Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis

Zhan Dong Zhao[1], Li Tan[1], Allan Marshall Showalter[2,3], Derek Thomas Anthony Lamport[4] and Marcia Jane Kieliszewski[1,3,*]

[1]*Department of Chemistry and Biochemistry,* [2]*Department of Environmental and Plant Biology and*
[3]*Molecular and Cellular Biology Program, Ohio University, Athens, OH 45710, USA, and*
[4]*School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK*

Summary

Functional analysis of the hyperglycosylated arabinogalactan-proteins (AGPs) attempts to relate biological roles to the molecular properties that result largely from *O*-Hyp glycosylation putatively coded by the primary sequence. The Hyp contiguity hypothesis predicts contiguous Hyp residues as attachment sites for arabino-oligosaccharides (arabinosides) and clustered, non-contiguous Hyp residues as arabinogalactan polysaccharide sites. Although earlier tests of naturally occurring hydroxyproline-rich glycoproteins (HRGPs) and HRGPs designed by synthetic genes were consistent with a sequence-driven code, the predictive value of the hypothesis starting from the DNA sequences of known AGPs remained untested due to difficulties in purifying a single AGP for analysis. However, expression in tobacco (*Nicotiana tabacum*) of the major tomato (*Lycopersicon esculentum*) AGP, LeAGP-1, as an enhanced green fluorescent protein fusion glycoprotein (EGFP)-LeAGP-1, increased its hydrophobicity sufficiently for chromatographic purification from other closely related endogenous AGPs. We also designed and purified two variants of LeAGP-1 for future functional analysis: one lacking the putative glycosylphosphatidylinositol (GPI)-anchor signal sequence; the other lacking a 12-residue internal lysine-rich region. Fluorescence microscopy of plasmolysed cells confirmed the location of LeAGP-1 at the plasma membrane outer surface and in Hechtian threads. Hyp glycoside profiles of the fusion glycoproteins gave ratios of Hyp-polysaccharides to Hyp-arabinosides plus non-glycosylated Hyp consistent with those predicted from DNA sequences by the Hyp contiguity hypothesis. These results demonstrate a route to the purification of AGPs and the use of the Hyp contiguity hypothesis for predicting the Hyp *O*-glycosylation profile of an HRGP from its DNA sequence.

Keywords: arabinogalactan-protein, hydroxyproline-rich glycoprotein, O-glycosylatoin, glycosyl-phosphatidylinositol anchor, green fluorescent protein.

Introduction

Arabinogalactan-proteins (AGPs) are hyperglycosylated members of the hydroxyproline-rich glycoprotein (HRGP) superfamily and are broadly implicated in all aspects of growth and development, from fertilization to apoptosis (Bacic *et al.*, 1996; Gao and Showalter, 2000; Nam *et al.*, 1999). Many AGPs appear abundantly at the outer surface of the plasma membrane, where they are anchored by a C-terminal glycosylphosphatidylinositol (GPI) lipid (Oxley

and Bacic, 1999; Sherrier *et al.*, 1999; Svetek *et al.*, 1999; Youl *et al.*, 1998), allowing them to interact with other molecules of the extracellular matrix. Inevitably these interactions must involve the AGP carbohydrate components that occur as two distinct types of *O*-hydroxyproline (Hyp) glycosyl substituent: acidic or neutral arabinogalactan polysaccharides and small neutral oligosaccharides of arabinose (Fincher *et al.*, 1974;

Lamport, 1967; Lamport, 1984; Pope, 1977). These glyco-substituents occur along the polypeptide backbone in small clusters, forming conserved glycomodules that are therefore of functional significance (Shpak *et al.*, 1999; Shpak *et al.*, 2001), and so it is important to elucidate their arrangement, structure and roles in more detail. For example, the *O*-Hyp glycosylation code, which is based on correlations between amino acid sequences and their glycosubstituents, predicts the addition of oligosaccharides to contiguous Hyp and the addition of polysaccharides to clustered non-contiguous Hyp residues (Kieliszewski and Lamport, 1994; Kieliszewski *et al.*, 1992; Shpak *et al.*, 1999; Shpak *et al.*, 2001).

