

The Classical Arabinogalactan Protein Gene Family of Arabidopsis

Carolyn J. Schultz,^a Kim L. Johnson,^b Graeme Currie,^b and Antony Bacic^{a,b,1}

^a Cooperative Research Centre for Bioproducts, School of Botany, University of Melbourne, Parkville, Victoria 3010, Australia

^b Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, Victoria 3010, Australia

Arabinogalactan proteins (AGPs) are extracellular proteoglycans implicated in plant growth and development. We searched for classical AGPs in Arabidopsis by identifying expressed sequence tags based on the conserved domain structure of the predicted protein backbone. To confirm that these genes encoded bona fide AGPs, we purified native AGPs and then deglycosylated and deblocked them for N-terminal protein sequencing. In total, we identified 15 genes encoding the protein backbones of classical AGPs, including genes for AG peptides—AGPs with very short backbones (10 to 13 amino acid residues). Seven of the AGPs were verified as AGPs by protein sequencing. A gene encoding a putative cell adhesion molecule with AGP-like domains was also identified. This work provides a firm foundation for beginning functional analysis by using a genetic approach.

INTRODUCTION

Arabinogalactan proteins (AGPs) make up a large family of proteoglycans that have been implicated in various processes associated with plant growth and development, including embryogenesis and cell proliferation (Knox, 1995; Nothnagel, 1997). Much of the evidence relating to AGP function has been based on the use of monoclonal antibodies that react with carbohydrate epitopes on AGPs (Knox, 1997; McCabe et al., 1997; Toonen et al., 1997; Casero et al., 1998; and references therein). Although these carbohydrate-directed monoclonal antibodies are useful for investigating the distribution of AGPs in cells, such epitopes are likely to be present on many AGPs with different protein backbones rather than on a single AGP (Nothnagel, 1997). The only antibody that recognizes the protein backbone of both the native and deglycosylated forms of a single AGP is an antibody for LeAGP1, a classical AGP from tomato (Gao et al., 1999). This antibody was generated against a peptide that includes a 15-amino acid lysine-rich domain containing no proline residues. Unfortunately, attempts to generate antibodies to other classical AGP protein backbones are unlikely to be successful because of the high degree of glycosylation of the proteins. For example, the classical AGPs from *Pyrus communis* and *Nicotiana glauca*, PcAGP1 and NaAGP1, respectively (Chen et al., 1994; Du et al., 1994), have Pro/Hyp residues (the presumed sites of carbo-

hydrate attachment in the AGP protein backbone) at least every seventh amino acid residue. Therefore, generating antibodies that recognize the backbones of other classical AGPs may not be possible.

Other studies of AGPs have utilized the specific reaction between AGPs and the β -glucosyl Yariv reagent (β -GlcY) to study AGPs (Yariv et al., 1962). This reagent has several uses: (1) to determine the cellular distribution of AGPs by specific staining, (2) to purify AGPs by selective precipitation, and (3) to determine the effect of cross-linking of AGPs in living plants (reviewed in Nothnagel, 1997). The cross-linking effect of β -GlcY inhibits cell growth in suspension-cultured cells of rose (Serpe and Nothnagel, 1994). In Arabidopsis, β -GlcY reduces root growth and alters the morphology of epidermal cells (Willats and Knox, 1996). In lily, β -GlcY alters the structure of the pollen tube cell wall (Roy et al., 1998). Although the precise effect of the cross-linking of cell surface AGPs is unknown, perhaps the formation of large complexes at the cell surface prevents the normal assembly of molecules into the cell wall (Roy et al., 1998).

Both approaches for investigating AGP function have the drawback that they cannot be used to distinguish single AGPs—that is, all the AGP glycoforms with a single protein backbone. A genetic approach offers an alternative for determining AGP function. However, designing mutant screens specific to AGPs is difficult because the function of the molecules is not clearly defined. A few groups have succeeded in identifying putative AGP mutants when searching for other phenotypes. One AGP mutant, *rat1* (*resistant to Agrobacterium transformation*), was identified with a T-DNA tag in the promoter region of an AGP gene (Nam et al., 1999;

¹ To whom correspondence should be addressed. E-mail a.bacic@botany.unimelb.edu.au; fax 61-3-9347-1071.

Y. Gaspar, P. Gilson, S. Gelvin, and A. Bacic, unpublished results). Two mutants identified have decreased AGP contents, *diminuto* (*dim*) (Takahashi et al., 1995) and *root epidermal bulger* (*reb1-1*) (Ding and Zhu, 1997). *dim* mutants are reportedly defective in steroid biosynthesis and can be rescued by the addition of brassinolide (Klahre et al., 1998). Therefore, any reduction in the amount of AGPs in *dim* mutants is probably a secondary effect. The suggestion of a relationship between AGP and phenotype is stronger for *reb1-1* mutants. This mutant was originally identified on the basis of a root-swelling phenotype (Baskin et al., 1992), which can be mimicked by growing wild-type Arabidopsis in the presence of β -GlcY (Ding and Zhu, 1997). Whether the gene affected in *reb1-1* mutants is required for the synthesis of a particular AGP (i.e., is a gene encoding the protein backbone) or for the post-translational modification of the protein backbone of AGPs (i.e., for prolyl hydroxylation or glycosylation), or if indeed the gene is involved in AGP synthesis/processing at all, is not known.

The finding of glycosylphosphatidylinositol (GPI) anchors on AGPs (Youl et al., 1998; Oxley and Bacic, 1999; Sherrier et al., 1999; Svetek et al., 1999) offers a new framework for considering the function of AGPs. GPI anchors provide an alternative to transmembrane domains for anchoring proteins to cell surfaces. In plants, as in other eukaryotes, GPI anchors are found on many different proteins (Stöhr et al., 1995; Kunze et al., 1997; Takos et al., 1997; Sherrier et al., 1999). For some proteins, GPI anchors lead to increased lateral mobility in the membrane, to polarized transport to the apical surface of cells, or to exclusion from clathrin-coated pits (Hooper, 1997). Several GPI-anchored proteins from animals are implicated in signal transduction pathways (Peles et al., 1997; Kleeff et al., 1998; Resta et al., 1998). In these examples, signal transduction occurs through interactions with other membrane-bound proteins. Another possible mechanism by which GPI-anchored proteins may be involved in signaling is by the phospholipase-mediated cleavage of the protein from its lipid anchor (Udenfriend and Kodukula, 1995a). This has the potential to generate both intra- and extracellular messengers by way of the lipid anchor or extracellular proteoglycan components, respectively. Structural characterization of the remnants of the GPI anchor present on PcAGP1 purified from the culture medium of suspension-cultured cells of pear suggests that the membrane-bound form is released by the action of a phospholipase (Oxley and Bacic, 1999). Proof that AGPs are indeed involved in cell signaling requires further experimentation.

Isolation of AGP mutants by reverse genetics techniques is one way to determine the function of AGPs. Most AGP genes cloned thus far have been from plants species that are either poorly suited to genetic analysis (e.g., pear and pine) or lack a well-developed system for reverse genetics experiments (e.g., tobacco and tomato). The study of AGPs in Arabidopsis offers an opportunity to identify AGP mutants by using the tools available through the Arabidopsis Biologi-

cal Resource Center (ABRC), such as T-DNA-tagged lines (Azpiroz-Leehan and Feldman, 1997; Campisi et al., 1999). Before the search for mutants can begin, we must characterize the AGPs of Arabidopsis biochemically.

The starting point for the search for Arabidopsis AGPs was a collection of five expressed sequence tags (ESTs) representing putative AGP protein backbone genes, *AtAGP1* to *AtAGP5*, previously identified by Schultz et al. (1998). These ESTs were identified on the basis of structural features of the deduced protein backbone. A sixth putative Arabidopsis AGP gene was recently identified by Sherrier et al. (1999). The proteins encoded by each of these six clones have all the features of classical AGPs: an N-terminal signal sequence and a region rich in Pro/Hyp, Ala, Ser, and Thr, followed by a C-terminal signal for the addition of a GPI anchor.

Our work combines a proteomic approach with a genomic approach to confirm that the putative Arabidopsis AGP genes encode bona fide AGPs. Seven protein backbones were identified by N-terminal protein sequencing of β -GlcY-precipitated and chemically deglycosylated AGPs. We identified 16 different AGP genes from the DNA sequence databases, including a group of AG peptides with very short protein backbones. Expression studies showed that AGPs are found in all Arabidopsis tissues.

RESULTS

Arabidopsis Has Many AGP Protein Backbones

AGPs were purified from Arabidopsis leaves and roots by using a method to selectively purify both plasma membrane-bound and soluble AGPs. The solubilized AGPs were precipitated with β -GlcY and subsequently separated by reversed-phase (RP)-HPLC. Figure 1A shows that native AGPs were eluted from the RP-HPLC column as several poorly resolved peaks. Multiple RP-HPLC runs were performed for both leaf and root tissues. The material in the fractions obtained by RP-HPLC were analyzed by SDS-PAGE (data not shown). Staining the gels with β -GlcY indicated that AGPs were present in each fraction. The material in these fractions did not stain with Coomassie blue, however, indicating that few (if any) contaminating proteins were present in each fraction.

Our experience of sequencing AGP protein backbones has been that RP-HPLC peaks from different tissues but having the same retention times can contain the same AGPs. To ensure sufficient material for amino acid sequence analysis, we pooled the root and leaf AGPs with retention times between 6.5 and 8 min (Figure 1A, peak 2). The other major root fractions (peaks 1 and 3) were analyzed separately. Each fraction was chemically deglycosylated and enzymatically "deblocked" to remove modified glutamine (pyroglutamate) residues at the N termini of the protein

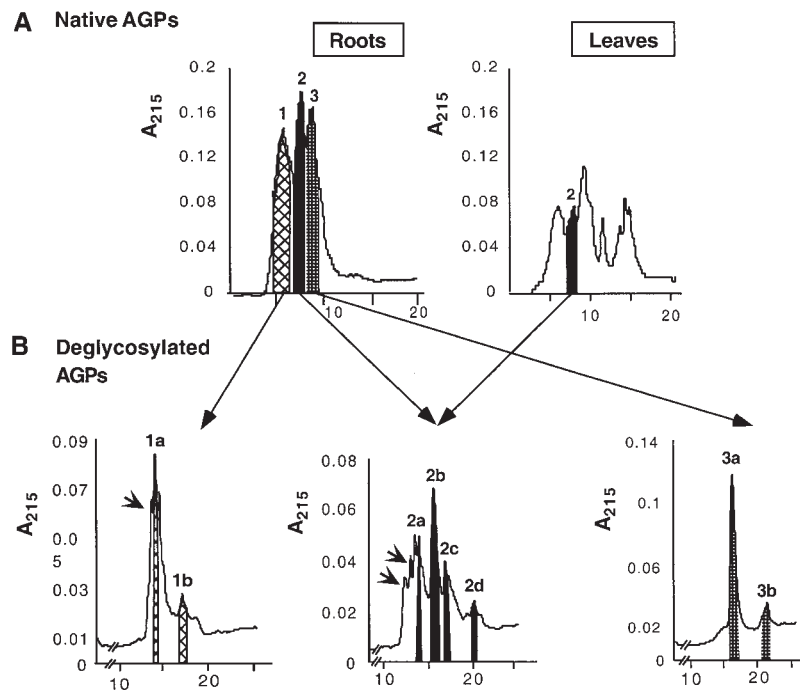


Figure 1. Separation of AGP Protein Backbones by RP-HPLC.

