

Molecular characterization of tomato fruit polygalacturonase

Raymond E. Sheehy¹, Judith Pearson², Colin J. Brady², and William R. Hiatt¹

¹ Calgene Inc., 1920 Fifth St, Davis, CA 95616, USA

² Plant Physiology Unit, CSIRO Division of Food Research and School of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia

Summary. Using the expression vector λ gt11 and immunological detection, cDNA clones of an endopolygalacturonase gene of tomato (*Lycopersicon esculentum* Mill.) were isolated and sequenced. The 1.6 kb cDNA sequence predicts a single open reading frame encoding a polypeptide of 457 amino acids. The PG2A isoform of tomato fruit endopolygalacturonase was purified and 80% of the amino acid sequence determined. The amino acid sequence predicted by the cDNA sequence was identical to the amino acid sequence of the PG2A isoform. The position of the codon for the N-terminal amino acid of mature PG2A in the open reading frame indicates the presence of a 71 amino acid N-terminal signal peptide which is post-translationally processed. The C-terminus of purified PG2A occurred 13 amino acids before the stop codon in the cDNA suggesting that C-terminal processing of PG2A may also occur. The nucleotide and amino acid sequence data predict a mature protein of 373 amino acids and a polypeptide molecular weight of 40279. The sequence contains four potential glycosylation sites. Northern analysis detected endopolygalacturonase mRNA in stage 3 (turning) and stage 6 (red) ripening fruit, but not in leaves, roots, or green fruit of normal cultivars or in mature fruit of the *rin* mutant.

Key words: Polygalacturonase – cDNA sequence – Protein sequence – Protein processing – Tomato fruit

Introduction

The ripening of fleshy fruits involves changes in the steady state levels of specific messenger RNAs (Christoffersen et al. 1982; Speirs et al. 1984; Mansson et al. 1985; Slater et al. 1985). This suggests that ripening is regulated by concurrent gene activation or suppression. Direct genetic control of ripening is also suggested by the recognition of mutations which eliminate, or greatly suppress, climacteric ripening (Tigchelaar et al. 1978).

Of the mRNAs that increase with the onset of ripening, two have been identified. In avocados, the mRNA coding for an endo β -1,4-glucan hydrolase (cellulase, EC 3.2.1.4) is barely detectable in preclimacteric fruit but is abundant in climacteric fruit (Christoffersen et al. 1984). Two-dimensional gel analysis suggests that multiple forms of the enzyme exist in fruit (Tucker et al. 1985). Additional forms

of cellulase which differ in M_r appear to represent processing intermediates (Bennett and Christoffersen 1986). Although cellulase cDNAs have been isolated (Christoffersen et al. 1984; Tucker et al. 1985), the structural relationships between the various enzyme forms and the corresponding gene(s) has not been completely established.

During tomato fruit ripening, the mRNA for poly (1,4- α -D-galacturonide) glycan hydrolase (EC 3.2.1.15, polygalacturonase or PG) becomes abundant (DellaPenna et al. 1986). This enzyme also exists in several molecular forms (Ali and Brady 1982). In vitro translation of PG mRNA and immunoprecipitation of the product has demonstrated a polypeptide with a M_r of approximately 54000 (DellaPenna et al. 1986; Sato et al. 1984; Biggs et al. 1986). The smallest form of the enzyme recovered from ripe tomato fruit is PG2A with an apparent M_r of approximately 43000 (Ali and Brady 1982) or 46000 (Tucker and Grierson 1982) including a carbohydrate moiety presumably added post-translationally (Ali and Brady 1982; Moshrefi and Luh 1983). These results suggest that the precursor polypeptide undergoes significant post-translational processing. There has been some difficulty in recognizing PG mRNA by in vitro translation and immunoprecipitation (Grierson et al. 1985; Slater et al. 1985) indicating the need for a more complete characterization of PG cDNAs.

In this paper, we describe the isolation of PG cDNAs, and compare the cDNA sequence with the amino acid sequence of the PG2A isoform. The comparison indicates that an N-terminal signal peptide of 71 amino acids is removed post-translationally, and also suggests that C-terminal processing of PG2A occurs.

Materials and methods

Plant material. Tomato plants (*Lycopersicon esculentum* cv Caligrande and cv Campbell 33) were field grown. The varieties UC82B and cv Ailsa Craig were grown under greenhouse conditions. The *rin* mutation was incorporated into cv Ailsa Craig.

Isolation of poly (A)⁺ RNA. Frozen tissues were ground to a powder under liquid nitrogen and homogenized with a Brinkmann polytron. Total RNA was isolated as described by Colbert et al. (1983) using the extraction buffer described by Facciotti et al. (1985). Polysaccharides were precipitated from total RNA preparations (Mansson et al.

1985) prior to the preparation of poly (A)⁺ RNA as described by Maniatis et al. (1982).

Construction of cDNA libraries. Synthesis of cDNA from poly (A)⁺ RNA prepared from ripe fruit of cv Caligrande was performed as described by Gubler and Hoffman (1983) with the following modifications: the reaction mixture for synthesis of the first strand included 1 mM dGTP, 1 mM dATP, 1 mM dTTP, 0.5 mM dCTP, 20 μ Ci 5'-[α -³²P] dCTP, 0.5 units/ μ l RNasin (Promega), 4 μ g tomato poly (A)⁺ RNA, and 80–100 units of reverse transcriptase (Life Sciences). The reaction was stopped with 2 μ l 500 mM EDTA, then precipitated with 10 μ g tRNA, 1 vol 4 M ammonium acetate, and 2.5 vol of ethanol overnight on dry ice. Second strand synthesis was performed from approximately 500 ng of the first strand reaction product.