Previous tests of this Hyp contiguity hypothesis have employed naturally occurring HRGPs (Kieliszewski *et al.*, 1995) as well as novel HRGPs designed as single repetitive glycomodules via a synthetic gene approach (Shpak *et al.*, 1999; Shpak *et al.*, 2001). Thus dipeptidyl Hyp in a proline-rich HRGP from Douglas fir (*Pseudotsuga menziesii*; Kieliszewski *et al.*, 1995) was the major site of *O*-Hyp arabinosylation, while Hyp that occurred singly was only rarely arabinosylated. Similarly, synthetic genes encoding dipeptidyl and tetrapeptidyl Pro gave rise to Hyp that was only arabinosylated (Shpak *et al.*, 2001). In contrast, in HRGPs composed entirely of Ser-Hyp-Ser-Hyp (Shpak *et al.*, 1999) or Ala-Hyp-Ala-Hyp repeats (L.T. and M.J.K., unpublished results), 100% of the Hyp residues had arabinogalactan polysaccharide adducts in HRGPs.

Considering that more than 50% of eukaryotic proteins are glycoproteins (Apweiler *et al.*, 1999), prediction of potential glycosylation sites starting from a DNA sequence has important implications for the interpretation of plant genome and proteomics projects. However, the predictive value of the Hyp contiguity hypothesis for naturally occurring AGPs starting from a DNA sequence is hitherto untested, for two reasons. First, the multiplicity of AGPs with similar compositions precludes the isolation of a single AGP in sufficient purity and quantity for analysis; second, the identification of glycosylation site occupancy by the isolation and characterization of AGP glycopeptides is a formidable task due to the complexity of the degradation products, hence our initial synthetic gene approach (Shpak *et al.*, 1999; Shpak *et al.*, 2001). We approached the first problem by devising a new route to AGP purification; the second by the experimental determination of Hyp glycoside profiles as these accurately reflected the Hyp glycosylation site occupancy in the simple repetitive glycomodules produced by our initial synthetic gene approach (Shpak *et al.*, 1999; Shpak *et al.*, 2001). Here we extend this approach to the complex LeAGP-1, to ascertain if the Hyp glycoside profiles would corroborate those predicted from its DNA sequence.

We chose to characterize LeAGP-1, which is expressed abundantly in reproductive and rapidly expanding tissues of whole plants as well as tomato cell suspension-cultured cells (Gao *et al.*, 1999; Li and Showalter, 1996a; Li and Showalter, 1996b). These tomato AGPs were partially characterized earlier (Lamport, 1970) after isolation of an AGP fraction by isoelectric focusing. The fraction contained AGPs with an amino acid composition similar to that of LeAGP-1, and also contained galactose and arabinose in approximately equal amounts, which is typical for AGPs. LeAGP-1 is also characterized by genomic and cDNA cloning and represents a widespread type of AGP with paralogs and orthologs in *Arabidopsis* and *Nicotiana*, respectively (Gilson *et al.*, 2001). Indeed, the Hyp/Pro-rich regions of the tobacco (*Nicotiana alata*) ortholog shares 74% protein sequence identity with tomato LeAGP-1 (Gilson *et al.*, 2001). Unprocessed LeAGP-1 is a modular HRGP that consists of four distinct regions: an N-terminal signal sequence for secretion; followed by a Hyp-rich AGP domain that is interrupted by a short lysine-rich basic region; and finally a hydrophobic C-terminal sequence identified as a GPI-anchor addition signal sequence (Takos *et al.*, 2000).

Earlier we isolated a LeAGP-1 fraction after deglycosylation of an AGP mixture isolated from the growth medium and cytosol of tomato suspension-cultured cells (Gao *et al.*, 1999; Lamport, 1970). LeAGP-1 was the major AGP polypeptide present. However, as described here, despite extensive fractionation, *intact* LeAGP-1 could be purified to the point of an 'enriched component' in a mixture consisting exclusively of AGPs, but not as a single glycoprotein pure enough for an unambiguous glycosylation profile or for functional analysis. As an alternative route to Le-AGP-1 purification and functional analysis, we used transgenic enhanced green fluorescent protein (EGFP)-LeAGP-1 fusion glycoproteins expressed in tobacco. Green fluorescent protein not only aided detection, but also increased the hydrophobicity of the transgenic AGPs, enabling their facile separation from multiple endogenous hydrophilic AGPs.

As an essential part of a modular approach to the functional analysis of AGPs, we designed three variants of an EGFP-LeAGP-1 fusion protein: a complete AGP; one lacking the lysine-rich region; and another lacking the putative GPI-anchor signal sequence. We engineered their heterologous expression in bright yellow 2 (BY-2) tobacco cell suspension cultures because, unlike our tomato culture lines, BY-2 cells undergo facile *Agrobacterium*-mediated transformation. We were able to purify each variant from the culture medium of the transformed cells, and found that the Hyp-glycoside profiles of all three transgenic AGPs were indeed

consistent with the predictions of the Hyp contiguity hypothesis.

