Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10305–10309, October 1994
Plant Biology

Molecular cloning of a gene encoding an arabinogalactan-protein from pear (*Pyrus communis*) cell suspension culture

CHAO-GUANG CHEN*, ZHAO-YAN PU*, ROBERT L. MORITZ†, RICHARD J. SIMPSON†, ANTONY BACIC*, ADRIENNE E. CLARKE*‡, AND SHAIO-LIM MAU*

*Cooperative Research Centre for Industrial Plant Biopolymers and Plant Cell Biology Research Centre, University of Melbourne, Parkville, Victoria 3052, Australia; and †Joint Protein Structure Laboratory, Ludwig Institute for Cancer Research (Melbourne Branch), and The Walter and Eliza Hall Institute of Medical Research, Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

Contributed by Adrienne E. Clarke, June 21, 1994

ABSTRACT **Arabinogalactan-proteins (AGPs) are proteoglycans containing a high proportion of carbohydrate (typically >90%) linked to a protein backbone rich in hydroxyproline (Hyp), Ala, Ser, and Thr. They are widely distributed in plants and may play a role in development. The structure of the carbohydrate of some AGPs is known in detail but information regarding the protein backbone is restricted to a few peptide sequences. Here we report isolation and partial amino acid sequencing of the protein backbone of an AGP. This AGP is a member of one of four major groups of AGPs isolated from the filtrate of pear cell suspension culture. A cDNA encoding this protein backbone (145 amino acids) was cloned; the deduced protein is rich in Hyp, Ala, Ser, and Thr, which together account for >75% of total residues. It has three domains, an N-terminal secretion signal, a central hydrophilic domain containing all of the Pro residues, and a hydrophobic C-terminal domain that is predicted to be a transmembrane helix. Approximately 93% of the Pro residues are hydroxylated and hence are potential sites for glycosylation.**

Arabinogalactan-proteins (AGPs) occur predominantly in the intercellular spaces of plant tissues but are also associated with membranes, some cytoplasmic organelles, and the cell wall (for reviews see refs. 1–5). AGPs bind to and are precipitated by the β-glucosyl Yariv reagent (6). The function of AGPs is not established, but they may be involved in development, cell–cell interactions, and plant defense.

The carbohydrate component of AGPs is generally composed of arabinose and galactose with minor amounts of other sugars. Linkage analysis is consistent with a structure based on a 3-linked β-galactosyl backbone, branched through C(O)6 to 6-linked galactosyl side chains. The arabinose is most often present as terminal residues. The protein is usually a minor component with characteristically high levels of hydroxyproline (Hyp), Ala, and Ser (for exceptions see refs. 7 and 8). Relatively little is known about the structure of the protein core of AGPs; only a few peptide sequences are available (7, 9–11). In this paper, we describe the isolation of a cDNA encoding an AGP protein backbone, based on peptide sequences obtained from one of the AGPs present in culture filtrate of pear cells.§

METHODS

Isolation and Analysis of AGPs from Pear Cell Culture Filtrate. Pear (*Pyrus communis*) cell suspension culture was initiated from fruit (12). The culture medium was separated from cells and depleted of pectins (13). High molecular mass material was precipitated with ethanol (4 vol) and dissolved in water, and AGPs were precipitated with the β-glucosyl Yariv reagent (14). The precipitate was dissociated with $Na_2S_2O_4$, desalted, freeze-dried, and dissolved in 6 M guanidinium hydrochloride/0.1 M Tris·HCl, pH 8.0. The sample was applied to a Superdex-75 FPLC column equilibrated with 0.1 M Tris·HCl, pH 8.0/8 M urea and run in the same buffer (15). The V_o fraction was collected, desalted, and freeze-dried.

AGPs were dissolved in 0.1% trifluoroacetic acid (TFA) and fractionated by reversed-phase (RP) HPLC and size-exclusion FPLC as shown in Fig. 1 *A–C*. AGPs were detected and quantified by the gel diffusion test against the β-glucosyl Yariv reagent (16). Carbohydrate linkages were analyzed by methylation and GC/MS (17).

AGPs were deglycosylated using anhydrous HF (18) and fractionated by size-exclusion FPLC and RP-HPLC according to Fig. 1 *D–E*. The protein backbones were digested with thermolysin, and the products were separated on a C_{18} microbore HPLC column (Ultrasphere ODS, 2.1 × 250 mm) and eluted with a gradient of acetonitrile in 0.1% aqueous TFA. Individual peaks were repurified and sequenced (19). Amino acid analyses were performed as described by Simpson *et al.* (20).