(A) Separation of native (glycosylated) AGPs. RP-HPLC profiles of AGPs prepared by precipitation with β -GlcY. Multiple separations were performed for each tissue type, and individual fractions (shaded) with the same retention time were pooled (as indicated by arrows) for subsequent purification.

(B) Separation of deglycosylated AGP protein backbones. RP-HPLC profiles of chemically deglycosylated and N-terminal-deblocked AGP protein backbones from the fractions shaded in **(A)**. The retention time for each peak is shown in Table 1. Arrowheads indicate peaks with the same retention time as the enzyme used in the deblocking step (pyroglutamate aminopeptidase).

The x axis is retention time in minutes. The y axis is absorbance at 215 nm.

backbones (Du et al., 1994). Deglycosylated protein backbones were further separated by RP-HPLC and sequenced by Edman degradation (Figure 1B and Table 1).

The material in fraction 1a (Figure 1B) gave the sequence AOAOTOTATOOOATOOOV (Table 1, where O represents hydroxyproline). This sequence matched the deduced mature protein sequence of AtAGP4 (Schultz et al., 1998). Fraction 1b included two protein backbones, as shown by the two amino acid peaks, one major (1b-1) and one minor (1b-2), at each cycle of sequencing. The respective sequences, AOAOSOTTTVTPOOV and AOGOAOTRSOLPSOA (Table 1), did not match the sequences of AtAGP1-AtAGP5 (Schultz et al., 1998) or AtAGP6 (Sherrier et al., 1999). However, when used to search the sequence databases, the 1b-1 sequence matched a genomic clone that encodes a "classical" AGP protein backbone. This gene was designated *AtAGP7* and has a corresponding EST (ATTS3245). The 1b-2 sequence matched an EST (193B7T7) representing a gene designated *AtAGP10* (Table 1). The N-terminal peptide sequence obtained from fraction 2a was the same as the peptide sequence obtained from fraction 1a (Table 1). This is not

unexpected because the deglycosylated fractions 1a and 2a have very similar retention times (14.2 and 14.5 min, respectively) and the native AGPs in peaks 1 and 2 are poorly resolved (Figure 1A). The N-terminal peptide sequences obtained for all the other fractions were matched to either ESTs or genomic sequences by the strategy outlined above (Table 1).

Some of the Arabidopsis AGP Protein Backbones Are Very Short

In a separate experiment, native AGPs from leaf tissue (retention time 7 min) were deglycosylated, deblocked, and separated by RP-HPLC. A fraction eluting at a retention time of 4.9 min was obtained (data not shown), the N-terminal sequence of which was XXAOAO(S/A)OTS (Table 2), indicating that this fraction contained at least two peptides. In the first two sequencing cycles, no single amino acid was abundant, and the minor peaks were difficult to distinguish from background signals. However, the amino acids present in cycle 1

Table 1. N-Terminal Peptide Sequence of AGPs Isolated from Arabidopsis

Fraction ^a	Time ^b	N-Terminal Sequence ^c	Gene ^d
1a	14.2	A-O-A-O-T-O-T-A-T-O-O-O-A-T-O-O-O-V	<i>AtAGP4</i>
1b-1	17.2	A-O-A-O-S-O-T-T-T-V-T-P-O-O-V	<i>AtAGP7</i>
1b-2	17.2	A-O-G-O-A-O-T-R-S-O-L-P-S-O-A	<i>AtAGP10</i>
2a	14.5	A-O-A-O-T-O-T-A-T-O	Same as 1a
2b-1	16.1	A-O-A-O-S-O-T-T-T-V-T-P	Same as 1b-1
2b-2	16.1	A-O-T-S-O-O-T-A-A-O-A-O ^e	<i>AtAGP9?</i>
2c	17.4	Used for mass spectrometry	
2d	20.5	A-O-A-O-A-O-T-T-V-T-P-O-O-T-A	<i>AtAGP2</i>
3a	16.4	A-O-T-S-O-O-T-A-A-O-A-O ^e	Same as 2b-2
3b	21.3	A-O-A-O-A-O-T-T-V	Same as 2d

^aFractions of deglycosylated and deblocked AGP backbones, as shown in Figure 1B.

^bRetention time in minutes.

^cN-terminal protein sequence, as determined by Edman degradation. Amino acids are represented by the single-letter code, and O represents hydroxyproline.

^dName given to the gene that matches each isolated protein backbone.

^eThe amino acid residue threonine is present instead of the underlined alanine in the protein predicted by the genomic and cDNA sequences of *AtAGP9*.

included L, V, A, S, and T, and those in cycle 2 included L, V, T, and E. In cycles 3 to 6 and cycles 8 to 10, a single amino acid residue was distinguished (Table 2); in the seventh cycle, S was the major amino acid residue, but A was also present. No signal was observed in the 11th and 12th cycles. The peptide sequence APAP(S/A)PTS identified five distinct EST sequences in the databases, which we ordered from the ABRC and sequenced. Each EST was full length and encoded a different protein backbone having the features of a classical AGP. The unusual feature of these AGPs is that the predicted mature protein backbone is only 10 to 13 amino acid residues long (Table 2). Seven of these amino acids are conserved in all five AGP protein backbones, which is why we were able to sequence this fraction even though it contained a mixture of several closely related AGPs.

The Classical AGP Gene Family in Arabidopsis Contains at Least 15 Distinct Genes

Database searching with the seven N-terminal protein sequences (Tables 1 and 2) identified 15 genes that encode classical AGPs. More genes than protein sequences were discovered because the database searches identified sequences that are similar (i.e., contain some mismatches) as well as sequences that are identical. An EST representing each gene was ordered from the ABRC and sequenced. Of the 15 ESTs, 13 were full-length cDNA clones; see Table 3 for GenBank accession numbers of the full sequences. The deduced proteins all had the features of classical AGPs: an N-terminal signal sequence and a region rich in Pro/Hyp, Ala, Ser, and Thr, followed by a C-terminal signal for the addition of a GPI anchor. The two exceptions were *AtAGP7*

and *AtAGP11*. The EST for *AtAGP11* lacks the first two nucleotides of the coding sequence (based on the genomic sequence). The single EST identified for *AtAGP7* is both a partial cDNA clone and a hybrid clone (see Methods for explanation).

Table 2. Peptide Sequencing Identifies AG Peptides

Source	Peptide Backbone Sequence
4.9 min ^a	X-X- A-O-A-O-S-O-T-S ^b
<i>AtAGP12</i>	T-E- A-O-A-O-S-O-T-S ^{c,d}
<i>AtAGP13</i>	V-E- A-O-A-O-S-O-T-S ^c
<i>AtAGP14</i>	V-D- A-O-A-O-S-O-T-S ^c
<i>AtAGP15</i>	S-E- A-O-A-O-S-O-T-S -G-S ^{c,d}
<i>AtAGP16</i>	S-L- A-O-A-O-A-O-T-S ^{c,d,e}

^aPeptide fraction from leaves with a retention time of 4.9 min (after deglycosylation and deblocking).

^bAmino acids are represented by the single-letter code; X indicates very low signal such that no single amino acid residue could be distinguished. In the seventh cycle, Ser was the major amino acid residue and Ala was a minor component.

^cPredicted mature peptide backbone assuming N- and C-terminal signals are removed and all Pro residues are modified to Hyp(O). Residues common to all sequences are in boldface type.

^dThe mature peptide backbone is predicted to begin with a glutamine residue. This residue is not shown because it would be removed from the peptide in the enzymatic deblocking step.

^eThe protein predicted by this clone includes a GPI anchor consensus cleavage site followed by a hydrophobic domain (see Figure 2C), but it is not predicted to be GPI-anchored based on PSORT prediction of cellular localization (Nakai and Horton, 1999). However, purification results suggest that *AtAGP16* is likely to be GPI-anchored (see Discussion).

Table 3. AGP Genes Encode Protein Backbones of Variable Length and with Varying Amino Acid Composition

Gene	Accession Number ^a	Mature Protein ^b	Predicted GPI Signal Cleavage Site	Amino Acid Composition (% w/w) ^c				
				Pro/Hyp	Ala	Ser	Thr	Others (>5% Each)
<i>AtAGP1</i>	AF082298	85	G↓GA	27	19	20	5	None
<i>AtAGP2</i>	AF082299	86	S↓AA	33	17	11	16	V, G
<i>AtAGP3</i>	AF082300	95	D↓SA	36	16	8	13	V
<i>AtAGP4</i>	AF082301	90	S↓AA	40	23	10	14	V
<i>AtAGP5</i>	AF082302	88	N↓AA	33	15	17	11	G, Q
<i>AtAGP6</i>	AJ012459	106	S↓GA	21	23	23	10	D, K
<i>AtAGP7</i>	AF195888	85	N↓AA	37	18	15	14	V
<i>AtAGP8</i>	AF195889	369 ^d	N↓AA	27	21	21	8	None
<i>AtAGP9</i>	AF195890	151	N↓GA	44	14	14	12	V
<i>AtAGP10</i>	AF195891	86	N↓AA	33	15	13	15	G
<i>AtAGP11</i>	AF195892	92	S↓GA	20	22	19	6	D, E, K
<i>AtAGP12</i>	AF195893	11	S↓DA	27	18	18	18	E, Q
<i>AtAGP13</i>	AF195894	10	S↓DA	30	20	20	10	E, V
<i>AtAGP14</i>	AF195895	10	S↓DA	30	20	20	10	D, V
<i>AtAGP15</i>	AF195896	13	S↓SA	23	15	31	8	G, E, Q
<i>AtAGP16</i>	AF195897	11 ^e	S↓DG	27	27	18	9	Q, L

^aThe GenBank accession number given to the complete sequence of the EST (for each gene) obtained from the ABRC. *AtAGP6* was also reported by Sherrier et al. (1999).

^bNumber of amino acid residues in the mature protein backbone. See Methods for predictions of N- and C-terminal signals.

^cAmino acid composition of the predicted mature protein backbones.

^dThis putative AGP has four domains, and the second and fourth domains are AGP-like (Figure 2D). The amino acid composition was determined on the combined AGP-like domains only (77 amino acid residues in total).