Cloning of double-stranded cDNA in λ gt11 and λ gt10. The double stranded cDNA was *Eco*RI methylated as described by the manufacturer (New England Biolabs). The cDNA ends were blunted using the Klenow fragment of polymerase I in 66 mM Tris HCl, 10 mM MgCl₂, 100 μ M dithiothreitol (DTT), 100 μ M dGTP, dATP, dTTP, dCTP, pH 7.5 at room temperature for 1 h. Phosphorylated *Eco* RI linkers (2 μ g) were added to the cDNA in 10 μ l ligase buffer (50 mM Tris, 10 mM MgCl₂, 20 mM DTT, 1 mM ATP, 5 mg/ml bovine serum albumin, pH 7.5). T₄ DNA ligase (1 Weiss unit, Promega) was added and the reaction incubated for 6 h at 15°C. An additional Weiss unit of T₄ DNA ligase in 10 μ l ligase buffer was then added and the reaction incubated for 24 h at 15–19°C. Following digestion with 100 units of *Eco*RI (New England Biolab) for 6–8 h, excess linkers and cDNAs smaller than 500 bp were removed by chromatography on Bio-Gel A-50 m (100–200 mesh) and the remaining cDNA ligated to *Eco*RI-cleaved and dephosphorylated λ gt10 and λ gt11 vector DNA (Stratagene) as described by Huynh et al. (1985). In vitro packaging reactions were performed with Gigapack extracts (Stratagene) as described by the vendor. Greater than 5×10^6 recombinants were obtained in λ gt11 and λ gt10 at approximately a 90% insertion rate.

Library screening. The unamplified λ gt11 library was screened immunologically as described (Young and Davis 1983; Huynh et al. 1985). Nitrocellulose overlays were saturated with 0.05 M Tris, 0.15 M NaCl, 0.02% NaN₃, 0.1% Triton X-100, 5% non-fat dry milk, pH 8.0 and incubated 30 min at room temperature with the same buffer containing an anti-PG2 antibody preparation diluted 1:1000. Bound antibody was detected with an alkaline phosphatase conjugated second antibody (Promega) as described by the vendor. The λ gt10 library was screened with a ³²P-labeled PG-specific probe as described by Huynh et al. (1985). Positive plaques were purified by successive plating and phage DNA was prepared as described by Maniatis et al. (1982).

RNA and DNA blot hybridizations. RNA and DNA blots were performed essentially as described by Mansson et al. (1985) using Gene Screen Plus (New England Nuclear). ³²P-labeled probes for Northern analysis were prepared from Bluescribe vectors (Stratagene) as described by the vendor. DNA fragments used as ³²P-labeled probes for Southern analysis were recovered from low melting temperature agar-

ose gels and purified by phenol extraction and chromatography on Elutip-d Columns (Schleicher & Schuell). DNA fragments were ³²P-labeled by nick translation (Bethesda Research Laboratories) as described by the vendor.

DNA sequencing. Initial sequencing of cDNA inserts was performed using single stranded template from M13 Bluescribe Minus vector (Stratagene) constructs prepared as described by the vendor. DNA sequencing was performed by the methods of Sanger et al. (1977) or Maxam and Gilbert (1980). Overlapping sequences were obtained by subcloning restriction fragments as described by Maniatis et al. (1982) from the Bluescribe constructs into M13mp18 and 19 vectors (Yanisch-Perron et al. 1985). Additional sequencing was performed on sequential deletions produced from inserts cloned into Bluescribe Minus using exonuclease III and mung bean nuclease (Stratagene) as described by the vendor. Resulting fragments were subcloned into M13 vectors for sequencing.

PG purification for antibody production. Cell wall bound proteins were prepared from ripe tomato fruit of cv UC82B as described by Tucker and Grierson (1982). The ammonium sulphate precipitate was dissolved in 150 mM NaCl and then dialyzed against the same buffer. The protein solution was fractionated on a TSK 3000/2000 HPLC column in 10 mM Tris, 10 mM NaCl, pH 7.2 at a flow rate of 0.5 ml/min. Fractions containing PG activity (Reisfeld et al. 1962) were pooled and fractionated over a hydroxylapatite HPLC column using a linear gradient of 10 mM–350 mM PO₄, pH 6.8 and a flow rate of 1 ml/min. The peak containing PG activity contained two polypeptides (PG2A and PG2B), and this preparation was used to inject rabbits. PG for booster injections was prepared by resolving the cell wall bound protein preparation on SDS polyacrylamide gels (Laemmli 1970). The region containing PG protein (PG2A and PG2B) was excised, frozen and ground in liquid N₂.