Isolation of cDNA Clones. A 68-base oligonucleotide, 5′-GCAAAATCACCAACAGCAACACCACCAACAGCAACACCACCATCAGCAGTATATAGTGAGTCGTATTA-3′, was synthesized. The first part of the sequence codes for the AGP peptide A-K-S-O-T-A-T-O-O-T-A-T-O-O-S-A-V (Table 3), using A in the third codon position for all amino acids and the codon TCA for Ser residues. The second part of the sequence (underlined) corresponds to the T7 RNA polymerase promoter (in reverse orientation). This oligonucleotide was annealed to a complementary oligonucleotide to form a double-stranded DNA fragment, from which an antisense RNA probe was synthesized using the T7 RNA polymerase and [α-^{32}P]UTP.

A cDNA library in λZap (Stratagene) was constructed from mRNA isolated from pear suspension culture cells and screened by overnight hybridization with the RNA probe at 42°C in 2× SSPE (standard saline/phosphate/EDTA)/50% deionized formamide/0.5% dried milk powder/1% SDS/50 μg of denatured salmon sperm DNA per ml/10% PEG-6000 (21). Filters were washed at 55°C twice for 15 min in 2× SSC (standard saline citrate)/0.1% SDS and twice for 15 min in 1× SSC/0.1% SDS before exposure to film. Positive clones were sequenced.

RESULTS

Separation and Partial Characterization of AGPs. The majority (72%) of AGPs loaded onto the RP-HPLC column was recovered in the unbound fraction. Most of the bound ma-

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

Abbreviations: AGP, arabinogalactan-protein; Hyp or O, hydroxyproline; RP, reversed-phase.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. U14009).

terial was recovered in a major peak (retention time, 6.43 min) with a trailing edge (Fig. 1*A*). This peak contained 27% of the AGPs loaded onto the column, and two minor peaks (retention time, 14.95 and 30.64 min) accounted for 0.1% and 0.9%, respectively.

The major peak was collected, reapplied to the same column, and eluted with a shallow gradient (Fig. 1*B*). Two major peaks (fractions 1 and 2) were resolved, separately collected, and subjected to size-exclusion FPLC. Fraction 1 eluted as a single broad peak (Fig. 1*C2*) and fraction 2