^eThe amino acid composition was calculated assuming that *AtAGP16* is GPI-anchored.

Figure 2 shows the complete DNA sequences and deduced protein sequences of four AGP genes. The sequences shown in Figure 2 were chosen because they are most relevant to the Discussion. The mature protein backbone of the classical AGPs (as deduced from DNA sequences) varied from 10 to 151 amino acid residues (Table 3), but in all cases it was rich in Pro/Hyp, Ala, Ser, and Thr. Table 4 shows that most of the AGPs, excluding the AG peptides (*AtAGP12* to *AtAGP16*, with short protein backbones), have <40% amino acid identity to each other.

An AGP-like Molecule May Be a Cell Adhesion Molecule

Another EST (197B15M4) was identified that could have encoded a classical AGP protein backbone; however, it was not a full-length clone. This gene was designated *AtAGP8*. When the full-length genomic sequence of *AtAGP8* was identified (AC005396), the deduced protein was 420 amino acid residues long (Figure 2D). The *AtAGP8* protein differs from the previously identified classical and nonclassical AGPs (Du et al., 1996) by having two AGP-like regions and two β -immunoglobulin (Ig)-H3/fasciadin domains (Kawamoto et al., 1998), and it is predicted to be GPI-anchored. None of the protein backbone sequences reported here match that clone.

AtAGP10 Is GPI-Anchored

A putative GPI anchor signal sequence is present on all of the Arabidopsis classical AGP protein backbones deduced from the genes reported here (Table 3). Electrospray ionization–mass spectrometry (ESI-MS) analysis was used to determine whether the C terminus of a representative AGP protein backbone was post-translationally modified in a manner consistent with the predicted addition of a GPI anchor. The fraction used for ESI-MS was the deglycosylated and deblocked fraction 2c (see Figure 1B, middle panel). This fraction has a retention time similar to that of fraction 1b (17.4 versus 17.2 min, respectively) and therefore may contain one (or both) of the same AGPs as fraction 1b (i.e., *AtAGP7* or *AtAGP10*). The anhydrous hydrogen fluoride treatment used to remove the carbohydrates also removes the glycan and phosphate components of the GPI anchor, leaving an ethanolamine at the C terminus of the protein backbone. ESI-MS showed that a species of average molecular mass 8379 D was present in fraction 2c (Figure 3). This mass is consistent with the mass of the protein predicted to be encoded by the gene *AtAGP10* (8377.1 D for Ala₂₃ to Asn₁₀₇; see Figure 2A), assuming that *AtAGP10* is post-translationally modified as expected: that is, the N-terminal signal sequence is cleaved between Ala₂₁ and Gln₂₂; Gln₂₂ is modified to pyroglutamate and removed in the

enzymatic deblocking step; the GPI anchor signal is cleaved at the predicted cleavage site between Asn₁₀₇ and Ala₁₀₈; an ethanolamine residue is attached to Asn₁₀₇; and 22 of the 26 Pro residues are hydroxylated. The mass obtained by ESI-MS is not consistent with the predicted mass of any of the other Arabidopsis AGPs identified in this study, including AtAGP7 (data not shown).

Differential Expression of Arabidopsis AGPs

AGP gene-specific probes were used to investigate the differential expression of these AGP genes. This experiment was designed to give an indication of expression profiles of the individual genes and was not intended as an exhaustive survey of all tissue types and developmental stages (see Discussion). The genes selected for analysis were based on the initial availability of cDNA clones to make probes. Three replicate RNA gel blots were hybridized with gene-specific probes for *AtAGP3*, *AtAGP4*, and *AtAGP5*, and the blots were exposed for the same length of time (Figure 4A). The blots contained RNA from leaves, roots, flowers, and siliques. Equal amounts of total RNA (10 µg) were loaded per lane. However, the intensity of ethidium bromide staining of rRNAs suggests that each root sample contains approximately half as much RNA as the other samples (Figure 4E). *AtAGP4* was expressed in all tissues and was most abundant in roots and flowers. mRNA corresponding to *AtAGP3* and *AtAGP5* was less abundant and was detected only in roots (*AtAGP3*) or in flowers and siliques (*AtAGP5*). Each of the three blots was stripped and reprobbed three times, first with *AtAGP8*, *AtAGP1*, and *AtAGP2* (Figure 4B), then with *AtAGP9*, *AtAGP12*, and *AtAGP16* (Figure 4C), and finally with *AtAGP7*, *AtAGP14*, and *AtAGP10* (Figure 4D). In the fourth reprobbed, the background was too high to obtain any information for *AtAGP7* and *AtAGP14* (data not shown). *AtAGP1*, *AtAGP8*, *AtAGP9*, *AtAGP10*, and *AtAGP16* were most abundant in flowers; *AtAGP10* was also quite abundant in siliques and roots. Most of the AGP genes were expressed in at least two of the tissues examined, but there were differences in the relative extents of expression as well as in the tissue type in which each gene was expressed.

Table 4. Comparison of Amino Acid Identity between Different AGPs

	AGP1 ^a	AGP2	AGP5	AGP6	AGP13
AGP1	—	—	—	—	—
AGP2	41	—	—	—	—
AGP3	41	67	39	33	55
AGP4	33	48	41	33	60
AGP5	31	36	—	—	—
AGP6	34	42	32	—	—
AGP7	35	49	38	35	70
AGP8	43	32	26	24	55
AGP9	37	42	39	32	55
AGP10	36	49	62	28	50
AGP11	25	36	27	80	36
AGP12	55	64	55	55	90
AGP13	45	45	55	45	—
AGP14	50	60	60	50	90
AGP15	50	54	54	54	82
AGP16	55	78	67	55	70
Na1 ^b	41	42	49	43	64
Pc1 ^c	35	45	40	31	45

^aAmino acid identities are shown as percentages. Comparisons were made using the predicted mature protein backbones (excluding N- and C-terminal signals; see Methods).

^bNa1 is NaAGP1 from *N. alata* (Du et al., 1994).

^cPc1 is PcAGP1 from *P. communis* (Chen et al., 1994).

DISCUSSION

We have taken advantage of the Arabidopsis EST and genomic sequencing efforts (Höfte et al., 1993; Newman et al., 1994; Flanders et al., 1998) to identify genes encoding the protein backbones of classical AGPs. Initially, we identified five putative classical AGP genes, *AtAGP1* to *AtAGP5*, by comparing the structure of the deduced protein backbones with AGPs from other plant species (Schultz et al., 1998). We have now identified a total of 15 AGP genes that encode classical AGPs (Table 3). Because AGPs share little sequence identity (Du et al., 1996; Table 4) and because AGPs had not been purified from Arabidopsis, it was necessary to

Figure 2. (continued).

nal (Udenfriend and Kodukula, 1995b). However, most signals for addition of a GPI anchor do not usually have as many hydrophilic residues following the hydrophobic domain (see Discussion).

(D) Full-length sequence for *AtAGP8* and the deduced amino acid sequence. The genomic sequence is shown (GenBank accession number AC005396; there are no introns). The 3' end of the sequence shown here is the poly(A) attachment site in the longest partial EST (141C1T7) for this gene, which is 100% identical to residues 329 to 1437 of the genomic sequence. The AGP-like domains are in white letters and the β-Ig-H3/fasciclin domains (Kawamoto et al., 1998) are underlined with dots. The amino acid residues in the H1 and H2 subdomains of the β-Ig-H3/fasciclin domain are in italics (and underlined with dots). Potential sites for N-linked glycosylation are boxed. This protein also contains three potential protein kinase C phosphorylation sites and eight potential casein kinase 2 phosphorylation sites (data not shown).

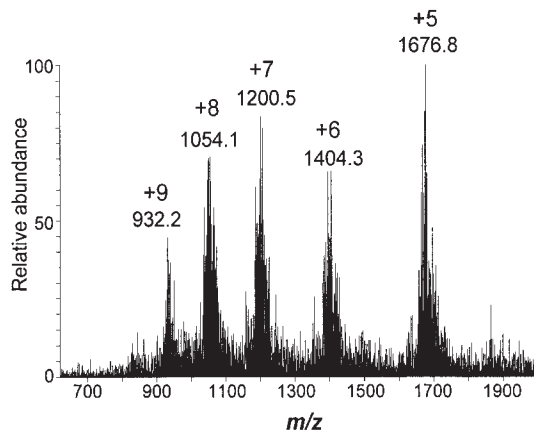


Figure 3. ESI-MS of the Purified AtAGP10 Protein Backbone.

Fraction 2c (Figure 1B) was analyzed by ESI-MS. The molecular mass was determined for each charge state, and the average was 8379 D. Numbers above the ion clusters indicate the charge state of the corresponding ion and the m/z value of the most intense ion.

isolate and sequence AGP protein backbones to confirm that the putative AGP genes encoded bona fide AGPs. To date, we have obtained seven different N-terminal protein sequences from purified AGPs (Tables 1 and 2), and these correspond to the predicted mature protein backbone of seven or eight of the 15 classical AGP genes (see below for

explanation). More genes than protein sequences were discovered because the database searches identified not only sequences that are identical but also those that are only similar (i.e., contain mismatches). We have not identified any full-length clones corresponding to nonclassical AGPs (Du et al., 1996). Only one of the seven protein sequences (fraction 2b-2, Table 1) did not precisely match a gene in the current Arabidopsis database. However, it is very similar (matching 11 of 12 amino acid residues) to the mature protein of the classical AGP deduced from *AtAGP9*. Additional sequencing (of protein backbones and genomic DNA) will be required to determine whether *AtAGP9* encodes the AGP protein backbone found in fraction 2b-2. Of the other six protein sequences, two matched the proteins deduced from the genes *AtAGP2* and *AtAGP4* (Schultz et al., 1998). The other four N-terminal protein sequences matched AGP genes newly identified in the Arabidopsis database. The two peptides inferred from the deglycosylated leaf fraction (retention time 4.9 min; Table 2) matched to five different genes. The observation of several amino acid residues in cycles 1, 2, and 7 of sequencing suggests that at least three of the corresponding AGPs (*AtAGP12*, *AtAGP16*, and *AtAGP13* and/or sometimes *AtAGP14*) were present in the fraction reported in Table 2. This combined proteomic and genomic information provides greater confidence that the sequences designated as putative AGPs on the basis of DNA sequence do actually encode AGP protein backbones.