Antibody preparation. One rabbit was given four injections (Antibodies, Inc.) of PG (125 μ g/injection) over a one month period, then two booster injections of PG (approximately 150 μ g each) at two week intervals. Serum was obtained two weeks after the last injection. Immunoglobulins in the serum were adsorbed to Protein A-Sepharose (Sigma), washed with 0.1 M phosphate buffer, pH 7.0 and eluted with 0.1 M glycine, pH 3. The eluted immunoglobulins were then incubated overnight at 4°C with covalently-linked PG2A-Sepharose (Pharmacia) in 0.01 M Tris, 150 mM NaCl, 0.5% Tween-20, pH 7.5 (TBST). The gel was washed with TBST and eluted with 0.2 M glycine, pH 2.75. The eluate was dialyzed against 20 mM Tris, 150 mM NaCl, pH 8.0.

Purification of PG2A for protein sequencing. PG2A (Ali and Brady 1982) was isolated from ripe fruit of cv Campbell 33. Fruit pericarp tissue was homogenized in 3 vol of 1.3 M NaCl, 0.05 M sodium phosphate, 40 mM 2-mercaptoethanol, pH 4.5 and incubated at 4°C for 2 h. The supernatant was recovered by centrifugation (11000 $\times$ g, 30 min) and made 40% saturated with ammonium sulphate. After 1 h at 4°C, the extract was centrifuged (13000 $\times$ g, 30 min) and the supernatant made 80% saturated with ammonium sulphate. The precipitate was recovered (13000 $\times$ g, 30 min), dissolved in 0.05 M sodium phosphate, 1 mM DTT, pH 6.0

and dialyzed for 24 h against the same buffer. The dialyzed extract was centrifuged ($40000 \times g$, 30 min) and the supernatant loaded on a 25×300 mm column of CM-Sepharose CL 6B (Pharmacia) equilibrated at 2°C with 0.05 M sodium phosphate, pH 6.0. The column was washed with equilibration buffer and bound proteins were eluted with a linear gradient of NaCl to 0.7 M . Two zones of PG activity emerged; the earlier, major peak contained PG2A and PG2B and this zone was concentrated by ultrafiltration (Amicon YM10 membrane) and applied to a 25×900 mm column of Sephacryl S-200 (Pharmacia) equilibrated at 2°C with 0.05 M sodium phosphate, 0.5 M NaCl, 1 mM DTT, pH 6.0. The column was eluted at 30 ml/h and the single active zone that emerged was concentrated by ultrafiltration and loaded onto a 9×100 mm column of Concanavalin A-Sepharose 4B (Pharmacia) which was eluted with 0.05 M sodium phosphate, 0.5 M NaCl, 1 mM MgCl_2 , 1 mM CaCl_2 , pH 6.0, and then with a gradient to 25 mM α -methylmannoside in the same buffer. PG2A emerged ahead of PG2B but the separation was incomplete. The PG2A-enriched zone was concentrated and loaded onto a 15×900 mm column of Sephadex G-75, superfine (Pharmacia), equilibrated at 2°C with 0.5 M sodium phosphate, 0.5 M NaCl, 1 mM DTT, pH 6.0. The column was eluted at 6 ml/h with PG2B emerging ahead of PG2A.

S-carboxymethylation. The zone containing only PG2A was dialyzed against 40 mM NH_4HCO_3 and lyophilized. The lyophilized protein (1.5 mg) was dissolved in 2 ml 6 M guanidine, 0.1 M Tris, 5.4 mM EDTA, pH 8.6 and 2-mercaptoethanol ($50\text{ }\mu\text{l}$) was added. The sample was stirred in an N_2 atmosphere at 40°C for 4 h and 134 mg of iodoacetic acid, neutralized with KOH, was added and the sample was left to react at 20°C for 20 min. The sample was dialyzed against several changes of water and lyophilized.

Cyanogen bromide cleavage. S-carboxymethylated protein ($500\text{ }\mu\text{g}$) was dissolved in 1 ml 70% formic acid and 3 mg of cyanogen bromide was added. The sample was stirred at 20°C for 24 h, diluted with 9 ml water and lyophilized. The residue was dissolved in $200\text{ }\mu\text{l}$ 0.05% trifluoroacetic acid (TFA), and the cleavage products separated on a Shodex WS 802.5 gel permeation column with 0.05% TFA in 50% acetonitrile.

Lysine peptides. Endoproteinase Lys-C (Boehringer Mannheim GmbH, $8\text{ }\mu\text{g}$) was added to 1 mg S-carboxymethylated protein dissolved in $100\text{ }\mu\text{l}$ 0.1 M Tris, pH 8.35 and the solution incubated at 40°C for 2 h. The sample was centrifuged, and peptides in the supernatant were separated using a Brownlee C18 RP, $5\text{ }\mu\text{m}$, $4.6 \times 250\text{ mm}$ column and a gradient of $0\text{--}70\%$ acetonitrile in 0.1% TFA.

Amino acid sequencing. Protein and peptide sequencing was performed on a gas phase protein sequencer (Applied Biosystems Model 470A). The released phenylthiohydantoin-amino acids were analyzed by separation on a Zorbax 'Gold Series', C-18, $3\text{ }\mu\text{m}$, $6.2 \times 80\text{ mm}$ HPLC column using a gradient of $22\text{--}58\%$ acetonitrile in 12.5 mM ammonium acetate. C-terminal analysis with carboxypeptidase was as described by Takashima et al. (1986) and the phenylisothiocyanate derivatives of released amino acids were analyzed by the method of Henrikson and Meredith (1984).