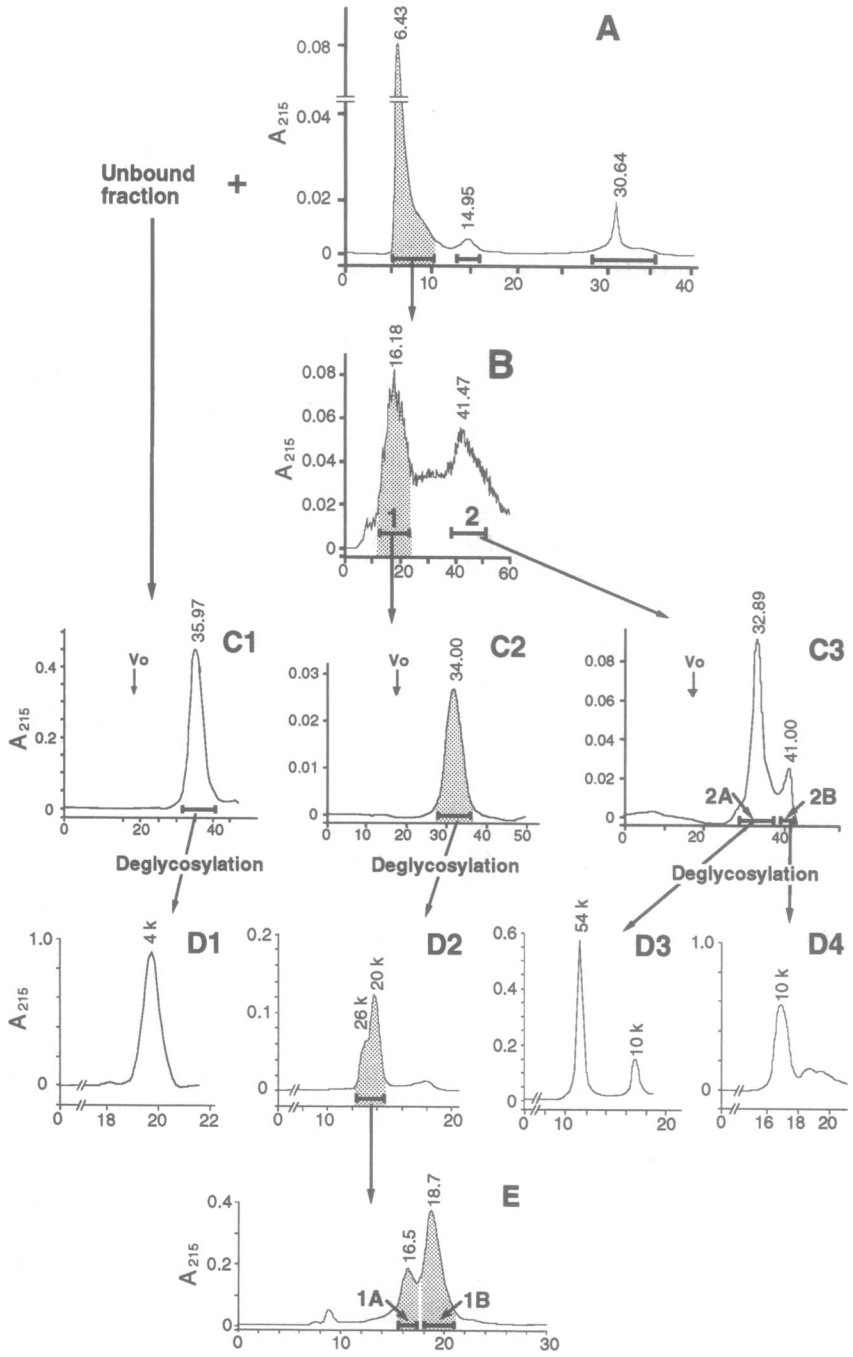

FIG. 1. Separation of AGPs and isolation of their protein backbones. (*A*) RP-HPLC (RP-300 column, 4.6 × 100 mm) profile of AGPs prepared by precipitation with the β-glucosyl Yariv reagent. AGPs were loaded and the column washed with solvent A (0.1% TFA in water). The unbound fraction was collected (not shown). The bound material was eluted with a linear gradient (0–100% solvent B; flow rate, 1 ml/min; 60 min) (solvent B: 60% acetonitrile in solvent A). Individual fractions from five separate runs were pooled for subsequent purification. (*B*) RP-HPLC (RP-300 column, 4.6 × 100 mm) profile of AGPs from the major bound peak shown in *A* (retention time, 5.0–10.57 min). Bound material was eluted with a shallow gradient (0–15% solvent B; flow rate, 1 ml/min; 60 min). Two fractions (1 and 2) were separately collected and subjected to size-exclusion FPLC. (*C*) Superose-6 FPLC profiles of AGPs in the unbound fraction from *A* and two eluted fractions from *B*. Samples were eluted in 25% acetonitrile/0.2 M KCl/5 mM KH_2PO_4 (flow rate, 0.4 ml/min). The unbound fraction and fraction 1 gave single peaks; fraction 2 resolved into two peaks (peaks 2A and 2B). (*D*) Superdex-75 FPLC profiles of protein backbones derived from AGPs in *C* by HF deglycosylation. Samples were eluted in the same buffer used in *C* (flow rate, 0.8 ml/min). The size of the protein was estimated from standard protein markers (Pharmacia). (*E*) RP-HPLC (RP-300 column, 4.6 × 100 mm) profile of protein backbones from *D2*. Combined fractions from *D2* were loaded and material was eluted with a linear gradient as described for *A*. Two peaks (peaks 1A and 1B) were collected for N-terminal sequencing and thermolysin digestion. The *x* axis is retention time (min). The pathway for purification of the AGP fractions, from which peptide sequences were obtained, is stippled.

resolved into two components (peaks 2A and 2B, Fig. 1*C3*). Ion-exchange chromatography did not resolve fraction 1 into further components (not shown). The original unbound fraction from RP-HPLC was also subjected to size-exclusion FPLC and eluted as a single broad peak (Fig. 1*C1*). Arabinose and galactose were the major monosaccharides of each fraction (Table 1). Arabinose was present mainly in the terminal position with small amounts of 3-linked and 5-linked residues. Galactose was present mainly as 3,6-linked residues with smaller amounts of 3-linked, 6-linked, and terminal residues.