The mature protein backbones, deduced from the AGP gene sequences, range from ~950 D (for *AtAGP14*) to 16 kD (for *AtAGP9*). However, the majority of the native AGPs iso-

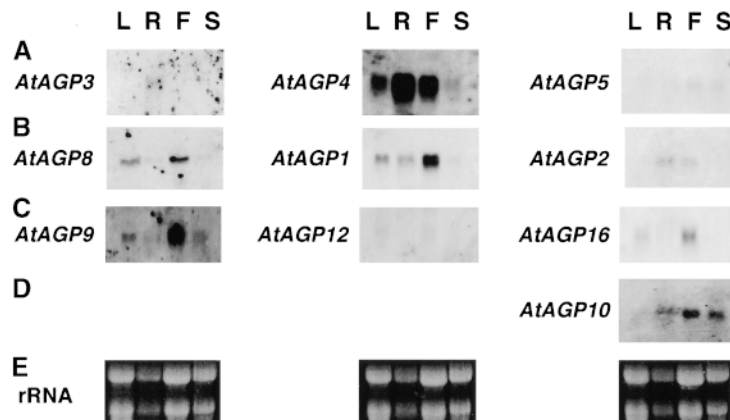


Figure 4. Expression of AGP Genes in Leaves, Roots, Flowers, and Siliques.

(A) Three replicate RNA gel blots were hybridized with gene-specific probes for *AtAGP3*, *AtAGP4*, and *AtAGP5*. Probes are indicated at left. Exposure times were identical for each of the three blots.

(B) Each blot was stripped and reprobbed with the gene-specific probe indicated at left. Exposure times were identical for each of the three blots but differed from those for (A).

(C) and (D) The blots were stripped again and reprobbed as indicated at left. Exposure times were identical for each of the three blots in (C).

(E) Ethidium bromide staining of the rRNAs.

L, leaves; R, roots; F, flowers; S, siliques.

lated from Arabidopsis are >115 kD, as based on their mobility through SDS–polyacrylamide gels (data not shown). The greater apparent molecular mass of the mature AGPs presumably reflects the addition of O-linked sugars to the Hyp residues in the protein backbone. Figure 5 compares the deduced protein backbone of several AGPs with the predicted structure of each native AGP.

AG Peptides: A Subgroup of Classical AGPs

Four of the genes (*AtAGP12* to *AtAGP15*) encode AG peptides with small protein backbones (Table 2). These genes provide proof that plant cells synthesize AG peptides de novo. Fincher et al. (1974) purified an AG peptide from wheat endosperm. However, whether this AG peptide is the result of proteolysis of a larger protein backbone or is synthesized de novo has not been determined. The estimated molecular mass of the wheat AG peptide is 20 kD, with an inferred peptide backbone size of a maximum of 20 amino acid residues (based on the relative proportion of carbohydrate to protein). We are confident that the Arabidopsis AG

peptides are not degradation products of larger AGPs because we have both the protein and the gene sequences.

The AG peptides *AtAGP12* to *AtAGP15* have the following features: they are GPI-anchored, they have short “classical” backbones of between 10 and 13 amino acid residues, and seven of the amino acid residues are conserved (Table 2). A fifth AG peptide, *AtAGP16*, also has a small protein backbone, but it is not clear whether this AGP is GPI-anchored or is attached to the plasma membrane with a transmembrane domain. Throughout this discussion, we consider *AtAGP16* a GPI-anchored AGP because (1) the Ala residue detected in cycle 7 of protein sequencing is found only in this AG peptide and (2) if *AtAGP16* contained a transmembrane domain, then it would be more hydrophobic than the other AG peptides—which is not the case.

The uncertainty surrounding *AtAGP16* highlights the limitations of using computer programs for determining the cellular location of proteins. The PSORT prediction of cellular localization (Nakai and Horton, 1999) suggests that *AtAGP16* has a transmembrane domain with a short cytoplasmic tail (12 amino acids). However, the unprocessed protein backbone of *AtAGP16* appears to include the major features of a

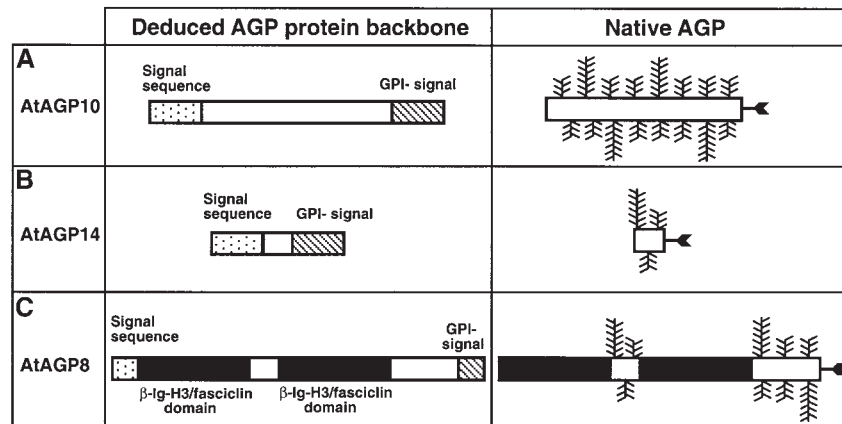


Figure 5. Schematic Representation of the Proteins Deduced from DNA Sequences and the Predicted AGP Structure after Processing and Post-Translational Modifications.

After the N-terminal signal sequence (dots) is removed, the C-terminal GPI anchor signal (diagonal stripes) is recognized, removed, and replaced by a C-terminal GPI anchor (indicated by arrow). Pro residues are hydroxylated to Hyp, and O-linked sugars are added to the Hyp residues.

(A) *AtAGP10* has a predicted protein backbone size of 86 amino acid residues. An estimated 22 of the 26 Pro residues are hydroxylated to Hyp (see Discussion), suggesting that as many as 22 sites could be available for the attachment of O-linked sugars chains (without even considering the possibility of adding O-linked carbohydrate chains to Ser and Thr residues). If carbohydrate chains (indicated by “feathers”) were attached to each of the Hyp residues, then carbohydrates would be attached to 25% of the amino acid residues in the AGP protein backbone. In this diagram, not all of the potential attachment sites are used.

(B) The AG peptide *AtAGP14* has a predicted backbone size of only 10 amino acid residues. N-Terminal sequencing of the protein backbone shows that all three of the Pro residues are hydroxylated to Hyp (Table 2). Thus, O-linked arabinogalactan chains are likely to be added to these Hyp residues such that the relative proportion of attachment sites for carbohydrate chains on the AG peptide is similar to that of other classical AGPs.

(C) *AtAGP8* has two AGP-like regions (open boxes) and two β -Ig-H3/fasciclin domains (black boxes). It is likely that the two AGP-like regions of *AtAGP8* probably are hydroxylated and glycosylated, as predicted for the other AGPs. The four regions of the predicted mature protein backbone of *AtAGP8* are scaled relative to each other. The two AGP-like domains (combined) are approximately the size of the *AtAGP10* protein backbone.

GPI anchor recognition signal, that is, a consensus cleavage site, followed by a hydrophobic domain (Figure 2C). The 12 amino acid residues following the hydrophobic domain (His₆₂-Phe₇₃) are not usually seen in GPI anchor recognition signals (see, e.g., AtAGP10 and AtAGP14, Figures 2A and 2B). However, as Wang et al. (1999) recently showed, a human folate receptor that is usually transmembrane-bound could be GPI-anchored by modifying the hydrophobic domain. GPI anchoring was achieved even though the hydrophobic domain was followed by a cytoplasmic tail of 50 amino acid residues. Thus, the C terminus of AtAGP16 could be a signal for the addition of a GPI anchor, as suggested by our preliminary results. Further experimentation is required to confirm this. If the PSORT prediction is correct, then AtAGP16 represents a new class of AGPs, distinct from the "classical" and "nonclassical" AGPs (Du et al., 1996). In either case, the short, Hyp-rich, glycosylated region should be on the external surface of the cell at the interface between the plasma membrane and the cell wall.

In Arabidopsis, the cellular location of the isolated AGPs is not known because we purified the AGPs in the presence of Triton X-100 to be able to obtain both membrane-bound and secreted AGPs. Based on immunolocalization studies and SDS-PAGE analysis, some AGPs are localized to the plasma membrane in Arabidopsis (Dolan and Roberts, 1995; Dolan et al., 1995; Sherrier et al., 1999). A large proportion of the Arabidopsis AGPs are probably released from the cell surface. In pear, only a small proportion (0.2%) of PcAGP1 is attached to the plasma membrane by way of its GPI anchor (Oxley and Bacic, 1999). Chemical analysis of PcAGP1 suggests that it is actively released into the extracellular matrix by either phospholipase D or phospholipase C and a phosphatase working sequentially (Oxley and Bacic, 1999). Whether the release of GPI-anchored AGPs from the plasma membrane is tightly regulated or is linked to the function of the AGPs is unknown.

Are AGPs Involved in Signaling?

The precise role that GPI anchors play in the function of AGPs remains unknown. Several GPI-anchored proteins from animals are implicated in signal transduction pathways (Peles et al., 1997; Kleeff et al., 1998; Resta et al., 1998). In these examples, signal transduction occurs through interactions with other membrane-bound proteins. These interactions can be with proteins in the same cell as the GPI-anchored protein or in neighboring cells (Peles et al., 1997). If GPI-anchored AGPs are involved in signal transduction pathways by way of interactions with other molecules, then these other molecules probably have both intra- and extracellular domains. An increasing number of plant proteins with these features are being identified, for example, wall-associated kinase, somatic embryogenesis receptor kinase, and *Clavata1* (reviewed in Lease et al., 1998).

Another possible mechanism by which GPI-anchored pro-

teins may be involved in signaling is by the phospholipase-mediated cleavage of the protein from its lipid anchor (Udenfriend and Kodukula, 1995a). This process has the potential to generate both intra- and extracellular messengers by way of the lipid anchor or extracellular proteoglycan components, respectively. If the Arabidopsis AG peptides (AtAGP12 to AtAGP16) are released from the cell surface and they are similar in size to the wheat AG peptide (molecular mass ~20 kD; Fincher et al., 1974), then they have the potential to act as signals that can diffuse easily through plant cell walls. Experimental evidence suggests that living plant cells have pores that allow globular molecules as large as 40 kD to pass through (Carpita et al., 1979). Although larger AGPs have the ability to move through the walls of cells specialized for secretion (e.g., roots and stigmas), this may not be the case for all plant cells. Kreuger and van Holst (1996) have suggested that large AGPs could be cleaved (by proteases or glycosidases) to enable them to act as diffusible signaling molecules. The evidence for the *de novo* synthesis of AG peptides provides a mechanism for AGPs to act directly as signaling molecules that can move from cell to cell. Proof that AGPs are indeed involved in cell signaling requires further experimentation.