Results

LeAGP-1 partial purification from the culture medium of tomato cell suspensions

Isoelectric focusing of tomato AGPs gave one major unresolved AGP fraction with a pl ranging from 4 to 7, and a minor AGP fraction with a pl of 2 (data not shown). LeAGP-1 occurred in the pl 4–7 fraction, judging by immunoassays of the recovered fractions.

Further fractionation of the focused AGPs (*pl* 4–7 range) by combinations of gel permeation, ion exchange and reversed phase chromatography yielded a peak that appeared chromatographically pure (Figure 1a) and was recognized by LeAGP-1-specific antibodies (Gao and Showalter, 2000). However, after hydrogen fluoride (HF) deglycosylation and reversed-phase chromatography (Figure 1b) the glycoprotein resolved as several polypeptide backbones, the contaminating peaks accounting for about 50% of the protein present. Judging from immunoreactivity, amino acid compositions and peptide sequences (Gao *et al.*, 1999; Figure 1), the major component was LeAGP-1 while several minor components were distinct, although closely related, AGPs rather than isoforms of LeAGP-1.

As unambiguous characterization of AGP Hyp glycosylation requires the pure glycoprotein, we adopted an alternative purification strategy based on expressing a transgenic LeAGP-1 in tobacco, an easily transformable, closely related solanaceous species.

Plasmid design and plant transformation

DNA sequence analysis of the plasmids confirmed in-frame insertion of the EGFP gene between the LeAGP-1 signal sequence and the AGP module (*EGFP-LeAGP-1*; Figure 2); that the GPI anchor sequence was replaced by a stop codon in the construction designed to express an EGFP-LeAGP-1 lacking the GPI-anchor (*EGFP-LeAGP-1ΔGPI*); and that the lysine-rich region was absent from *EGFP-LeAGP-1ΔK*.

Localization of the gene products

Cells expressing EGFP-LeAGP-1 (Figure 3a), EGFP-LeAGP-1ΔGPI and EGFP-LeAGP-1ΔK (not shown) showed an intense peripheral fluorescence. In plasmolysed EGFP-LeAGP-1 cells, the fluorescence remained at the surface of the protoplast, including the plasma membrane extensions (Hechtian threads; Hecht, 1912), while the cell wall did not fluoresce (Figure 3b).
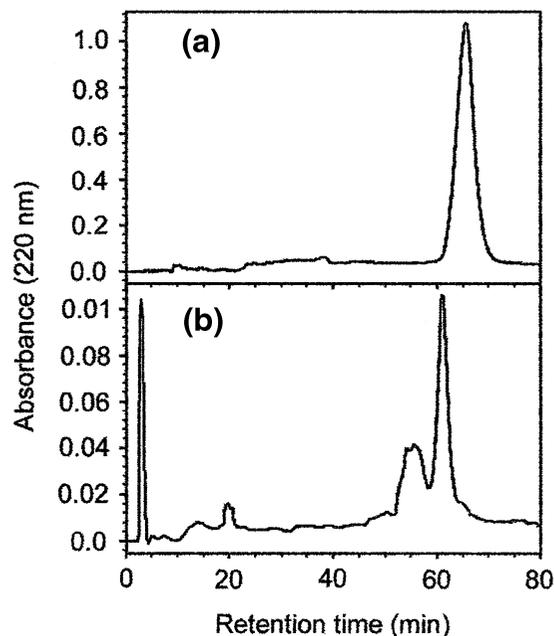

Figure 1. LeAGP-1-enriched fraction isolated from tomato medium (a)before and (b)after HF-deglycosylation.

(a) Combinations of isoelectric focusing, anion exchange and reversed-phase chromatography yielded a single symmetrical peak of constant composition that reacted with antibodies raised against the lysine-rich region of LeAGP-1. However, there was still significant contamination of LeAGP-1 with other AGPs (50% of the peak area was contaminating AGPs) that became obvious only after HF-deglycosylation and fractionation of the deglycosylated sample on PRP-1.

(b) The contaminating AGPs were probably not isoforms of LeAGP-1 judging by a peptide partial sequence from the contaminating 55 min peak: Ser-Ala-Tyr-Ile-Ser-Hyp-Ala-Hyp-Val-Ala-Ser-Hyp-Hyp.