Each fraction from size-exclusion FPLC (Fig. 1 *C1–C3*) was separately deglycosylated, and the resulting protein backbones were isolated by size-exclusion FPLC (Fig. 1 *D1–D4*). The apparent molecular mass of the proteins was different for each fraction. The unbound fraction gave a single peak (4 kDa; Fig. 1*D1*), which resolved into many proteins by RP-HPLC (not shown), and was not investigated further. The present study focuses on the protein backbone of fraction 1 [protein backbones from peaks 2A and 2B (54 kDa and 10 kDa, respectively) require further investigation].

Isolation and Analysis of Protein Backbones of AGPs Present in Fraction 1. RP-HPLC fractionation of the protein backbones derived from fraction 1 gave two poorly resolved peaks (20 kDa and 26 kDa) (Fig. 1*D2*). The material of both peaks was pooled and subjected to RP-HPLC and, again, two major, poorly resolved peaks were obtained (peak 1A, 16.5 min; peak 1B, 18.7 min) (Fig. 1*E*). The composition of fraction 1 and its two deglycosylated products was similar; 75% of the amino acids were Hyp, Ser, Ala, and Thr (Table 2). The mass of protein deduced from amino acid analysis of fraction 1 accounted for <1% of the total mass of AGPs in this fraction measured by the gel diffusion test. N-terminal amino acid sequencing of material in peak 1B gave the sequence Q-A-O-X-A-A; no sequence was obtained from material in peak 1A. The two components were digested separately with thermolysin. Sequences of three peptides were obtained from material in peak 1A and four were obtained from material in peak 1B; one sequence was common to both peaks (Table 3).

Isolation of a cDNA Encoding an AGP Protein Backbone. A pear cDNA library was screened with the antisense RNA probe encoding the AGP peptide and gave four positive clones. The sequence of the longest was 893 bp (Fig. 2). This cDNA is referred to as AGP*Pc*1 cDNA (indicating cDNA 1 from *P. communis*). Two other clones had sequences identical to the AGP*Pc*1 cDNA but were shorter at the 5′ end. The fourth clone had an unrelated sequence (not shown).

The AGP*Pc*1 cDNA sequence encodes a polypeptide of 145 residues, with a putative signal sequence of 23 amino acids (Fig. 3) and a predicted cleavage site between Ala^{23} and Glu^{24} (23). The predicted molecular mass of the mature protein is 11 kDa. The N-terminal sequence of the deduced mature protein, Q-A-P-G-A-A, matches that obtained from N-terminal sequencing of material in peak 1B, Q-A-O-X-A-A. In addition, the deduced amino acid sequence contains two stretches of sequence, A-K-S-P-T-A-T-P-P-T-A-T-P-P-S-A-V and V-T-A-P-T-P-S-A-S-P-P-S-S-T-P-A-S-T-P-A, which match the internal peptide sequences obtained from the same peak (Table 3). None of the peptide sequences obtained from material in peak 1A are represented in this cDNA. The sequence of the mature protein predicted from the cDNA has a high content of Pro (24.5%), Ala (21.3%), Ser (19.6%), and Thr (13.1%); the composition closely matches that obtained from amino acid analysis of fraction 1 before and after deglycosylation (assuming hydroxylation of the Pro residues, Table 2). All Pro residues deduced from the cDNA are modified to Hyp in the peptide sequences, and amino acid analysis (Table 2) shows that 93% of the Pro residues in the whole molecule are hydroxylated. There are no obvious motifs in the sequence; the Pro, Ala, Ser, and Thr residues are interspersed with each other and there are few runs of any single amino acid. There are no predicted N-glycosylation sites. The codon usage for Pro is strongly biased toward CCA (73.3%); the codon for Ala is biased to a lesser extent toward GCT (44.8%); there is no significant bias in codon usage for other amino acids.

The C-terminal region of 22 amino acid residues is hydrophobic (Fig. 3) and predicted to be a transmembrane helix

Table 1. Linkage analysis of AGP fractions

Monosaccharide and deduced linkage	mol % Unbound fraction (Fig. 1*C1*)	Fraction 1 (Fig. 1*C2*)	Fraction 2 (Fig. 1*C3*) 2A	Fraction 2 (Fig. 1*C3*) 2B
Ara*f*				
Terminal	34	36	24	18
3-	3	3	4	4
5-	2	3	1	1
Gal*p*				
Terminal	7	8	12	14
3-	5	4	8	5
6-	10	10	8	23
3,6-	38	36	44	35

Ara*f*, arabinofuranose; Gal*p*, galactopyranose.