An AGP-like Molecule May Be Involved in Cell Adhesion

AtAGP8 encodes an AGP-like protein backbone but is different from the previously identified classical and nonclassical AGPs (Du et al., 1996). *AtAGP8* includes an N-terminal signal sequence and a C-terminal signal for adding a GPI anchor; however, the mature protein backbone can be separated into four regions (Figures 2D and 5C). Two of these are AGP-like regions, in that they are rich in Pro/Hyp, Ala, Ser, and Thr (shaded residues, Figure 2D), but *AtAGP8* also contains two β -Ig-H3/fasciclin domains (sequence underlined with dots, Figure 2D). β -Ig-H3/fasciclin domains are found in proteins from animals, insects, algae, and bacteria and are thought to be involved in cell adhesion (Kawamoto et al., 1998). At least five other Arabidopsis genes encode proteins with both Pro/Hyp-rich and β -Ig-H3/fasciclin domains (data not shown). β -Ig-H3/fasciclin domains consist of two long repeats, and in most cases these long repeats include two highly conserved domains, H1 and H2 (Kawamoto et al., 1998). In *AtAGP8*, the first β -Ig-H3/fasciclin-like domain contains an H1 domain and the second β -Ig-H3/fasciclin domain contains both an H1 and an H2 domain (italicized residues, Figure 2D). Figure 6 shows an alignment of H1 domains from proteins found in plants, *Volvox*, mice, and *Drosophila*. The AGP-like regions of *AtAGP8* are most likely hydroxylated and glycosylated (Figure 5C), but amino acid and carbohydrate analyses of purified *AtAGP8* will be necessary to confirm this.

Proteins with β -Ig-H3/fasciclin domains appear to have special roles in development. The cell adhesion molecule algal-CAM from *Volvox* is important in embryo development

AtAGP8	T	V	F	A	P	S	D	E	A	F
PtX14A9	T	I	F	A	P	S	D	A	A	F
Algal-CAM	T	I	F	V	P	T	D	E	A	F
Mm β -Ig-H3.1	T	V	F	A	P	T	N	E	A	F
Mm β -Ig-H3.2	T	I	F	A	P	S	N	E	A	F
Dm Fas1.1	T	I	F	V	P	T	N	E	A	F
Dm Fas1.2	T	I	L	A	P	S	N	E	A	F

Figure 6. Alignment Showing the Conserved H1 Sequence of Proteins Containing β -Ig-H3/Fasciclin Domains.

AtAGP8 is described in this study. PtX14A9 is the deduced protein of a putative xylem-specific AGP gene from *Pinus taeda* (GenBank accession number U09556; Loopstra and Sederoff, 1995). Algal-CAM is a cellular adhesion molecule from *Volvox carteri* (X80416; Huber and Sumper, 1994). Mm β -Ig-H3.1 and Mm β -Ig-H3.2 are two of the four H1 domains present in a β -Ig-H3 protein from mouse that inhibits cell attachment in vitro (L19932; Skonier et al., 1994). Dm Fas1.1 and Dm Fas1.2 are two of the four H1 domains from the neural cell adhesion molecule, fasciclin 1, of *Drosophila* (M20545; Zinn et al., 1988). Identical amino acid residues are listed in white letters, conserved amino acids are shaded with dark gray boxes, and similar amino acids are stippled.

(Huber and Sumper, 1994) and the fasciclins of *Drosophila* are involved in neuronal development (reviewed in Prokop, 1999). Fas1, a GPI-anchored form of the glycoprotein fasciclin, functions in a signaling pathway with a specific cytoplasmic tyrosine kinase, Abelson (Abl), even though there is no direct physical interaction between Fas1 and Abl. This relationship was determined on the basis of the phenotype of the *fas1/abl* double mutant, which has major defects in the pathways of axons (Elkins et al., 1990).

Arabidopsis Has a Large Gene Family Encoding GPI-Anchored AGPs

For most of the Arabidopsis classical AGPs, it was possible to choose a single cleavage site (ω) based on probabilities of amino acids at the ω , $\omega + 1$, and $\omega + 2$ positions in other eukaryotic systems (Udenfriend and Kodukula, 1995b). One of these sites, Asn \downarrow Ala-Ala, which is present in AtAGP7, AtAGP8, and AtAGP10, is identical to that determined experimentally for NaAGP1 from *N. alata* (Youl et al., 1998). In all, eight putative cleavage recognition sites were identified (see Table 3); thus, AGPs from Arabidopsis should provide a good system for studying attachment and processing of GPI anchors in plants.

AGPs Do Not Contain a Signature Motif

We were unable to find a motif that identifies a protein as an AGP backbone. Generally, the AGP protein backbones have

Pro residues alternating with Ala, Ser, Thr, or Val rather than sequences of three or more Pro residues, as is commonly found in extensins (Kieliszewski and Lamport, 1994). The only predicted Arabidopsis AGP with an extensin-like motif was AtAGP9, which had five Ser-Pro₃ motifs. AtAGP9 is not likely to be an extensin for the following reasons: it is precipitated by β -GlcY; it has motifs that are AGP-like, such as five Thr-Pro₃ motifs and two Ala-Pro₃ motifs; and it has no Ser-Pro₄ motifs or Tyr residues, which are commonly found in extensins. None of the predicted AGPs has the motifs typically associated with the Pro/Hyp-rich glycoproteins: Pro-Pro-Xaa-Yaa-Lys or Pro-Pro-Xaa-Lys (Sommer-Knudsen et al., 1998).

However, some motifs were found in several AGPs. One motif, PAPAP, was found in AtAGP1 to AtAGP3, AtAGP8, AtAGP9, and AtAGP16. Another motif, ATPPPP, occurs five times in AtAGP4, three times in AtAGP9, and once in AtAGP7. This motif is also found in AGPs from cotton and pine (John and Keller, 1995; Loopstra and Sederoff, 1995). Both PAPAP and ATPPPP were present in the N terminus of several AGPs and were sequenced at the protein level (Table 1, fractions 2d and 1a, respectively). In these examples, all of the Pro residues are hydroxylated to Hyp. Of the seven AGP protein backbones sequenced, most of the Pro residues were hydroxylated, with two exceptions: TTVTPOO in AtAGP2 and AtAGP7 and SOLPSO in AtAGP10. In extensins and other Hyp-rich glycoproteins, KP, YP, and FP are not hydroxylated, whereas PV, SPPPP, AP, and PA are always hydroxylated (Kieliszewski and Lamport, 1994; Sommer-Knudsen et al., 1998). There is no simple rule for "TP," given that one Pro residue is not hydroxylated in TTVTPOO, whereas the Pro residues in ATPPPP and PTP are hydroxylated (see Table 1). The sequence LP rather than LO, which is present in AtAGP10, is also found in the protein backbones of both NaAGP1 (Du et al., 1994) and the 120-kD glycoprotein (Schultz et al., 1997). However, more protein sequencing will be required to determine whether LP is ever hydroxylated.

AGPs Are Abundant

The RNA gel blot experiments show that most of the Arabidopsis AGP genes are expressed in at least two of the tissue types examined, but there are differences in the relative extent of expression for each gene. This experiment was designed to give a preliminary indication of the expression profiles of the individual genes and was not intended as an exhaustive survey of all tissue types and developmental stages. One reason we did not attempt an exhaustive analysis of AGP expression by way of RNA gel blot analysis is that mRNA and protein expression are not always correlated. Human liver cells, for example, show only a 0.48 correlation factor between mRNA and protein concentrations (Anderson and Seilhamer, 1997). The other reason is the advent of such techniques as microarray technology (Schen

et al., 1995). Microarray analysis will be important for studying AGP gene expression because it allows use of sensitive oligonucleotide-based microarrays with perhaps 4 to 10 unique oligonucleotides per gene to ensure gene specificity (Kehoe et al., 1999). RNA in situ hybridization will also be an important tool, especially for determining cell type specificity of each AGP gene.

In most cases, the AGPs purified from *Arabidopsis* tissues were well represented in the EST database with from five to >10 ESTs. There were two notable exceptions: AtAGP7, for which only one partial (hybrid) EST (ATTS3245) was identified, and AtAGP10, for which only two ESTs were identified. We were surprised that only two ESTs for AtAGP10 were found because the transcript is relatively abundant in roots and flowers (Figure 4D) and because mRNA from both root and flower tissues was used in the construction of the cDNA library, λ PRL2, from which most of the ESTs were obtained (Newman et al., 1994).

All of the N-terminal peptide sequences from fraction 2, which contained AGPs from both leaf and root tissues (Figure 1A), were also found in fractions 1 and 3, which were from root tissue only. This result highlights the difficulties in separating native AGPs. None of the leaf-only fractions from the RP-HPLC separation shown in Figure 1A contained enough material for N-terminal sequencing (data not shown). Therefore, most, if not all, of the N-terminal peptide sequences in fraction 2 probably came from the root sample only. For all of the AGPs purified from roots only—AtAGP2, AtAGP4, AtAGP7, AtAGP9, and AtAGP10—the appropriate gene-specific transcript was detected by RNA gel blot analysis, although the transcript was not necessarily abundant (see, e.g., AtAGP2, Figure 4). On the basis of RNA gel blot analysis, we expect that AtAGP4 and AtAGP9 will also be relatively abundant in leaves.

Why So Many AGP Genes?

The observation that many of the *Arabidopsis* AGP genes are expressed in two or more tissues rather than being restricted to a single tissue type is consistent with other multigene families (reviewed in Meagher et al., 1999). Often this is attributed to genetic redundancy. However, increasing evidence indicates that multigene families allow organisms to modulate their response to the environment to greater effect (Pickett and Meeks-Wagner, 1995; McAdams and Arkin, 1999; Meagher et al., 1999). Indeed, multigene families may have evolved because networked biological systems are more robust than systems with a single highway (Barkai and Leibler, 1997). To achieve the benefits of this phenomenon, termed isovariant dynamics, the coexpressed members of the gene family must have at least one different activity (e.g., interactions with other proteins or cofactors) (Meagher et al., 1999).

AGPs have the potential for specific interactions with other molecules, including other AGPs, because of differences in both protein and carbohydrate composition. Deter-

mining the molecule or molecules with which each AGP interacts and identifying the structural features that are important for these interactions will be exciting and challenging areas of AGP research. Progress in these areas has the potential to proceed rapidly by using a reverse genetics approach in *Arabidopsis*.

METHODS

Plant Material

Wild-type *Arabidopsis thaliana* (Columbia-0 strain CS1092; Arabidopsis Biological Research Center [ABRC], Columbus, OH) plants were used. For isolation of arabinogalactan proteins (AGPs), most of the plants were grown in liquid medium (Reiter et al., 1992) except where noted. Gamborg's B-5 medium (Gibco BRL; cat. No. 21153-028) was prepared in 100-mL volumes and placed in 250-mL flasks. Seeds were surface-sterilized and washed three times with ultrafiltered water. The seeds (20 to 30) were pipetted into the flasks, grown on a shaker table (shaking at 120 rpm) at 26°C on a 16-hr-light/8-hr-dark cycle, and then harvested after 12 to 14 days. Roots and leaves were separated, frozen in liquid nitrogen, and freeze-dried. All plants to be used for RNA extraction were grown in soil in a greenhouse. The root and leaf tissues to be extracted were harvested at 4 to 6 weeks (just before bolting). Flower samples included closed buds and fully opened flowers.