The gradient featured in (a) was 0–12% buffer B over 72 min; LeAGP-1 eluted in ≈10% buffer B. The gradient featured in (b) was 0–40% buffer B over 100 min; LeAGP-1 eluted at 60 min in ≈25% buffer B.

Expression of EGFP-LeAGP-1ΔK gave a similar distribution of fluorescence. However, transformants expressing EGFP-LeAGP-1ΔGPI secreted the fusion glycoprotein and also exhibited general fluorescence throughout the cell (not shown), presumably due to the intracellular biosynthesis and transit of the fusion glycoprotein.

Transgene products, purification and yields

After filtration of the medium from cells expressing the fusion glycoproteins, fractionation of the concentrated growth medium on phenyl sepharose yielded single green fluorescent peaks (not shown), further purified via PRP-1 reversed-phase chromatography (Figure 4).

Yields (weight recovered) of the purified fusion glycoproteins from the culture medium varied between cell lines. The non-GPI-anchored EGFP-LeAGP-1ΔGPI cells gave 4.9–6.3 mg l^{-1} culture medium, consistently more than cells expressing the putatively GPI-anchored fusion

BamHI
AGG ATC CGG TCT ATA TTT TCT TTA GCT ACC ATG GAT AGA AAA TTT GTA TTT CTA GTA TCA ATT TTG
                                    M   D   R   K   F   V   F   L   V   S   I   L

Xmal
TGC ATT GTA GTG GCA AGT GTC ACG GGT CAG ACA ACC CGG GTA CCG GTC GCC ACC ATG –EGFP-
C   I   V   V   A   S   V   T   G   Q   T   T   R   V   P   V   A   T   M

BsrGl
GAG CTG TAC ACG GGT CAG ACA CCT GCC GCA GCA CCC GTT GGT GCT AAG GCT GGT ACT ACT CCA
E   L   Y   T   G   Q   T   P   A   A   A   P   V   G   A   K   A   G   T   T   P

CCA GCT GCT GCA CCA ACT AAG CCG AAA ACT CCT GCT CCC GCA ACA GCA CCA GCC TCG GCC CCA
P   A   A   A   P   T   K   P   K   T   P   A   P   A   T   A   P   A   S   A   P

CCT ACA GCT GTT CCT GTT GCT CCA GTA ACC GCT CCA GTT ACT GCT CCC ACT ACA CCT GTT GTT
P   T   A   V   P   V   A   P   V   T   A   P   V   T   A   P   T   T   P   V   V

GCT GCA CCC GTA TCA GCA CCA GCA AGT TCT CCA CCA CTT AAA GCA CCA GCA AGT TCT CCA CCA
A   A   P   V   S   A   P   A   S   S   P   P   L   K   A   P   A   S   S   P   P

GTA CAA TCT CCA CCA GCT CCA GCT CCA GAG GTA GCT ACA CCG CCA GCT GTT TCT ACT CCA CCG
V   Q   S   P   P   A   P   A   P   E   V   A   T   P   P   A   V   S   T   P   P

GCT GCA GCT CCA GTT GCT GCA CCT GTT GCT TCG GAG ACA ACT CCA GCT CCG GCT CCT AGC AAA
A   A   A   P   V   A   A   P   V   A   S   E   T   T   P   A   P   A   P   S   K

Nsil
GGA AAA GTA AAG GGA AAG AAG GGA AAG AAA CAC AAT GCA TCA CCA GCA CCT TCT CCC GAT ATG
G   K   V   K   G   K   K   G   K   K   H   N   A   S   P   A   P   S   P   D   M

ATG AGC CCA CCT GCA CCT CCT TCC GAA GCT CCT GGA CCT AGC ATG GAC TCC GAT TCA GCT CCC
M   S   P   P   A   P   P   S   E   A   P   G   P   S   M   D   S   D   S   A   P

AGC CCA TCT CTT AAC GAT GAG AGT GGA GCA GAG AAA TTG AAG ATG CTG GGA AGT TTG GTA GCT
S   P   S   L   N   D   E   S   G   A   E   K   L   K   M   L   G   S   L   V   A