Table 2. Amino acid composition (mol %) of fraction 1 AGP, its deglycosylated backbones, and the protein deduced from the AGP*Pc*1 cDNA

Amino acid	Fraction 1 (Fig. 1*C2*)	Peak 1A (Fig. 1*E*)	Peak 1B (Fig. 1*E*)	Deduced protein*
Hyp	28.4	24.5	24.4	
Pro	2.1	1.3	1.8	24.5
Asx	2.8	3.1	2.9	1.6
Glx	4.9	4.9	3.6	0.8
Ser	20.4	22.1	17.6	19.6
Gly	3.0	2.3	4.0	4.1
His	0.3	0.5	0.8	0.0
Arg	0.6	1.2	0.6	0.8
Thr	10.1	9.3	10.8	13.1
Ala	18.5	19.7	21.6	21.3
Tyr	0.4	1.0	0.3	0.0
Val	3.2	3.2	4.1	4.9
Met	0.0	0.0	0.0	0.0
Ile	0.4	0.2	0.8	3.2
Leu	1.9	1.7	1.9	1.6
Phe	0.0	0.0	0.1	1.6
Lys	3.1	5.1	4.7	2.4
Cys	ND	ND	ND	0.0
Trp	ND	ND	ND	0.0

ND, not determined.
*Excluding the 23-amino acid signal sequence.

Table 3. Peptide sequences from fraction 1

Peak	Sequence
1A	L-S-O-K-K-S-O-T-A-O-S-O-S-(S)-T-O-O-T-(T)
	V/S-P/S-X-O-V-Q-S-O-A-S-O-O-O-T-(T)
	X-X-O-O-A-A-O-(V)-X-A-O/S
1B	Q-A-O-X-A-A* (N-terminal)
	A-K-S-O-T-A-T-O-O-T-A-T-O-O-S-A-V*
	V-T-A-O-T-O-S-A-S-O-O-S-S-T-O-A-(S)-T-X-A*
	V-T-A-O-T-O-S-A-S-O-O-S-S-T*
	L-S-O-K-K-S-O-T-A-O-S-O-S-(S)-T-O-O-T

All residues of ambiguous assignments are shown; uncertain residues are in parentheses. "X" indicates no signal or an unknown residue. "O" represents hydroxyproline.
*Sequences included in the cDNA.

```
CTCTCTCGCTCACTCATCAAATTCTCTCTCTCTCTCTCTCTCTCTCTTTCTCTCTCTCTC  60
TCTCTAAAAATGAAGATGGGTTTTGCAGGGTTCCAAGTTTTGATGGTTTTGGGTCTGTTG 120
         M  K  M  G  F  A  G  F  Q  V  L  M  V  L  G  L  L    17

GCCACATCATGCATAGCCCAAGCCCCAGGAGCAGCACCCACAGCTTCACCCCCAACCGCA 180
A  T  S  C  I  A  Q  A  P  G  A  A  P  T  A  S  P  P  T  A    37
                  ↑     O  X

AAGTCGCCAACCGCCACCCCACCAACCGCCACACCGCCATCAGCCGTACCAGTTCCATCA 240
K  S  P  T  A  T  P  P  T  A  T  P  P  S  A  V  P  V  P  S    57
      O           O  O           O  O

CCCAGCAAAACACCAACCGCGTCACCAACTCCATCACCAGTGACAGCACCAACCCCAAGT 300
P  S  K  T  P  T  A  S  P  T  P  S  P  V  T  A  P  T  P  S    77
                                          O     O

GCCTCCCCACCATCTTCCACACCAGCTTCCACCCCAGCTTCCACTCCAGCAGCTAAGTCT 360
A  S  P  P  S  S  T  P  A  S  T  P  A  S  T  P  A  A  K  S    97
      O  O           O           X

CCATCGTCGTCAGCTGCTCCCTCAGGCTCAAGCCCGAACTCCCCACCGGCTGACGCTATT 420
P  S  S  S  A  A  P  S  G  S  S  P  N  S  P  P  A  D  A  I   117

CCTCCAAGTGGCACCTCCGCCATCAGCCGCGTTGCTATTGCTGGAACTGCTCTTGCTGGA 480
P  P  S  G  T  S  A  I  S  R  V  A  I  A  G  T  A  L  A  G   137

GTTTTCTTCGCGATTGTGTTGGCTTAGATTCATGGGATTTGCTCTTTCGGGTTTTCCTAT 540
V  F  F  A  I  V  L  A  ***                                  145

TGGTCCACGTGGAGACTCACATCTGCTCTTAGATCTGGGTTTTGATGGACGGTCGAGATC 600
TATTAATTTCTTTTTATTTTGTTGCTTATTTTCGTAATGTTTTTTGTATTTTTGTTTAAC 660
TCTGTTTTCATGCCATATGGTGATTATTGGTTTGGCAGTCTATGGTGGATTTGGACGGTC 720
GTGATGTGATTAATTATGGTGATTCATTGTTTTAGAGTTGACAAGTGCACCCATTTGTAG 780
ATGAGTCGTTGGATGTACATCTGTCCGATCATAGTTTAATAAAACAGTTTGTCATTCTTT 840
TTCTTATGGATCTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  893
```