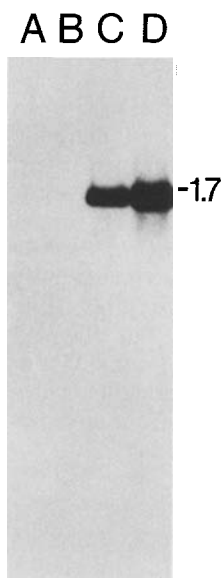


Fig. 1. Northern blot analysis of tomato RNA with PG-specific probe C3. Poly(A)⁺ RNA was prepared from the pericarp of the following fruit: lane A, stage 1 (green) cv UC82B; lane B, *rin* mutant cv Ailsa Craig 50 days post anthesis; lane C, stage 6 (red) cv UC82B; lane D, stage 3 (turning) cv UC82B

Results

Isolation of PG cDNA clones

Twelve immunopositive clones were isolated from the $\lambda\text{gt}11$ library after screening approximately 200 000 plaque-forming units. The cDNA insert from one clone (C3 = 0.9 kb) was used to probe mRNA prepared from normal fruit at the green, turning, and red stages of ripening, and from mature *rin* fruit (Fig. 1). C3 hybridized to a mRNA species with the size and expression pattern expected for PG. An abundant mRNA species of approximately 1.7 kb was detected in turning and red fruit, but not green fruit of cv UC82B or mature fruit of cv Ailsa Craig containing the *rin* mutation. Leaves and root tissues from cv UC82B also did not contain detectable levels of PG mRNA (data not shown).

Phage DNA was prepared from the remaining 11 clones, digested with *EcoRI* and *HindIII*, and subjected to Southern blot analysis using clone C3 as a probe. Two additional cDNA clones (N1 and P1) cross-hybridized with C3. The $5'$ *BamHI-EcoRI* fragment of the larger clone (P1 = 1.3 kb) was used as a probe to screen the $\lambda\text{gt}10$ library. Positive signals from 31 plaques resulted from a screen of 20 000 clones. Phage DNA prepared from ten of the clones was digested with *BamHI* and *EcoRI*, and subjected to Southern analysis using the $5'$ *BamHI-EcoRI* fragment of P1 as a probe. Eight of the ten clones contained a *BamHI-EcoRI* fragment approximately 300 bp larger than the $5'$ *BamHI-EcoRI* fragment of P1 (data not shown). Clone F1 (1.6 kb) is representative of this group. The alignment and partial restriction maps of C3, P1, and F1 are shown in Fig. 2.

Sequence of PG cDNAs

The cDNA inserts from C3, P1, and F1 were subcloned into M13 vectors and sequenced by the strategy shown in Fig. 2. The sequence of F1 is detailed in Fig. 3. Sequenced regions of C3 and P1 were identical to the sequence of F1 except at position 1438 of P1 where C replaced G. The sequence of C3 in this region was not determined. Identification of translational start and stop codons