Eagl
GGA TGG GCT GTG ATG AGC TGG CTC TTG TTC TAA AGC GGC CGC ACC GCG
G   W   A   V   M   S   W   L   L   F   Stop

Figure 2. DNA sequence of *EGFP-LeAGP-1* and the corresponding primary sequence.
We inserted the EGFP (green) between the LeAGP-1 signal sequence (red) and the rest of LeAGP-1 (black, pink and blue). Pink highlights the lysine-rich region; blue denotes the putative GPI anchor-addition sequence. Underlined are the regions of this gene common to the oligonucleotides used to introduce the restriction sites (bold and labeled), enabling the construction of plasmids *pUC-SS^tom-EGFP-LeAGP-1*, *pUC-SS^tom-EGFP-LeAGP-1ΔK*, and *pUC-SS^tom-EGFP-LeAGP-1ΔGPI*. The chymotrypsin-labile Tyr residue engineered into the fusion glycoprotein to allow removal of EGFP from LeAGP-1 is encoded in the *Bsr*GI restriction site.
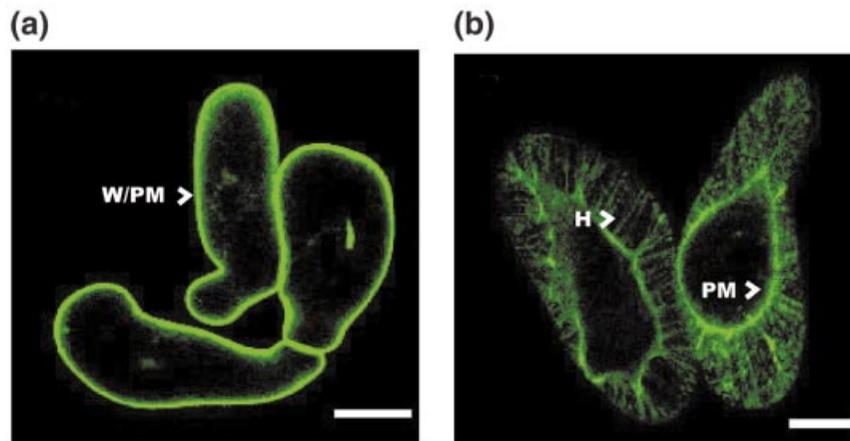

Figure 3. Tobacco cells expressing EGFP-LeAGP-1 before (a) and after (b) plasmolysis.
(a) EGFP fluorescence was observed at the cell wall–plasma membrane interface (W/PM); the localization of EGFP at the membrane (PM) rather than in the cell wall was obvious only after the membrane was retracted during plasmolysis in 5% sodium chloride.
(b) EGFP-LeAGP-1 also occurs in the Hechtian threads (H), connections between the wall and the membrane.
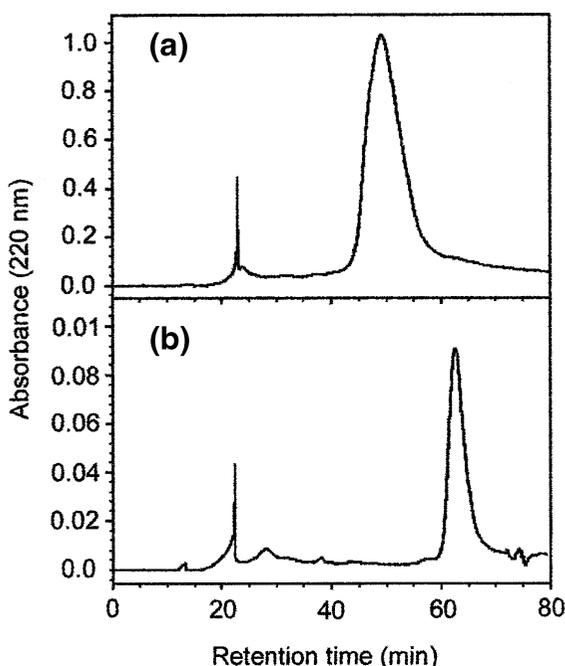Scale bars, 20 μm.

Figure 4. Purification of EGFP-LeAGP-1 by reversed-phase chromatography on PRP-1.

(a) Glycosylated tobacco EGFP-LeAGP-1 yielded a single symmetrical peak after purification by hydrophobic interaction chromatography (not shown) followed by PRP-1 chromatography.

(b) After HF-deglycosylation EGFP-LeAGP-1 still yielded a single peak on PRP-1, in contrast to heterogeneous tomato LeAGP-1 shown in Figure 1(a,b).