Purification of AGPs

To extract AGPs, 8 g (fresh weight) of freeze-dried root or leaf tissue was ground to a fine powder in liquid nitrogen. Roots were harvested from both soil-grown plants and plants grown in liquid culture (at a ratio of 1:4 (w/w), respectively) and combined. Ground tissue was added to 8 mL of extraction buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 0.1% β -mercaptoethanol, and 1% (w/v) Triton X-100) and incubated at 4°C for 3 hr. Samples were centrifuged for 10 min at 14,000g. The supernatant was precipitated with 5 volumes of ethanol (at 4°C, overnight). The pellet was resuspended by vortex-mixing in 5 mL of 50 mM Tris-HCl, pH 8.0. The insoluble material was removed by centrifugation, and the supernatant was retained. The pellet was resuspended in an additional 5 mL of 50 mM Tris-HCl, pH 8.0, and the soluble material (after centrifugation) was pooled with the first supernatant. The buffer-soluble material was freeze-dried overnight to concentrate the sample. The dried samples were resuspended in 250 to 500 μ L of 1% (w/v) NaCl and transferred to 2-mL microcentrifuge tubes. AGPs were precipitated with the β -glucosyl Yariv reagent (β -GlcY) (Gane et al., 1995) by mixing the resuspended samples in an equal volume of β -GlcY (2 mg mL⁻¹) in 1% (w/v) NaCl and leaving overnight at 4°C. The insoluble Yariv-AGP complex was collected by centrifugation at 14,000g in a microcentrifuge for 1 hr. The β -GlcY was removed by washing the pellet three times in 1% (w/v) NaCl and then twice in methanol. The pellet was dried, dissolved in a minimum volume of dimethyl sulfoxide, and mixed with solid sodium dithionite. Water was added with vortex-mixing until the mixture became a clear yellow color. The resulting yellow solution was then desalted on a PD-10 column (Pharmacia) that had been equilibrated with water, and the eluate was freeze-dried.

HPLC

Reversed-phase (RP)–HPLC was performed with a Brownlee Aquapore RP-300 column (C_8 , 2.1×100 mm; Applied Biosystems, Foster City, CA) attached to a Beckman System Gold HPLC (Beckman Instruments, Brea, CA), which consisted of a model 126 solvent delivery system and a model 168 diode array detector. Samples (1 mL) were loaded onto the RP-300 column, which was equilibrated with 0.1% (v/v) trifluoroacetic acid (TFA). AGPs (native and deglycosylated) were eluted and collected from the column with a linear gradient of solvent B (80% acetonitrile in 0.1% TFA): from 0 to 30% solvent B in 30 min and then from 30 to 100% in 30 min at a flow rate of 1 mL min^{-1} . Chromatography was monitored by absorption at 215 and 280 nm. In the experiment that identified the AG peptides, the native AGPs were separated with a different linear gradient of solvent B: from 0 to 100% solvent B in 60 min.

Deglycosylation of AGPs by Anhydrous Hydrogen Fluoride

AGPs were deglycosylated with anhydrous hydrogen fluoride (HF), according to the method of Mort and Lampion (1977). RP-HPLC fractions containing native AGPs were dried in a rotary evaporator overnight. AGPs were resuspended in anhydrous methanol in a closed system that had been evacuated before the anhydrous HF was added. Two sets of conditions were used for different sets of deglycosylation: in the first experiment (Figure 1), samples were incubated for 2 hr at room temperature and the HF was removed under vacuum; the pellet obtained was resuspended in 0.1% TFA, desalted on a PD-10 column equilibrated in 0.1% TFA, and freeze-dried. In the second experiment (Table 2), samples were incubated 4 hr on ice, the HF was removed under vacuum, and the pellet was resuspended in 2 M Tris-HCl, pH 7.5, in 8 M guanidine hydrochloride. Samples were desalted on Sephadex G-10 columns (Pharmacia) equilibrated in 0.1% TFA and freeze-dried.

N-Terminal Deblocking by Pyroglutamate Aminopeptidase

A deblocking step was included in the purification protocol as a precautionary measure. Glutamine residues are frequently cyclized into pyroglutamate, which cannot be cleaved during Edman degradation sequencing (Mozdzanowski et al., 1998). Of the seven AGP protein backbones sequenced at the protein level, eight were predicted to start with a glutamine residue, so this step was probably critical. Removal of pyroglutamate was performed at 37°C for 12 hr with the enzyme pyroglutamate aminopeptidase (Boehringer Mannheim) using 20 μg of enzyme per nmol of peptide, as described by Du et al. (1994).

Protein Sequencing

N-Terminal protein sequencing was performed by automated Edman degradation sequencing on a sequencer (model LF 3400; Beckman Instruments) with on-line analysis on a Beckman System Gold HPLC.

Electrospray Ionization–Mass Spectrometry

Mass spectrometry was performed with a Finnigan MAT LCQ ion trap mass spectrometer (ThermoQuest, San Diego, CA). Electrospray conditions were as follows: heated capillary = 200°C ; tube lens voltage = 40 V; capillary voltage = 30 V; sheath gas of nitrogen at

290 kPa; needle voltage = 4.3 kV. Mass spectra were acquired approximately every second with a maximum ion time of 500 μsec and an average of 2 microscans per spectrum. The sample was preconcentrated just before mass spectrometry by the HPLC attached directly to the electrospray source. The samples were eluted from a reversed-phase C_8 , $5 \text{ cm} \times 300\text{-}\mu\text{m}$ column (LC Packings, San Francisco, CA) with a gradient from 0.5% acetic acid in water to 0.5% acetic acid and 20% water in acetonitrile. The Applied Biosystems 140B solvent delivery system used a pre-split to deliver a solvent flow through the column of $\sim 6 \mu\text{L/min}$. The total eluate from the reversed-phase column passed directly into the electrospray mass spectrometer.

AGP Gene Identification and Sequencing

Peptide sequences obtained by Edman degradation sequencing were used to search DNA sequence databases using either tFastA (Pearson and Lipman, 1988) or tBlastn (Altschul et al., 1990) search algorithms. In some cases, it was necessary to alter the parameters (e.g., turning off filtering tools or reducing the word size). Bacterial strains containing the expressed sequence tag (EST) of interest were ordered from the ABRC through the Arabidopsis Information Management System (AIMS) website (<http://aims.cps.msu.edu/aims/>). Plasmid DNA was prepared by using the alkaline lysis/polyethylene glycol precipitation protocol, modified for use in dye terminator cycle sequencing, according to the manufacturer's (Applied Biosystems) instructions. Cycle sequencing was performed with an ABI Prism dye terminator cycle sequencing mix (Applied Biosystems), according to the manufacturer's instructions, by using primers specific to the vectors (i.e., SP6, T7, or T3). The GenBank accession number for each full-length (or longest) cDNA clone (sequencing performed in our laboratory) is given in Table 3. Except for the single intron in *AtAGP9* and *AtAGP16*, there was 100% identity between the EST and the genomic sequences.

For each gene, the EST and genomic accession numbers are reported as follows: gene name, EST number, and GenBank accession number of genomic clone if available (with bacterial artificial chromosome number and position of coding sequence [CDS] in parentheses). *AtAGP1*, ATTS0200, AB008268 (MSJ1, CDS 46,802..47,197); *AtAGP2*, 157F11T7, AC006592 (F14M13, CDS 55,491..55,885); *AtAGP3*, 171M2T7; *AtAGP4*, 147O10T7; *AtAGP5*, 190F4T7; *AtAGP6*, 171N22T7; *AtAGP7*, ATTS3245 (partial/hybrid EST; 1..270 matches the 3' end of the *AtAGP7* gene [71 bp of coding sequence and 199 bp of 3' untranslated region], 275 to 1254 matches a genomic clone on chromosome 4 [AL078637 96,905..98,903, minus introns], *AtAGP7* is on chromosome 5), AB011479 (MNA5, CDS 41,174..40,781); *AtAGP8*, 141C1T7 (partial EST), AC005396 (F4L23, CDS 3879..5141); *AtAGP9*, 172L5T7, AC005396 (T26I20, CDS 20,171..20,680, 21,182..21,246); *AtAGP10*, 193B7T7, AC005359 (F23J3, CDS 44,591..44,974); *AtAGP11*, 305B10T7 (partial EST), AC009325 (F4P13, CDS 25,127..24,716); *AtAGP12*, 210H14T7, AP000603 (MRP15, 63,017..63,199); *AtAGP13*, 126P21T7, AL049171 (T25K17, CDS 61,524..61,345); *AtAGP14*, 176B14T7, AB019234 (MKN22, CDS 18,677..18,859); *AtAGP15*, 244F14T7; and *AtAGP16*, 130D3T7, AC005397 (F11C10, CDS 2814..2697, 2542..2435).

Computer Analysis of Sequences

Most of the sequence analysis programs used were provided by ANGIS (Australian National Genomic Information Service, Sydney, Australia), including the Genetics Computer Group suite of programs (e.g.,

FINDPATTERNS, GAP), the Staden programs (e.g., protein interpretation program [PIP]), and multiple sequence alignments programs. All analyses were performed by using the default parameters. PIP was used to determine the molecular mass of the unmodified protein backbone of AtAGP10 (Ala₂₃ to Asn₁₀₇). The predicted mass for AtAGP10 after modification (8377.1 D) was determined by adding +16 D for each Pro residue modified to Hyp; 43.3 D was added for the C-terminal ethanolamine residue. Multiple sequence alignments obtained with ClustalW (Thompson et al., 1994) were uninformative because of the large number of gaps and the low sequence identity between the AGPs (data not shown). Therefore, to show the similarity between the AGPs, pairwise comparisons were performed by using GAP (Table 4). GAP uses the algorithm of Needleman and Wunsch (1970) to find the alignment of two "complete" sequences that maximizes the number of matches and minimizes the number of gaps. A penalty (of 3) is assigned for introducing a gap and a smaller penalty (0.1) is added for each extension of the gap. When one sequence is longer than the other, the best alignment of the shorter sequence with the longer sequence is found (with the appropriate penalties for introducing a gap). All gaps and the "unaligned" sequence at the beginning and end of the longer sequence are ignored when calculating the percentage identity. This explains why the AG peptides, for example, AtAGP13, have relatively high similarity to most of the AGPs (Table 4). Although this analysis has limitations, it has the advantage of maximizing the length of sequence compared by using all of the smallest sequence. To predict the cleavage sites of N-terminal signal sequences, we used SignalP (<http://www.cbs.dtu.dk/services/SignalP/index.html>; Nielsen et al., 1997). Prediction of GPI anchor addition was performed with PSORT (<http://psort.nibb.ac.jp:8800/>; Nakai and Horton, 1999). Prediction of the cleavage site (ω) for GPI anchor addition was determined manually based on the probabilities of possible amino acid at positions ω , $\omega + 1$, and $\omega + 2$ (Udenfriend and Kodukula, 1995b). To determine the presence of motifs, FINDPATTERNS was used. All AGPs were checked by FINDPATTERNS, with use of the motifs specified in Discussion. The profilescan server http://www.isrec.isb-sib.ch/software/PFSCAN_form.html (Hofmann et al., 1999) was used to identify the β -immunoglobulin (Ig)-H3/fasciclin domain. We also used FINDPATTERNS to search for the H1 and H2 subdomains of β -Ig-H3/fasciclin (Kawamoto et al., 1998), allowing for three mismatches of the consensus sequences (for H1, TV/LF/LA/VPT/SN/DXF/W; for H2, NGVI/HHXI/VDXV/LL/I, where X = any amino acid residue).