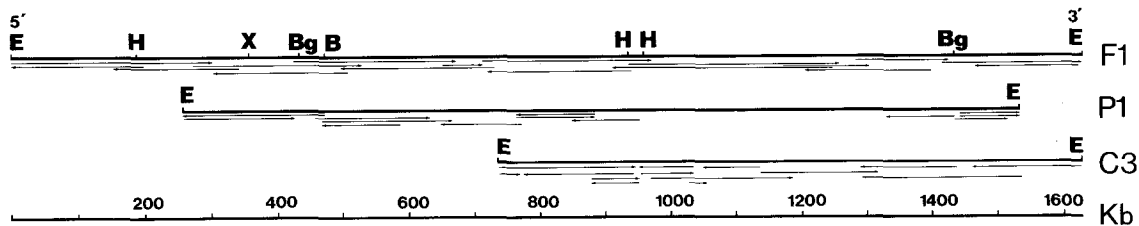


Fig. 2. Partial restriction map and sequencing strategy for PG cDNA clones. Cleavage sites for restriction endonucleases are: B, *Bam*HI; Bg, *Bg*II; E, *Eco*RI; H, *Hind*III; X, *Xba*I. Sites were determined from the DNA sequences. Arrows indicate the regions and direction of DNA sequencing

```

          30          60          90
GAATTC AATGAC AAGTTT AAAA ACCATACCATATAACAATATATCATGGTTATCCAAAGGAATAGTATTCTCCTTCTCATTATTATTTTTCCTTCATCAATTTCAACTTGTAAGCAA
150      180      210
TGTTATTGATGACAATTTATTCAACAGTTTATGATAATATCTTGAACAAGAAATTTGCTCATGATTTTCAAGCTTATCTTTCTATTGAGCAAAAATATTGAAAGCAACAATAATAT
270      300      330
TGACAAGGTTGATAAAAATGGGATTAAAGTGATTAAATGTACTTAGCTTTGGAGCTAAGGGTGATGAAAAACATATGATAATATGCATTGAGCAAGCATGGAATGAAGCATGTTTCATC
390      420      450
TAGAACACCTGTTCAATTTTGGTTCTTAAAAACAAGAAATATCTTCTCAAGCAAAATCACCTTTTCAGGTCCATGCAGATCTTCTATTTCAGTAAAGATTTTGGATCCTTAGAAGCATC
510      540      570
TAGTAAATTTCAGACTACAAAGATAGAAGGCTTTGGATTGCTTTTGATAGTGTTCAAAATTTAGTTGTGGAGGAGGAGAACTATCAATGGCAATGGCAAGTATGGTGGCCAAAGTTC
630      660      690
TTGCAAAATAAATAAATCACTGCCATGCAGGGATGCACCAACGGCCTTAACCTTCTGGAATTGCAAAAAATTTGAAAGTGAATAATCTAAAGAGTAAAAATGCACAACAAATTCATATCAA
750      780      810
ATTTGAGTCATGCACTAATGTTGTAGCTTCAAATTTGATGATCAATGCTTCAGCAAGAGGCCAAATACCTGATGGAGTCCATGATCAAACTCAATATATTTCAAATATCTGATACATAT
870      900      930
TATTGGAACAGGTGATGATTGTATTTCATTTGTTCTGGATCTCAAAATGTGCAGGCCACAAATATTACTTGTGGTCCAGGTGATGGTATAAGTATTGGAAGCTTAGGATCTGGAATTC
990      1020      1050
AGAAGCTTATGTGTCTAATGTTACTGTAAATGAAGCCAAAATATCGGTCCGAAAATGGAGTTAGGATCAAGACTTGGCAGGGAGGATCTGGACAAGCTAGCAACATCAAATTTCTGAA
1110      1140      1170
TGTGGAATGCAAGACGTTAAGTATCCATAAATTATAGACCAAAATATTGTGATCGAGTTGAACCATGTATACAACAGTTTTCAGCAGTTCAAGTGAATAATGTTGTATGAGAATAT
1230      1260      1290
CAAGGGCACAAAGTGCAACAAAGGTGGCCATAAAATTTGATTGCAGCACAAATTTCCATGTGAAGGAATTATAATGGAGAATATAAAATTTAGTAGGGGAAAGTGGAAAACCATCAGAGGC
1350      1380      1410
TACGTGCAAAATGTCCATTTTAAACATGTGTAACATGTTACACCACTGCACCTTCACATGAATAATTTAGAGGATGAAGCTCTTTGTATAATTATTAATTTATATCTATAGATCTTGAA
1470      1500      1530
TATATAGCAGATATGATATATCACATAAACTATATCTATGTATTGAATAATTATTATTAATATGTACGGATTGAAGTTTAAATAAGACTACTATGATTCTTCTATTCTCTAGTCA
1590
AAAGTTTGACGATTGTACTTTTAAATGTACAAAATAATAAAATGGTTATTTATATG

```

Fig. 3. Nucleotide sequence of PG cDNA clone F1. The underlined codons represent the start (47–49) and stop (1418–1420) of the open reading frame. Potential polyadenylation signal sequences are also underlined

predicts a single open reading frame from base position 47 to position 1417, encoding a polypeptide of 50057 Da. The open reading frame is followed by a 3' untranslated region of 200 bp which contains two consensus polyadenylation signal sequences. The location of the poly(A) tract for F1 was not determined. However, a poly(A) tract for C3 was located starting at base position 1618 and for P1 starting a position 1522. The position of the mRNA cap site has not yet been determined. The 5' terminal sequence of F1 (GAATTC) may represent an *Eco*RI linker rather than PG cDNA sequence.

Amino acid sequence

PG2A was purified free of PG2B and other proteins and partially sequenced as shown in Fig. 4. The sequences of cyanogen bromide and endoproteinase Lys-C generated peptides fitted precisely within the open reading frame of F1 (Fig. 3) and confirmed the identity of F1 as a PG cDNA. The amino acid sequence was determined for 81% of the polypeptide sequences predicted by F1. Regions that were distant from methionine and lysine cleavage sites were unconfirmed (shown in Fig. 4 by dotted lines). The sequence shown in Fig. 4 is based on the cDNA sequence. The amino acid sequences determined for purified PG2A and predicted from the cDNA sequence agree precisely, except for the expectation that residue 233 in the sequenced protein may be lysine since peptide K16 is presumed to have been liberated by the cleavage of a lysine carboxyl bond. The isolated PG2A protein terminated at a cysteine-

threonine-serine sequence 13 amino acids before the predicted stop codon. The PG2A protein C-terminal sequence was established by (1) the failure of CNB2 and K24, and also of a peptide commencing from glutamate 365 (not shown in Fig. 4) to sequence beyond serine 373, (2) the absence of tyrosine in CNB2 and (3) by carboxypeptidase analysis which gave the serine-threonine-cysteine-histidine sequence.

The amino-terminus of PG2A corresponds to the codon GGG at base positions 260–262. The remaining open reading frame encodes a polypeptide of 386 amino acids (M_r 41833) to the stop codon, with 373 amino acids (M_r 40279) before the C-terminal serine of PG2A. PG2A is a glycoprotein, and four potential glycosylation sites are identified in Fig. 4. Three of these were confirmed by peptide sequencing.

The pre-sequence of PG

The open reading frame of F1 contains information for 71 amino acids upstream from the N-terminus of the mature protein. The predicted amino acid sequence is shown in Fig. 5.