The gradient in (a) was 0–10% buffer B (see Experimental procedures) for 10 min then 70% buffer B by 70 min. EGFP-LeAGP-1 eluted in ≈50% buffer B. Due to the presence of EGFP in the fusion glycoprotein, a much higher concentration of acetonitrile was required for elution of EGFP-LeAGP-1 from the column than for tomato LeAGP-1, which eluted in 10% buffer B (Figure 1a). The gradient in (b) was 0–80% buffer B for 100 min. Deglycosylated EGFP-LeAGP-1 was eluted in ≈60% buffer B, being a much more hydrophobic protein than deglycosylated LeAGP-1 (Figure 1b) due to EGFP.

glycoproteins: EGFP-LeAGP-1, 0.46–1.1 mg l^{-1}; EGFP-LeAGP-1ΔK, 0.53–1.2 mg l^{-1}.

Characterization of the transgene products

Amino acid composition and protein sequence analysis. The fusion glycoproteins were purified to a constant composition and deglycosylated, and EGFP was removed by chymotryptic cleavage (at the Tyr residue introduced by the *Bsr*GI restriction site; Figure 2). The compositions of transgenic LeAGP-1, LeAGP-1ΔK and LeAGP-1ΔGPI after EGFP removal (Table 1) matched those predicted from the gene sequences, with most of the Pro residues hydroxylated to Hyp. In addition, Edman degradation of LeAGP-1 yielded a partial N-terminal
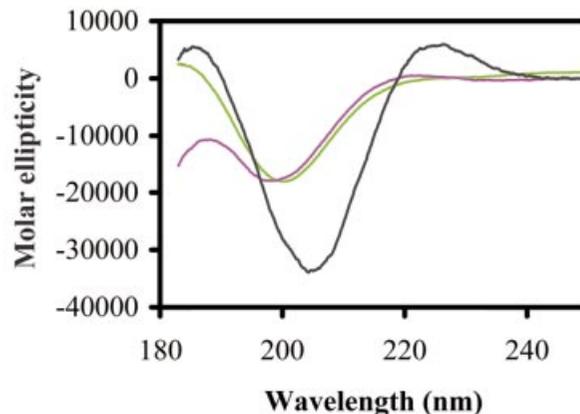
Figure 5. Circular dichroism spectra of the tobacco LeAGP-1 module before and after deglycosylation, compared to a polyhydroxyproline standard.

EGFP was removed from EGFP-LeAGP-1 by chymotryptic digestion and the glycoprotein isolated by reversed-phase chromatography on PRP-1. The black curve with a minimum at 206 nm and a maximum at 225 nm corresponds to the polyhydroxyproline standard in the polyproline II conformation, a left-handed helix with three residues/turns and a pitch of 0.94 nm (Homer and Roberts, 1979). Both glycosylated LeAGP-1 (pink) and deglycosylated LeAGP-1 (green) favor a 'random coil conformation'; however, the presence of arabinosides in LeAGP-1 shifted the spectrum minimum from 203 nm (green) to 201 nm (pink) (Shpak *et al.*, 2001), and the presence of polysaccharide contributed to the second minimum at 183 nm (pink) (Shpak *et al.*, 2001).

sequence, TGQTOAAAXVGAKAGTTOOAAO (X denotes a blank cycle), that also confirmed its identity.

Sugar composition. Neutral sugar analyses, uronic acid assays (Table 2) and linkage analyses showed that all three of the fusion glycoproteins were, like other AGPs, rich in galactose (1,3-linked Gal; 1,6-linked Gal; 1,3,6-linked Gal; terminal galactose) and arabinose (1,2-linked Ara; 1,5-linked Ara; terminal Ara), with lesser amounts of terminal rhamnose and 4-linked glucuronic acid. All the fusion glycoproteins were precipitated by the Yariv reagent (Yariv *et al.*, 1962).

Hyp-glycoside profiles of the three fusion glycoproteins showed that 52–56% of the total Hyp had polysaccharide substituents. The rest of the Hyp was either non-glycosylated, or had arabinoside substituents (Table 3).

Circular dichroism

Both glycosylated and deglycosylated LeAGP-1 fusion glycoproteins (without EGFP) showed random coil conformations, with the ≈200 nm minimum shifted slightly to a lower wavelength in the glycosylated module (Figure 5). The glycosylated LeAGP-1 module also showed a second minimum at 183 nm, a minimum that occurs only in HRGPs containing arabinogalactan polysaccharide substituents (Shpak *et al.*, 2001).