FIG. 2. Nucleotide and predicted amino acid sequences of AGP*Pc*1 cDNA. The putative secretion signal peptide is underlined with dots. Deduced peptide sequences that match peptide sequences obtained experimentally are underlined; the Pro residues modified to Hyp are indicated by "O." Residues not assigned at sequencing are represented as "X." The putative poly(A) signal is double-underlined. The predicted transmembrane helix is in italics. The stop codon is indicated by ***.

(24, 25). The border between the C-terminal transmembrane helix and the extracellular domain contains several potential sites for proteolytic cleavage.

Expression of Sequences Related to AGP*Pc*1 in the Tissues of Various Plants. At low stringency, AGP*Pc*1 cDNA hybridized to transcripts of approximately 0.9 and 1.8 kb in all plants analyzed. The 1.8-kb transcript probably corresponds to 18S rRNA, and, at high stringency, hybridization to this species was lost. The transcript of 0.9 kb was detected at high stringency, only in lanes containing RNA from cultured cells (Fig. 4).

DISCUSSION

Evidence for Multiple AGPs in the Pear Cell Culture Filtrate. AGPs in extracts of plant tissues commonly appear as smears at the high molecular mass range in SDS/PAGE gels and often resolve into several overlapping components by size-exclusion or ion-exchange chromatography. This may reflect either true heterogeneity in the protein backbones or in glycosyl substituents or technical limitations in separating polydisperse proteoglycans. In the present study, several groups of AGPs were separated from the pear cell culture filtrate (Fig. 1). One of the groups of AGPs (fraction 1) appeared as a single component during a number of chromatographic procedures but gave two protein backbones of different apparent molecular mass on deglycosylation (Fig. 1 *D2* and *E*). The cDNA obtained encodes one of these proteins

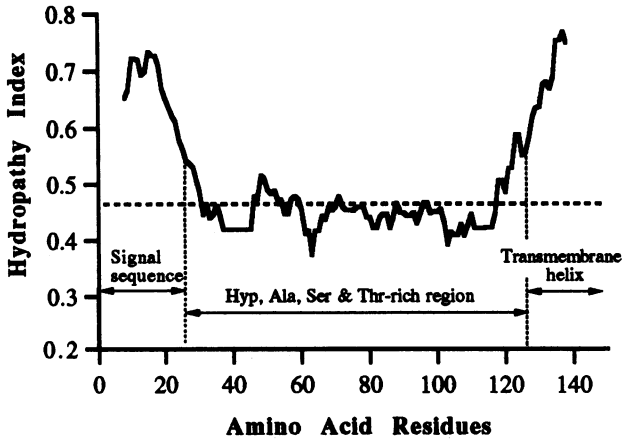

FIG. 3. Hydropathy plot of the deduced amino acid sequence of AGP*Pc*1. The hydropathy values of each amino acid were determined using an interval of 15 amino acids (22). The scale is normalized. Values above the dashed line indicate hydrophobic regions, and values below the dashed line represent hydrophilic regions.

