolation and Northern blotting

RNA was isolated from unwounded and wounded carrot tissue as described (Haffner *et al.*, 1978). An oligo-(dT)-cellulose affinity column was used to isolate poly(A)$^+$ RNA (Aviv and Leder, 1972). RNA was separated in 1% agarose gels containing formaldehyde and transferred to nitrocellulose paper (Lehrach *et al.*, 1977; Thomas, 1980).

DNA probes and hybridization

Insert DNA was separated by agarose gel electrophoresis, electro-eluted, and nick-translated with ^{32}P-labeled nucleotides (Maniatis *et al.*, 1982). For strand-specific probes, single-stranded DNA was isolated from M13 phage or from defective M13 phages (in the case of subclones in pEMBL8+ or pEMBL9+). Single-stranded DNA was annealed to a sequencing primer (GTAAAACGACGGCCAGT) or a hybridization primer (CACAATTCCACACAAC) and extended from the primer by the large fragment of *E. coli* DNA polymerase I with ^{32}P-labeled nucleotides as described by Hu and Messing (1982). Hybridizations were performed under the following conditions: nitrocellulose blots were pre-hybridized for 2 h at 42°C in a hybridization solution containing 40% formamide, 3.6 × SSC (1 × SSC is 0.15 M NaCl/0.015 M Na citrate, pH 7.0), 3 × Denhardt's mixture (1 × is 0.02% each of bovine serum albumin, Ficoll 400 and polyvinylpyrrolidone), 20 mM Tris-HCl pH 7.9, 5 mM EDTA, 0.2% SDS and 50 μg/ml

sheared and denatured calf thymus DNA. Then the blots were incubated with labeled DNA in a fresh hybridization buffer at 42°C or 37°C for 24 h. Blots were rinsed once with hybridization buffer, then washed at 65°C with 2 × SSC containing 0.2% SDS three times for 30 min each and finally washed with 2 × SSC at 65°C.

S1 nuclease mapping

S1 nuclease mapping of RNA transcripts was done as described (Maniatis *et al.*, 1982) with 10 μg of poly(A)$^+$ RNA, single-stranded DNA probe and 1000 units of S1 nuclease at 37°C for 30 min. DNA sequencing gels were used to size the S1 nuclease-resistant fragments.

Acknowledgements

We thank Dr Allan Showalter for performing the genomic Southern experiment and Steve Janovec for his excellent editorial assistance in preparing this manuscript. The work was supported by NSF PCM8104516, DOE F602-84ER13255, and an unrestricted grant from the Monsanto Company to J.E.V.