RNA Gel Blot Analysis

RNA was isolated by a protocol described by McClure et al. (1990). Total RNA (10 μ g) was electrophoresed through 1.2% agarose gels containing formaldehyde and transferred to nylon membrane, as previously described (Schultz et al., 1997). Single-stranded digoxigenin-labeled probes were prepared by a two-stage polymerase chain reaction (PCR) protocol (Myerson, 1991), as described by Schultz et al. (1997). For each gene, the first (nonlabeling) PCR round included the appropriate forward and reverse primers (see below). In the second (labeling) round of PCR, only the reverse primer was used. Primers were synthesized by Life Technologies (Gaithersburg, MD) and were of standard purity. The 5' and 3' primers were designated as forward (F) and reverse (R) primers, respectively. The sequence of each primer is as follows: AGP1-F1, 5'-GTGTTTGTCTTCTCGCTGCTC3'; AGP1-R1, 5'-AATGAATCATCATCTCTCTCAC-3'; AGP2-F1, 5'-TTCTAAGCAATGCAAGCTTTG-3'; AGP2-R1, 5'-TGCTCTATGTT-CATCTCATCC-3'; AGP3-F1, 5'-TCAGGTTTCTATCTCTCTCGTC-3';

AGP3-R1, 5'-TACAATCAGAAGCTTCTTCCCTC-3'; AGP4-F1, 5'-CCAAAGAGAAAGAGAGAGAAATG-3'; AGP4-R1, 5'-CCTCTACAC-AACCATATGAAGC-3'; AGP5-F1, 5'-CGTAACAATGGCCTCCAA-ATCC-3'; AGP5-R1, 5'-GTGAATCTATTCGATGGGTC-3'; AGP8-F2, 5'-AATTCGATCTAACGACCTCTACC-3'; AGP8-R1, 5'-CAAACCTCAA-CAACACATAACCAC-3'; AGP9-F1, 5'-CTTTTCGCTATTGCTGTGA-TCTG-3'; AGP9-R1, 5'-CCTGCTATCTCCATCTCAAGCTC-3'; AGP10-F1, 5'-GTCGTTTTGCTCTTCTCGCTC-3'; AGP10-R1, 5'-GAC-GAATACAAATCCGGCTAAAG-3'; AGP12-F1, 5'-GTGCAAAAAG-AGGAGAAATGGAG-3'; AGP12-R1, 5'-CACACAACACATAGTAGT-CC-3'; AGP16-F1, 5'-TGGCGTCGAGAACTCCGTCAC-3'; and AGP16-R1, 5'-CTCCAGAAATCATAATCGAG-3'. Probe sizes amplified were *AtAGP1*, 469 nucleotides (nt); *AtAGP2*, 557 nt; *AtAGP3*, 695 nt; *AtAGP4*, 670 nt; *AtAGP5*, 453 nt; *AtAGP8*, 572 nt; *AtAGP9*, 730 nt; *AtAGP10*, 368 nt; *AtAGP12*, 341 nt; and *AtAGP16*, 383 nt. We are confident that the probes used are gene specific, as based on the different expression profiles of the transcripts detected by the two most closely related probes (Figure 4) used in this analysis, *AtAGP2* and *AtAGP3* (65% identity at the DNA level over the region of the probes).

Hybridization was performed at 42°C in a solution of 50% formamide, 5 \times SSPE (1 \times SSPE is 0.15 M NaCl, 10 mM sodium phosphate, and 1 mM EDTA), 7% SDS, 0.1% *N*-lauroyl sarcosine, 2% block (Boehringer Mannheim), and 20 mM sodium maleate, pH 7.5, with the digoxigenin probe at 10 ng/mL of hybridization solution. Blots were washed twice for 5 min at room temperature in 2 \times SSC (1 \times SSC is 0.15 M NaCl and 0.015 sodium citrate) and 0.1% SDS, followed by two 15-min washes at 65°C in 0.5 \times SSC and 0.1% SDS. Chemiluminescent detection of bound probe was with CSPD (3-(4-methoxyphosphoryl)-1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3,3,1.1^{3,7}]decan-4-yl) phenyl phosphate (Boehringer Mannheim), according to the manufacturer's instructions. Blots were stripped by washing the blots once (first two "strippings") or twice (third "stripping") for 30 min at 65°C in 50% formamide, 1% SDS, and 50 mM Tris-HCl, pH 8.0. After stripping, blots were rinsed in water and then in 2 \times SSC, after which they were either reprobed immediately or stored at 4°C in 2 \times SSC.

ACKNOWLEDGMENTS

This work was done jointly with funding provided by the Australian Government to the Cooperative Research Centre for Bioproducts and the Plant Cell Biology Research Centre, a Special Research Centre of the Australian Research Council. We are grateful to the ABRC at Ohio State University for providing ESTs and seeds. We are grateful to Dr. Shaio-Lim Mau for her extensive advice on the purification of AGPs. We thank Kris Ferguson for protein sequencing and general technical assistance. We thank Professor Adrienne Clarke for her support, encouragement, and ongoing input into these studies. We also thank Drs. Paul Gilson and Ed Newbigin for comments provided during the preparation of the manuscript.

Received April 7, 2000; accepted June 2, 2000.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

- Anderson, L., and Seilhamer, J.** (1997). A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537.
- Azpiroz-Leehan, R., and Feldman, K.A.** (1997). T-DNA insertion mutagenesis in Arabidopsis: Going back and forth. *Trends Genet.* **13**, 152–156.
- Barkai, N., and Leibler, S.** (1997). Robustness in simple biochemical networks. *Nature* **387**, 913–917.
- Baskin, T.I., Betzner, A.S., Hoggart, R., Cork, A., and Williamson, R.E.** (1992). Root morphology mutants in *Arabidopsis thaliana*. *Aust. J. Plant Physiol.* **19**, 427–437.
- Campisi, L., Yang, Y., Yi, Y., Heilig, E., Herman, B., Cassista, A.J., Allen, D.W., Xiang, H., and Jack, T.** (1999). Generation of enhancer trap lines in Arabidopsis and characterization of expression patterns in the inflorescence. *Plant J.* **17**, 699–707.
- Carpita, N., Sabulase, D., Montezinos, D., and Delmar, D.P.** (1979). Determination of the pore size of cell walls of living plant cells. *Science* **205**, 1144–1147.
- Casero, P.J., Casimiro, I., and Knox, J.P.** (1998). Occurrence of cell surface arabinogalactan-protein and extensin epitopes in relation to pericycle and vascular tissue development in the root apex of four species. *Planta* **204**, 252–259.
- Chen, C.-G., Pu, Z.-Y., Moritz, R.L., Simpson, R.J., Bacic, A., Clarke, A.E., and Mau, S.-L.** (1994). Molecular cloning of a gene encoding an arabinogalactan-protein from pear (*Pyrus communis*) cell suspension culture. *Proc. Natl. Acad. Sci. USA* **91**, 10305–10309.
- Ding, L., and Zhu, J.K.** (1997). A role for arabinogalactan-proteins in root epidermal cell expansion. *Planta* **203**, 289–294.
- Dolan, L., and Roberts, K.** (1995). Secondary thickening in roots of *Arabidopsis thaliana*: Anatomy and cell surface changes. *New Phytol.* **131**, 121–128.
- Dolan, L., Linstead, P., and Roberts, K.** (1995). An AGP epitope distinguishes a central metaxylem initial from other vascular initials in the Arabidopsis root. *Protoplasma* **189**, 149–155.
- Du, H., Simpson, R.J., Moritz, R.L., Clarke, A.E., and Bacic, A.** (1994). Isolation of the protein backbone of an arabinogalactan-protein from the styles of *Nicotiana glauca* and characterization of a corresponding cDNA. *Plant Cell* **6**, 1643–1653.
- Du, H., Clarke, A.E., and Bacic, A.** (1996). Arabinogalactan-proteins—A class of extracellular matrix proteoglycans involved in plant growth and development. *Trends Cell Biol.* **6**, 411–414.
- Elkins, T., Zinn, K., McAllister, L., Hoffmann, F.M., and Goodman, C.S.** (1990). Genetic analysis of a Drosophila neural cell adhesion molecule: Interaction of fasciclin 1 and Abelson tyrosine kinase mutations. *Cell* **60**, 565–575.
- Fincher, G.B., Sawyer, W.H., and Stone, B.A.** (1974). Chemical and physical properties of an arabinogalactan-peptide from wheat endosperm. *Biochem. J.* **139**, 535–545.
- Flanders, D.J., Weng, S., Petel, F.X., and Cherry, J.M.** (1998). AtDB, the *Arabidopsis thaliana* database, and graphical-web-display of progress by the Arabidopsis Genome Initiative. *Nucleic Acids Res.* **26**, 80–84.
- Gane, A.M., Craik, D., Munro, S.L.A., Howlett, G.J., Clarke, A.E., and Bacic, A.** (1995). Structural analysis of the carbohydrate moiety of arabinogalactan-proteins from stigmas and styles of *Nicotiana glauca*. *Carbohydr. Res.* **277**, 67–85.
- Gao, M., Kieliszewski, M.J., Lampert, D.T.A., and Showalter, A.M.** (1999). Isolation, characterization and immunolocalization of a novel, modular tomato arabinogalactan-protein corresponding to the LeAGP-1 gene. *Plant J.* **18**, 43–55.
- Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A.** (1999). The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**, 215–219.
- Höfte, H., et al.** (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* **4**, 1051–1061.
- Hooper, N.M.** (1997). Glycosyl-phosphatidylinositol anchored membrane enzymes. *Clin. Chim. Acta* **266**, 3–12.
- Huber, O., and Sumper, M.** (1994). Algal-CAMs: Isoforms of a cell adhesion molecule in embryos of the alga *Volvox* with homology to *Drosophila* fasciclin I. *EMBO J.* **13**, 4212–4222.
- John, M.E., and Keller, G.** (1995). Characterization of mRNA for a proline-rich protein of cotton fiber. *Plant Physiol.* **108**, 669–676.
- Kawamoto, T., Noshiro, M., Shen, M., Nakamasu, K., Hashimoto, K., Kawashima-Ohya, Y., Gotoh, O., and Kato, Y.** (1998). Structural and phylogenetic analyses of RGD-CAP/ β ig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. *Biochim. Biophys. Acta* **1395**, 288–292.
- Kehoe, D.M., Villand, P., and Somerville, S.** (1999). DNA microarrays for studies of higher plants and other photosynthetic organisms. *Trends Plant Sci.* **4**, 38–41.
- Kieliszewski, M.J., and Lampert, D.T.A.** (1994). Extensin: Repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J.* **5**, 157–172.
- Klahre, U., Noguchi, T., Fujioka, S., Takatsuto, S., Yokota, T., Nomura, T., Yoshida, S., and Chua, N.-H.** (1998). The Arabidopsis DIMINUTO/DWARF1 gene encodes a protein involved in steroid synthesis. *Plant Cell* **10**, 1677–1690.
- Kleeff, J., Ishiwata, T., Kumbasar, A., Friess, H., Büchler, M.W., Lander, A.D., and Korc, M.** (1998). The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *J. Clin. Invest.* **102**, 1662–1673.
- Knox, J.P.** (1995). The extracellular matrix in higher plants. 4. Developmentally regulated proteoglycans and glycoproteins of the plant cell surface. *FASEB J.* **9**, 1004–1012.
- Knox, J.P.** (1997). The use of antibodies to study the architecture and developmental regulation of plant cell walls. *Int. J. Cytol.* **171**, 79–120.
- Kreuger, M., and van Holst, G.-J.** (1996). Arabinogalactan-proteins and plant differentiation. *Plant Mol. Biol.* **30**, 1077–1086.
- Kunze, M., Riedel, J., Lange, U., Hurwitz, R., and Tischner, R.** (1997). Evidence for the presence of GPI-anchored PM-NR in leaves of *Beta vulgaris* and for PM-NR in barley leaves. *Plant Physiol. Biochem.* **35**, 507–512.
- Lease, K., Ingham, E., and Walker, J.C.** (1998). Challenges in understanding RLK function. *Curr. Opin. Plant Biol.* **1**, 388–392.