Discussion

The results presented in this paper provide a structural characterization of PG cDNAs and the PG2A protein. Comparison of the sequence of PG cDNA F1 (Fig. 3) with the partial amino acid sequence determined for PG2A (Fig. 4)

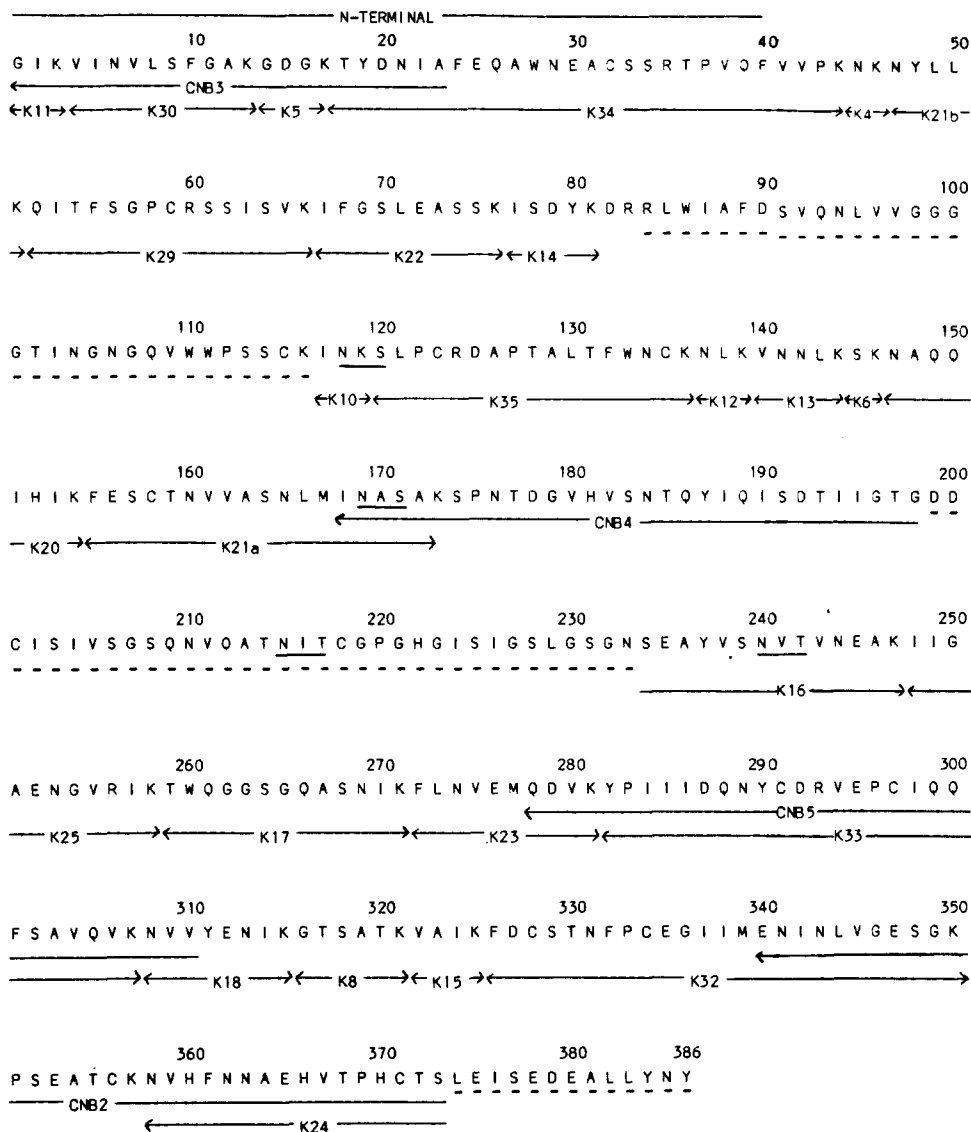


Fig. 4. Overlapping peptide sequences and predicted amino acid sequence of PG. The amino acid sequence predicted by the nucleotide sequence in Fig. 3 is given using the single letter amino acid code. The N-terminal amino acid sequence of PG2A is shown by the *solid line* above residues 1–39. Sequences of peptides derived from PG2A by cyanogen bromide treatment are labeled CNBn. Sequence of peptides released from PG2A by endoproteinase Lys-C are labeled Kn where n represents the order of elution from the peptide fractionating column; a and b after K21 refer to peptides incompletely separated by HPLC. The *dotted line* shows predicted sequence for which there is no confirmatory amino acid sequence. Putative glycosylation sites (N_xT or N_xS) are *underlined*

establishes that F1 spans the full length of the mature PG2A protein and also includes an N-terminal 71 amino acid signal peptide, and a 13 amino acid extension beyond the C-terminus of the isolated PG2A protein. The 71 amino acid signal peptide supports previous reports that PG is synthesized as a precursor polypeptide of about 54 kDa which is extensively processed after translation (Sato et al. 1984; Biggs et al. 1986; DellaPenna et al. 1986). The sequence includes four potential glycosylation sites, at which the glucosamine, mannose and fucose residues present in the mature protein (Moshrefi and Luh 1983) may be added.

PG2A is the most abundant form of PG in ripe fruit (Ali and Brady 1982). PG2B and the active subunit of PG1 (Moshrefi and Luh 1983) may have the same amino acid sequence as PG2A. At least two lines of circumstantial evidence support this idea. The active subunit of PG1 is electrophoretically identical to PG2A (Moshrefi and Luh 1983) and yields trypsin and chymotrypsin peptide patterns identical to those obtained from PG2A (Brady and Speirs, unpublished). PG2B has the same isoelectric point as PG2A (Ali and Brady 1982) but is larger (separable by gel electrophoresis and gel permeation chromatography) and has a higher affinity for the lectin Concanavalin A suggesting that

the difference in size reflects a higher carbohydrate content. One PG gene may, then, code for each of the recognized isoforms of PG.