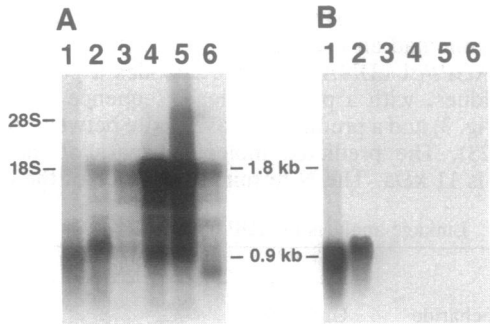

FIG. 4. Detection of mRNA related to AGP*Pc*1 in other plants. Total RNA was isolated from cultured cells of *P. communis* (lanes 1) and *Nicotiana plumbaginifolia* (lanes 2), shoots of *Brassica napus* (lanes 3), *Arabidopsis thaliana* (lanes 4), and *Lycopersicon esculentum* (lanes 5), and leaves of *Lolium temulentum* (lanes 6). Equal amounts of RNA (10 μg per lane) were fractionated on a formaldehyde agarose gel, transferred to a Hybond-N membrane, and hybridized with ^{32}P-labeled AGP*Pc*1 cDNA. (*A*) Hybridization and washing were performed at 55°C as described previously for library screening. (*B*) Hybridization and washing at 65°C. The amount of RNA loaded and the integrity of rRNA were confirmed by ethidium bromide staining. Sizes of the transcripts (0.9 and 1.8 kDa) and positions of the rRNA bands (18S and 28S) are indicated.

but does not include the amino acid sequences obtained from the other. The close association of these two AGPs through many chromatographic procedures and the similarity of their deglycosylated proteins demonstrate the difficulty of isolating an AGP with confidence that it represents a single species.

Isolation of a cDNA Encoding an AGP Protein Backbone. Several peptide sequences of AGPs, but no corresponding cDNAs, have been reported previously (7, 9–11). The difficulties in obtaining cDNAs derive from the high level of redundancy and high GC content of the codons for the major amino acids present, Ala, Pro, Ser, and Thr. In designing our probes, we reduced the redundancy by assigning only one codon (TCA) for Ser and by using A for the third codon positions for the other amino acids, thus creating a long "guessmer" oligonucleotide. This guess was based on the strong bias in dicots to A in the third position of Pro codons and the slight preference for A and T for other amino acids (26). A long oligonucleotide would tolerate some mismatches and the problems of mismatch were further reduced by using an antisense RNA probe; all the anticodons would thus have U at their third position, which binds A and to lesser extent G (21). This possible binding of U to two nucleotides also influenced our choice of A in the third codon positions in the guessmer. Another advantage of using RNA probes is that they can be labeled to a higher specific activity than oligonucleotide probes. The strategy was successful; the protein encoded by AGP*Pc*1 cDNA has an amino acid composition similar to that of fraction 1 and included the sequence of four peptides detected in the deglycosylated material.

The deduced protein sequence has three distinct domains (Figs. 2 and 3). The predicted signal sequence indicates that the molecule is secreted; N-terminal sequencing of the protein confirms that the signal is cleaved. The central part of the molecule contains all of the Pro residues, most of which are hydroxylated and are thus potential sites for O-glycosylation (27, 28). The abundance of Hyp residues is consistent with the high carbohydrate content (≈99%).

The third domain, at the C terminus, is hydrophobic and is predicted to be a transmembrane helix. This suggests that the molecule could be anchored within the plasma membrane, with the central part exposed at the extracellular face. In this case, the AGP purified from the culture filtrate could be derived from a membrane-bound precursor by proteolysis. Some plasma membrane-associated AGP epitopes are developmentally regulated (29–31), and proteolytic cleavage to release the extracellular domain could be responsible for its loss from the plasma membrane at a particular developmental stage. The hydrophilic and hydrophobic domains could also confer micelle-forming properties on the AGP.

There are four families of Hyp-rich proteins in plants: the extensins, the Hyp/Pro-rich proteins, the solanaceous lectins, and the AGPs (1–5, 32). The AGP protein backbone encoded by AGP*Pc*1 differs from that of other plant Hyp-rich proteins in that it lacks the Ser-Hyp_4 motif and high Tyr content of the extensins; the high Cys content of solanaceous lectins and the Pro-Hyp-Xaa-Yaa-Lys motifs of the Hyp/Pro-rich proteins. The dominant feature of the AGP sequence is that the four major residues (Pro, Ala, Ser, and Thr) are interspersed with each other, with few runs of any particular amino acid and no obvious motifs. The Ala-Hyp motif present in some AGP peptides (3, 5) is found for only 4 of the 26 Ala residues. RNA blot analysis shows that sequences related to AGP*Pc*1 are present in other plants, but strong hybridization was observed only in RNA of cultured cells.

This isolation of a cDNA encoding an AGP backbone now allows us to address the questions of their role as markers of cellular identity (30, 31) and inducers of differentiation and embryogenesis (29).

We thank Drs. Steve Read and Ed Newbigin for critical review of the manuscript and Mss. Judith Webster and Alison Gane and Drs. Ian Sims and David Oxley for helpful discussions. Dr. David McManus kindly provided the culture. Ms. Eva Lau performed the methylation analyses.