- Loopstra, C.A., and Sederoff, R.R.** (1995). Xylem-specific gene expression in loblolly pine. *Plant Mol. Biol.* **27**, 277–291.
- McAdams, H.H., and Arkin, A.** (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69.
- McCabe, P.F., Valentine, T.A., Forsberg, L.S., and Pennell, R.I.** (1997). Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot. *Plant Cell* **9**, 2225–2241.
- McClure, B.A., Gray, J.E., Anderson, M.A., and Clarke, A.E.** (1990). Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. *Nature* **347**, 757–710.
- Meagher, R.B., McKinney, E.C., and Kandasamy, M.K.** (1999). Isovariant dynamics expand and buffer the responses of complex systems: The diverse plant actin gene family. *Plant Cell* **11**, 995–1005.
- Mort, A.J., and Lampion, D.T.A.** (1977). Anhydrous hydrogen fluoride deglycosylates glycoproteins. *Anal. Biochem.* **82**, 289–309.
- Mozdzanowski, J., Bongers, J., and Anumula, K.** (1998). High-yield deblocking of amino termini of recombinant immunoglobulins with pyroglutamate aminopeptidase. *Anal. Biochem.* **260**, 183–187.
- Myerson, D.** (1991). Producing single-stranded DNA probes with *Taq* DNA polymerase: A high yield protocol. *BioTechniques* **10**, 35–38.
- Nakai, K., and Horton, P.** (1999). PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24**, 34–36.
- Nam, J., Mysore, K.S., Zheng, C., Knue, M.K., Matthyse, A.G., and Gelvin, S.B.** (1999). Identification of T-DNA tagged Arabidopsis mutants that are resistant to transformation by *Agrobacterium*. *Mol. Gen. Genet.* **261**, 429–438.
- Needleman, S.B., and Wunsch, C.D.** (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453.
- Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Somerville, C.** (1994). Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* **106**, 1241–1255.
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G.** (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6.
- Nothnagel, E.A.** (1997). Proteoglycans and related components in plant cells. *Int. Rev. Cytol.* **174**, 195–291.
- Oxley, D., and Bacic, A.** (1999). Structure of the glycosylphosphatidylinositol anchor of an arabinogalactan-protein from *Pyrus communis* suspension-cultured cells. *Proc. Natl. Acad. Sci. USA* **96**, 14246–14251.
- Pearson, W.R., and Lipman, D.J.** (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
- Peles, E., Nativ, M., Lustig, M., Grumet, M., Schilling, J., Martinez, R., Plowman, G.D., and Schlessinger, J.** (1997). Identification of a novel contactin-associated transmembrane receptor with multiple domains implicated in protein-protein interactions. *EMBO J.* **16**, 978–988.
- Pickett, F.B., and Meeks-Wagner, D.R.** (1995). Seeing double: Appreciating genetic redundancy. *Plant Cell* **7**, 1347–1356.
- Prokop, A.** (1999). Integrating bits and pieces: Synapse structure and formation in *Drosophila* embryos. *Cell Tissue Res.* **297**, 169–186.
- Reiter, R.S., Young, R.M., and Scolnik, P.A.** (1992). Genetic linkage of the Arabidopsis genome: Methods for mapping with recombinant inbreds and random amplified polymorphic DNAs (RAPDs). In *Methods in Arabidopsis Research*, C. Koncz, N.-H. Chua, and J. Schell, eds (Singapore: World Scientific), pp. 170–190.
- Resta, R., Yamashita, Y., and Thompson, L.F.** (1998). Ecto-enzyme and signaling functions of lymphocyte CD73. *Immunol. Rev.* **161**, 95–109.
- Roy, S., Jauh, G.Y., Hepler, P.K., and Lord, E.M.** (1998). Effects of Yariv phenylglycoside on cell wall assembly in the lily pollen tube. *Planta* **204**, 450–458.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O.** (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.
- Schultz, C.J., Gilson, P., Oxley, D., Youl, J., and Bacic, A.** (1998). GPI-anchors on arabinogalactan-proteins: Implications for signaling in plants. *Trends Plant Sci.* **3**, 426–431.
- Schultz, C.J., Hauser, K., Lind, J.L., Atkinson, A.H., Pu, Z.-Y., Anderson, M.A., and Clarke, A.E.** (1997). Molecular characterization of a cDNA sequence encoding the backbone of a style-specific 120kDa glycoprotein which has features of both extensins and arabinogalactan-proteins. *Plant Mol. Biol.* **35**, 833–845.
- Serpe, M.D., and Nothnagel, E.A.** (1994). Effects of Yariv phenylglycosides on *Rosa* cell suspensions: Evidence for the involvement of arabinogalactan-proteins in cell proliferation. *Planta* **193**, 542–550.
- Sherrier, D.J., Prime, T.A., and Dupree, P.** (1999). Glycosylphosphatidylinositol-anchored cell-surface proteins from Arabidopsis. *Electrophoresis* **20**, 2027–2035.
- Skonier, J., Bennett, K., Rothwell, V., Kosowski, S., Plowman, G., Wallace, P., Edelhoff, S., Disteché, C.M., Neubauer, M., Marquardt, H., Rodgers, J., and Purchio, A.F.** (1994). β Ig-h3: A transforming growth factor-beta-responsive gene encoding a secreted protein that inhibits cell attachment in vitro and suppresses the growth of CHO cells in nude mice. *DNA Cell. Biol.* **13**, 571–584.
- Sommer-Knudsen, J., Bacic, A., and Clarke, A.E.** (1998). Hydroxyproline-rich plant glycoproteins. *Phytochemistry* **47**, 483–497.
- Stöhr, C., Schuler, F., and Tischner, R.** (1995). Glycosyl-phosphatidylinositol-anchored proteins exist in the plasma membrane of *Chlorella saccharophila* (Krüger) Nadson: Plasma membrane-bound nitrate reductase as an example. *Planta* **196**, 284–287.
- Svetek, J., Yadav, M.P., and Nothnagel, E.A.** (1999). Presence of a glycosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. *J. Cell Biol.* **274**, 14724–14733.
- Takahashi, T., Gasch, A., Nishizawa, N., and Chua, N.-H.** (1995). The *DIMINUTO* gene of Arabidopsis is involved in regulating cell elongation. *Genes Dev.* **9**, 97–107.
- Takos, A.M., Dry, I.B., and Soole, K.L.** (1997). Detection of glycosyl-phosphatidylinositol-anchored proteins on the surface of *Nicotiana tabacum* protoplasts. *FEBS Lett.* **405**, 1–4.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J.** (1994). Clustal W: Improving the sensitivity of progressive multiple sequence align-

ment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Toonen, M.A.J., Schmidt, E.D.L., van Kammen, A., and de Vries, S.C. (1997). Promotive and inhibitory effects of diverse arabinogalactan proteins on *Daucus carota* L. somatic embryogenesis. *Planta* **203**, 188–195.

Udenfriend, S., and Kodukula, K. (1995a). How glycosyl-phosphatidylinositol-anchored membrane proteins are made. *Annu. Rev. Biochem.* **64**, 563–591.

Udenfriend, S., and Kodukula, K. (1995b). Prediction of ω site in nascent precursor of glycosylphosphatidylinositol protein. *Methods Enzymol.* **250**, 571–582.

Wang, J., Maziarz, K., and Ratnam, M. (1999). Recognition of the carboxyl-terminal signal for GPI modification requires transloca-

tion of its hydrophobic domain across the ER membrane. *J. Mol. Biol.* **286**, 1303–1310.

Willats, W.G.T., and Knox, J.P. (1996). A role for arabinogalactan-proteins in plant cell expansion—Evidence from studies on the interaction of beta-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. *Plant J.* **9**, 919–925.

Yariv, J., Raport, M.M., and Graf, L. (1962). The interaction of glycosides and saccharides with antibody to the corresponding phenylazo glycosides. *Biochem. J.* **85**, 383–388.

Youl, J.J., Bacic, A., and Oxley, D. (1998). Arabinogalactan-proteins from *Nicotiana glauca* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. *Proc. Natl. Acad. Sci. USA* **95**, 7921–7926.

Zinn, K., McAllister, L., and Goodman, C. (1988). Sequence analysis and neuronal expression of fasciclin I in grasshopper and *Drosophila*. *Cell* **53**, 577–587.