A comparison of the structures of cDNAs C3, P1, and F1 demonstrated almost identical DNA sequences for the regions analyzed. F1 and P1 differed by a single base in the 3' untranslated region. In addition, P1 and C3 contained poly(A) tracts at different locations, with C3 representing a longer transcript. C3 and F1 contain two hexanucleotide consensus sequences (AATAAA) for polyadenylation, and it is possible that either site can be used. All three cDNA clones contained the sequences encoding the 13 amino acid C-terminal extension. F1 has been established as a PG-specific probe and is being used to determine gene number and structure.

The sequence of F1 indicates that PG is initially synthesized as a polypeptide with a molecular weight of 50057 daltons. Removal of the signal peptide, presumably in the endoplasmic reticulum, will reduce this to a polypeptide of 41833 daltons. The addition and then the trimming of carbohydrate will follow. Recovered PG2A is 13 amino acids shorter at the C-terminal end than predicted by the F1 sequence. The C-terminus may be modified in the tissue,

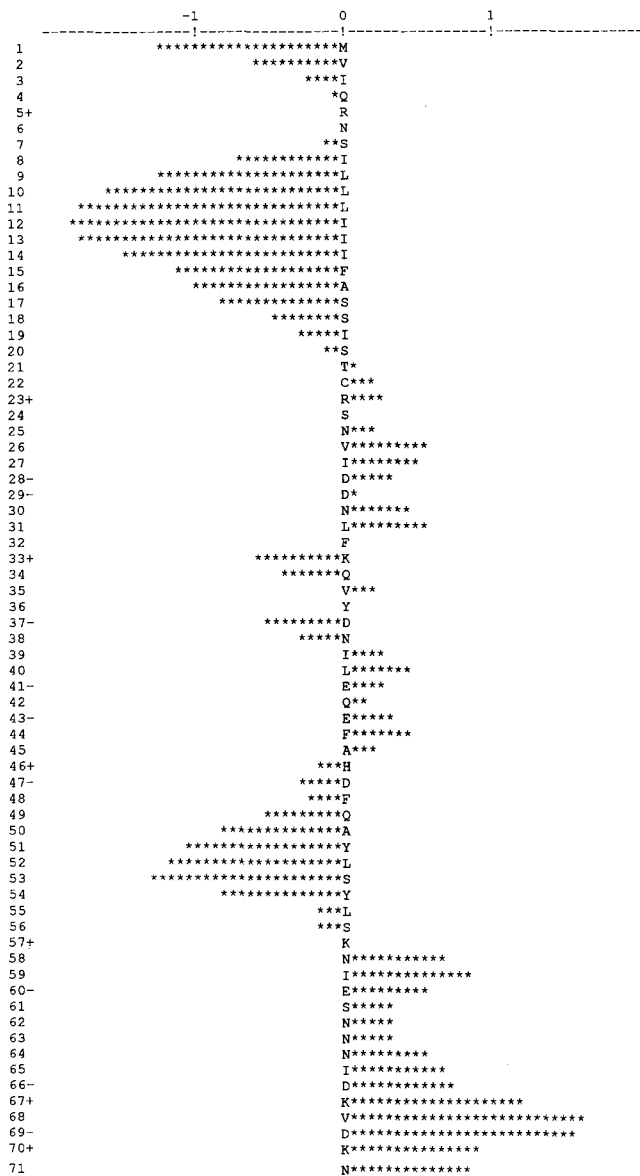


Fig. 5. Hydropath plot of the predicted PG signal polypeptide. The plot was produced from the PEP program of IntelliGenetics, Inc. with a windowing average across 6 residues. Charged groups (+, -) are shown against the residue number. The asterisks indicate the degree of hydrophobicity (<0) or hydrophilicity (>0) for each residue

or during isolation or purification. If the cleavage occurs during isolation, enzyme activity is retained because PG 2A, as sequenced, has the enzymatic specific activity reported by Ali and Brady (1982). It is not likely that PG 2B retains this C-terminal peptide for it would then separate from PG 2A during isoelectric focusing.

Signal peptides typically contain 15–25 amino acids (von Heijne 1984). The first 25 residues of the PG pre-sequence are characteristic of a signal peptide: there is a short positively charged N-terminal region, a strongly hydrophobic core and a region following the core which is more polar. Another 46 residues remain before the cleavage site. Signal peptides and transit peptides for localization to mitochondria and chloroplasts vary in size but are generally longer than secretory signal peptides and more com-

parable to the PG pre-sequence in size. Multiple domains and more than one processing step have been suggested for some organelle proteins (Smeekens et al. 1986; Chia and Arntzen 1986). The PG leader sequence, and, indeed, the C-terminal sequence may contain information relevant to the targeting of the enzyme to a cell-wall location. It is noteworthy, however, that avocado fruit β -1,4-glucanase (cellulase), which also is assumed to be located in the cell wall, appears to have a much shorter leader sequence (Bennett and Christoffersen 1986).

During preparation of this manuscript, the sequence of a tomato PG cDNA clone was reported by Grierson et al. (1986) and shown to be identical to F1 except at position 1438.