Table 1. Amino acid compositions of tobacco LeAGP-1 glycoproteins LeAGP-1ΔK, LeAGP-1ΔGPI and LeAGP-1 compared with deglycosylated LeAGP-1 isolated from tomato cultures and *LeAGP-1*c DNA (Gao *et al.*, 1999)

Amino acid	Composition (mol %)				
	Tobacco LeAGP-1ΔK	Tobacco LeAGP-1ΔGPI	Tobacco LeAGP-1	Tomato LeAGP-1	Tomato *LeAGP-1c* DNA
Hyp	24	25	24	29	–
Asx	5	4	4	2	3.5
Thr	10	10	10	10	8.8
Ser	11	10	11	12	11.7
Glx	5	5	5	3	3.5
Pro	2	3	2	1	27.5
Gly	2	3	2	5	3.5
Ala	24	22	26	21	23.4
Val	9	9	9	9	8.2
Met	2	1	2	0	1.7
Ile	1	0	0	1	0.0
Leu	2	1	2	1	1.2
Tyr	0	0	0	0	0.0
Phe	0	0	0	1	0.0
His	0	0	0	0	0.6
Lys	4	4	4	5	6.4
Arg	1	1	0.0	0.0	0.0
Trp	nd	nd	nd	nd	0.0

The composition of transgenic LeAGP-1 shows less Lys than predicted, while LeAGP-1ΔK shows more than the predicted 2.5 mol %. However, consistent with the DNA sequences of the transgenes, transgenic LeAGP-1 reacted strongly with antibodies raised against the lysine-rich region (Gao *et al.*, 1999), while LeAGP-1ΔK showed no reaction. EGFP was removed from the fusion glycoproteins before analysis.nd, Not determined.

Discussion

Arabinogalactan-proteins are difficult to purify because they are numerous, very closely related compositionally, and highly glycosylated, resulting in a range of glycoforms for each core polypeptide that defines an AGP. Because conventional methods of separation invariably yield fractions containing multiple AGP polypeptides, we selected an AGP and altered its properties by genetic engineering to facilitate its separation from all other AGPs. This approach is based on earlier work using synthetic genes to express simple repetitive AGP glycomotifs as AGP-EGFP fusion glycoproteins secreted into the growth medium of tobacco cultures (Shpak *et al.*, 1999). The EGFP tag increased the hydrophobicity of the hydrophilic expression products, enabling their separation from other hydrophilic AGPs. This suggested a general approach to the problem of purifying known AGPs.

We chose LeAGP-1 based on its constitutive expression as a major AGP of tomato cultures and its availability as clones. The LeAGP-1 variants were designed to contain an N-terminal EGFP preceded by an N-terminal secretion signal sequence (Figure 2). Such EGFP fusion glycoproteins also had the advantage of permitting continuous fluorescence monitoring of column eluates, as well as *in situ* visualization by confocal fluorescence microscopy.

Table 2. Glycosyl compositions of LeAGP-1 fusion glycoproteins expressed in tobacco

Glycosyl residue	Composition (mol %)		
	LeAGP-1	LeAGP-1ΔK	LeAGP-1ΔGPI
Rha	8	10	13
Ara	39	34	33
Gal	37	39	38
GlcUA	12	15	14
Man	4	2	2

Molar ratios were derived from gas–liquid chromatography after alditol acetate derivation and colorimetric uronic acid assays. The presence of glucuronic acid was established by methylation analysis. EGFP itself contains no sugar (Shpak *et al.*, 1999).

Tobacco BY-2 cells were chosen as the expression system because *Agrobacterium* readily yielded stable transformants, whereas our tomato cell lines did not. Also, we aim eventually to compare the glycosylation profiles of the same LeAGP-1 polypeptide in two different, though related, solanaceous species, tobacco and tomato. Furthermore, homologs of LeAGP-1, including tobacco, are widespread (Schultz *et al.*, 2000) and their expression is

Table 3. Relative amounts of Hyp polysaccharide, Hyp arabinosides and non-glycosylated Hyp predicted by the Hyp contiguity hypothesis, compared with profiles obtained experimentally from isolated tobacco LeAGP-1 fusion glycoproteins

Hyp glycoside	Percentage of total hydroxyproline			
	Predicted[a]	LeAGP-1	LeAGP-1ΔK	LeAGP-1ΔGPI
Hyp-PS	58–60	54	52	56
Hyp-Ara$_4$		4	7	2
Hyp-Ara$_3$	28	11	13	7
Hyp-Ara$_2$		10	8	17
Hyp-Ara		8	3	4
NG-Hyp	12–14	13	17	14

Hyp-PS, Hyp polysaccharide; Hyp-Ara$_n$, Hyp-arabinoside$_{1-4}$; NG-Hyp, non-glycosylated Hyp.
[a]The percentage of Hyp-PS predicted is the maximum assuming that each non-contiguous Hyp has a polysaccharide adduct.

correlated with cell expansion but repressed by wounding (Gilson *et al.*, 2001), as is the expression of LeAGP-1 (Gao and Showalter, 2000; Li and Showalter, 1996a).