References

Aaronson,S. (1970) *Ann. N.Y. Acad. Sci.*, **175**, 531-540.
Avid,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408-1412.
Barnes,W.M., Bevan,M. and Son,P.H. (1983) *Methods Enzymol.*, **101**, 98-122.
Bennett,M.D. and Smith,J.B. (1976) *Phil. Tran. R. Soc. Lond.*, **274**, 227-274.
Blattner,F.R., Williams,B.G., Blechl,A.E., Denniston-Thompson,K., Faber,H.E., Furlong,L.A., Grunwald,D.J., Kiefer,D.O., Moore,D.D., Sheldon,E.L. and Smithies,O. (1977) *Science (Wash.)*, **196**, 161-169.
Chen,J. and Varner,J.E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4399-4403.
Chrispeels,M.J. (1970) *Plant Physiol.*, **45**, 223-227.
Chrispeels,M.J. and Sadava,D. (1974) in Hay,E.D., King,T. and Papaconstantinou,J. (eds.), *Macromolecules Regulating Growth and Development*, Academic Press, NY, pp. 131-152.
Chrispeels,M.J., Sadava,D. and Cho,Y.P. (1974) *J. Exp. Bot.*, **25**, 1157-1166.
Clark,A.J., Clissold,P.M., Shawi,R.A., Beattie,P. and Bishop,P. (1984) *EMBO J.*, **3**, 1045-1052.
Cooper,J.B. and Varner,J.E. (1983) *Biochem. Biophys. Res. Commun.*, **112**, 161-167.
Cooper,J.B., Adair,W.S., Mecham,R.P., Heuser,J.E. and Goodenough,U.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5898-5901.
Dente,L., Ceasreni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645-1655.
Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz, R.A., DeRiel,J.K., Forget,B.G., Weissman,S.M., Slightom,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.C. and Proudfoot,N.J. (1980) *Cell*, **21**, 653-668.
Epstein,L. and Lamport,D.T.A. (1984) *Phytochemistry*, **23**, 1241-1246.
Esquerre-Tugaye,M.T., Lafitte,C., Mazau,D., Toppan,A. and Touze,A. (1979) *Plant Physiol.*, **64**, 320-326.
Esquerre-Tugaye,M.T. and Lamport,D.T.A. (1979) *Plant Physiol.*, **64**, 314-319.
Frischauf,A.M., Lehrach,H., Poustka,A.M. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827-842.
Fry,S.C. (1982) *Biochem. J.*, **204**, 449-455.
Haffner,M.H., Chin,M.B. and Lane,B.G. (1978) *Can. J. Biochem.*, **56**, 729-733.
Hu,N.T. and Messing,J. (1982) *Gene*, **17**, 271-278.
Hu,N.T., Peifer,M.A., Heidecker,G., Messing,J. and Rubenstein,I. (1982) *EMBO J.*, **1**, 1337-1342.
Kyte,J..and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105-132.
Lamport,D.T.A. (1965) *Adv. Bot. Res.*, **2**, 151-218.
Lamport,D.T.A. (1977) in Loewus,F.A. and Runeckles,B.C. (eds.), *Recent Advances in Phytochemistry*, Vol. II, Plenum Press, NY, pp. 79-115.
Lamport,D.T.A., Katona,L. and Roerig,S. (1972) *Biochem. J.*, **133**, 125-132.
Lehrach,H., Diamond,D., Wozney,J.M. and Boektker,H. (1977) *Biochemistry (Wash.)*, **16**, 4743-4751.
Lerner,M.R., Boyle,J.A., Mount,S.M., Wolin,S.L. and Steitz,J.A. (1980) *Nature*, **183**, 220-224.
Lycett,G.W., Delauney,A.J. and Croy,R.R.D. (1983) *FEBS Lett.*, **153**, 43-46.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
McNeil,M., Darvill,A.G., Fry,S.C. and Albersheim,P. (1984) *Annu. Rev. Biochem.*, **53**, 635-663.
Norrande,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101-106.
Rivin,C.J., Zimmer,E.A. and Walbot,V. (1982) in Sheridan,W.F. (ed.), *Maize for Biological Research*, Plant Molecular Biology Association, Virginia, pp. 161-164.
Roberts,K. (1979) *Planta*, **146**, 275-279.
Smith,J.J., Muldoon,E.P. and Lamport,D.T.A. (1984) *Phytochemistry*, **23**, 1233-1239.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Stuart,D.A. and Varner,J.E. (1980) *Plant Physiol.*, **66**, 787-792.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.
Toppan,A., Roby,D. and Esquerre-Tugaye,M.T. (1982) *Plant Physiol.*, **70**, 82-86.
Van Holst,G.J. and Varner,J.E. (1984) *Plant Physiol.*, **74**, 247-251.
Varner,J.E. and Cooper,J.B. (1983) in Ciferri,O. and Dure,L. (eds.), *Structure and Function of Plant Genomes*, Plenum Press, NY, pp. 463-480.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.
Vogeli,G., Ohkubo,H., Avvedimento,V.E., Sullivan,M., Yamada,Y., Mudryj, M., Pastan,I. and deCrombrugghe,B. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 777-783.
Von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17-21.

Received on 4 June 1985