Acknowledgements. We would like to thank J. Kenny and L.M. Müller for PG 2 purification for antibody production, J. Caton and B. Rose for DNA sequencing, C. Shewmaker for scientific discussions, and D. Stalker for conception of Calgene's effort.

References

- Ali ZM, Brady CJ (1982) Purification and characterization of polygalacturonases of tomato fruits. *Aust J Plant Physiol* 9: 155–169
- Bennett AB, Christoffersen RE (1986) Synthesis and processing of cellulase from ripening avocado fruit. *Plant Physiol* 81: 830–835
- Biggs MS, Harriman RW, Handa AK (1986) Changes in gene expression during tomato fruit ripening. *Plant Physiol* 81: 395–403
- Chia CP, Arntzen CJ (1986) Evidence for two step processing of nuclear-encoded chloroplast proteins during membrane assembly. *J Cell Biol* 103: 725–732
- Christoffersen RE, Warm E, Laties GG (1982) Gene expression during fruit ripening in avocado. *Planta* 155: 52–57
- Christoffersen RE, Tucker ML, Laites GG (1984) Gene expression during fruit ripening in avocado fruit: the accumulation of cellulase mRNA and protein as demonstrated by cDNA hybridization and immunodetection. *Plant Mol Biol* 3: 385–391
- Colbert JT, Hershey HP, Quail P (1983) Autoregulatory control of translatable phytochrome mRNA levels. *Proc Natl Acad Sci USA* 80: 2248–2252
- DellaPenna D, Alexander DC, Bennett AB (1986) Molecular cloning of tomato fruit polygalacturonase – analysis of polygalacturonase mRNA levels during ripening. *Proc Natl Acad Sci USA* 83: 6420–6424
- Facciotti D, O'Neal JK, Lee S, Shewmaker CK (1985) Light inducible expression of a chimeric gene in soybean tissue transformed with *Agrobacterium*. *Bio/Technology* 3: 241–246
- Grierson D, Slater A, Speirs J, Tucker GA (1985) The appearance of polygalacturonase in tomatoes: one of a series of changes in gene expression during development and ripening. *Planta* 163: 263–271
- Grierson D, Tucker GA, Keen J, Ray J, Bird CR, Schuch W (1986) Sequencing and identification of a cDNA clone for tomato polygalacturonase. *Nucleic Acids Res* 14: 8595–8603
- Gubler V, Hoffman BG (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263–269
- Heinrikson RL, Meredith SC (1984) Amino acid analysis by reverse phase high performance liquid chromatography: precolumn derivatization with phenylisothiocyanate. *Anal Biochem* 136: 65–74
- Huynh TV, Young RA, Davis RW (1985) Construction and screening cDNA libraries in λ gt10 and λ gt11. In: Glover DM (ed) *DNA cloning: A practical approach Vol I*. IRL Press Oxford, England, pp 49–78
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T₄. *Nature* 227: 680–685

- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Mansson PE, Hsu D, Stalker D (1985) Characterization of fruit-specific cDNAs from tomato. *Mol Gen Genet* 200:356–361
- Maxam A, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* 65:499–580
- Moshrefi M, Luh BS (1983) Carbohydrate composition and electrophoretic properties of tomato polygalacturonase isoenzymes. *Eur J Biochem* 135:511–514
- Reisfeld RA, Lewis US, Williams DE (1962) Disc electrophoresis of basic proteins and peptides on polyacrylamide gels. *Nature* 195:281–283
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5461–5463
- Sato T, Kusaba S, Nakagawa H, Ogura N (1984) Cell-free synthesis of a putative precursor of polygalacturonase in tomato fruits. *Plant Cell Physiol* 25:1069–1071
- Slater A, Mauders MJ, Edwards K, Schuch W, Grierson D (1985) Isolation and characterization of cDNA clones for tomato polygalacturonase and other ripening-related proteins. *Plant Mol Biol* 5:137–147
- Smeekens S, Bauerie C, Hageman J, Keegstra K, Weisbeck P (1986) The role of transit peptide in routing precursors toward different chloroplast compartments. *Cell* 46:365–375
- Speirs J, Brady CJ, Grierson D, Lee E (1984) Changes in ribosome organization and messenger RNA abundance in ripening tomato fruits. *Aust J Plant Physiol* 11:225–233
- Takashima K, Watanabe S, Yamada M, Mamiya G (1986) The amino acid sequence of the nonspecific lipid transfer protein from germinated castor bean endosperm. *Biochim Biophys Acta* 870:248–255
- Tigchelaar EC, McGlasson WB, Buescher RW (1978) Genetic regulation of tomato fruit ripening. *Planta* 155:64–67
- Tucker GA, Grierson D (1982) Synthesis of polygalacturonase during tomato fruit ripening. *Planta* 155:64–67
- Tucker ML, Christoffersen RE, Woll L, Laties GG (1985) Induction of cellulase by ethylene in avocado fruit. In: Roberts JA, Tucker GA (eds) *Ethylene and Plant Development*. Butterworths, London, pp 163–172
- von Heijne G (1984) How signal sequences maintain cleavage specificity. *J Mol Biol* 173:243–251
- Yanisch-Perron C, Vierra J, Messing J (1985) Improved M13 vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33:102–119
- Young RA, Davis RW (1983) Efficient isolation of genes by using antibody probes. *Proc Natl Acad Sci USA* 80:1194–1198

Communicated by R. Goldberg

Received December 8, 1986