1. Clarke, A. E., Anderson, R. L. & Stone, B. A. (1979) *Phytochemistry* **18**, 521–540.
2. Fincher, G. B., Stone, B. A. & Clarke, A. E. (1983) *Annu. Rev. Plant Physiol.* **34**, 47–70.
3. Showalter, A. M. & Varner, J. E. (1989) in *The Biochemistry of Plants*, ed. Marcus, A. (Academic, New York), Vol. 15, pp. 485–520.
4. Pennell, R. I. (1992) in *Perspectives in Plant Cell Recognition*, eds. Callow, J. A. & Green, J. R. (Cambridge Univ. Press, Melbourne, Australia), pp. 105–121.
5. Showalter, A. M. (1993) *Plant Cell* **5**, 9–23.
6. Yariv, J., Lis, H. & Katchalski, E. (1967) *Biochem. J.* **105**, 1–2.
7. Kieliszewski, M. J., Kamyab, A., Leykam, J. F. & Lamport, D. T. A. (1992) *Plant Physiol.* **99**, 538–547.
8. Baldwin, T. C., McCann, M. C. & Roberts, K. (1993) *Plant Physiol.* **103**, 115–123.
9. Jermyn, M. A. & Guthrie, R. (1985) in *AGP News*, ed. Clarke, A. E. (Univ. of Melbourne, Melbourne, Australia), Vol. 5, pp. 4–25.
10. Gleeson, P. A., McNamara, M., Wettenhall, T. E. H., Stone, B. A. & Fincher, G. B. (1989) *Biochem. J.* **264**, 857–862.
11. Komalavilas, P., Zhu, J.-K. & Nothnagel, E. A. (1991) *J. Biol. Chem.* **266**, 15956–15965.
12. Pech, J.-C., Latché, A. & Fallot, J. (1979) *Physiol. Plant* **46**, 260–264.
13. Scott, J. E. (1965) in *Methods in Carbohydrate Chemistry*, ed. Whistler, R. L. (Academic, New York), Vol. 5, pp. 38–44.
14. Clarke, A. E., Gleeson, P. A., Jermyn, M. A. & Knox, R. B. (1978) *Aust. J. Plant Physiol.* **5**, 702–722.
15. Gane, A. (1994) Ph.D. thesis (Univ. of Melbourne, Melbourne, Australia).
16. van Holst, G.-J. & Clarke, A. E. (1985) *Anal. Biochem.* **148**, 446–450.
17. Lau, E. & Bacic, A. (1993) *J. Chromatogr.* **637**, 100–103.
18. Mort, A. J. & Lamport, D. T. A. (1977) *Anal. Biochem.* **82**, 289–309.
19. Grego, B., van Driel, I. R., Stearne, P. A., Goding, J. W., Nice, E. G. & Simpson, R. J. (1985) *Eur. J. Biochem.* **148**, 485–491.
20. Simpson, R. J., Moritz, R. L., Nice, E. C., Grego, B., Yoshizaki, F., Sugimura, Y., Freeman, H. C. & Murata, M. (1986) *Eur. J. Biochem.* **157**, 497–506.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
22. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
23. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
24. Klein, P., Kanehisa, M. & DeLisi, C. (1985) *Biochim. Biophys. Acta* **815**, 468–476.
25. Rao, J. K. M. & Argos, P. (1986) *Biochim. Biophys. Acta* **869**, 197–214.
26. Murray, E. E., Lotzer, J. & Eberle, M. (1989) *Nucleic Acids Res.* **17**, 477–498.
27. McNamara, M. K. & Stone, B. A. (1981) *Lebensm.-Wiss. Technol.* **14**, 182–187.
28. Kawasaki, S. (1987) *Plant Cell Physiol.* **28**, 925–935.
29. Kreuger, M. & van Holst, G.-J. (1993) *Planta* **189**, 243–248.
30. Pennell, R. I., Janniche, L., Kjellbom, P., Scofield, G. N., Peart, J. M. & Roberts, K. (1991) *Plant Cell* **3**, 1317–1326.
31. Knox, J. P., Linstead, P. J., Peart, J., Cooper, C. & Roberts, S. K. (1991) *Plant J.* **1**, 317–326.
32. Kieliszewski, M. J. & Lamport, D. T. A. (1994) *Plant J.* **5**, 157–172.