Transformed cells exhibited a uniform peripheral green fluorescence when turgid; on plasmolysis the fluorescence remained largely with the retracted protoplast, indicating an even, non-punctate distribution of LeAGP-1 over the membrane surface. However, numerous thin, highly fluorescent extensions of the membrane – Hechtian threads (Buer *et al.*, 2000; Hecht, 1912; Lamport, 1965) connected the shrunken protoplast to the barely discernible non-fluorescent cell wall (Figure 3). This dramatically confirms that regions of firm attachment exist between the plasma membrane and the cell wall. Such regions (Garcia-Gomez *et al.*, 2000) may represent specific adhesion microdomains involved in cell wall assembly and wall metabolism, or anchor sites that facilitate mitosis and cytokinesis (Cleary, 2001).

Non-GPI-anchored EGFP-LeAGP-1ΔGPI was freely secreted. However EGFP-LeAGP-1 and EGFP-LeAGP-1ΔK, both targeted to the outer surface of the plasma membrane, also appeared in the culture medium – although in lesser amounts – perhaps due to phospholipase cleavage of the GPI-anchor (Gaspar *et al.*, 2001), as transgenic LeAGP-1 and LeAGP-1ΔK (both without EGFP) did not interact with liposomes in a GPI-anchor assay (Camolezi *et al.*, 1999; J. Xu, Z.D.Z. and M.J.K., unpublished results).

The accumulation of the LeAGP-1 fusion glycoproteins in the culture medium simplified the purification of each variant. This is the first demonstrated isolation of a known arabinogalactan-protein and its purification to homogeneity. We were able to purify bulk amounts of the LeAGP-1 fusion glycoproteins, thereby enabling their biochemical characterization including their Hyp glycosylation specifics.

Hydroxyproline-rich glycoprotein Hyp glycosylation depends primarily on the sequence-dependent specificities of both prolyl hydroxylases and *O*-Hyp glycosyltransferases. Plant prolyl hydroxylases follow sequence rules that differ markedly from the Gly-Pro-Xaa triplet collagen sequence code that animal prolyl hydroxylases recognize. Plants and some algae hydroxylate most, but not all, proline residues of proline-rich proteins targeted to the appropriate ER subdomain (Choi *et al.*, 2000; Ferris *et al.*, 2001). The known exceptions in plants include Lys-Pro, Ile-Pro, Leu-Pro, Tyr-Pro, Phe-Pro and Gly-Pro (Goodrum *et al.*, 2000; Kieliszewski and Lamport, 1994); furthermore, transgenic plants do not hydroxylate collagen expression products (Ruggiero *et al.*, 2000).

Judging from known glycosylated sequences of HRGPs, including synthetic gene expression products and the LeAGP-1 glycosylation profiles presented here, Hyp glycosyltransferases of plants recognize a fairly simple sequence code based on Hyp contiguity or non-contiguity. Hence the two general modes of Hyp glycosylation: arabinosylation of contiguous Hyp; and galactosylation of clustered non-contiguous Hyp residues.

Here we consider two aspects of Hyp *O*-glycosylation: the sites glycosylated and their glycosubstituents. The Hyp glycosylation code putatively determines both the sites glycosylated along the polypeptide, and the type of glycosubstituents at each site. Previous results obtained using naturally occurring HRGPs (Kieliszewski *et al.*, 1995) and synthetic genes to express simple repetitive Hyp glycomodules (Shpak *et al.*, 1999; Shpak *et al.*, 2001) support the premise that contiguous Hyp directs the addition of oligoarabinosides, while non-contiguous Hyp, particularly when clustered, directs the addition of arabinogalactan polysaccharide. A 100% addition of arabinogalactan polysaccharide to clustered non-contiguous Hyp in modules of repetitive Ser-Hyp-Ser-Hyp (Shpak *et al.*, 1999) or Ala-Hyp-Ala-Hyp (L.T. and M.J.K., unpublished results), compared with the exclusive addition of arabinosides to repetitive modules containing only contiguous Hyp (Shpak *et al.*, 2001), indicates that Hyp glycosylation is non-random and therefore code-driven. Here we extended those results by comparing the predicted and experimentally determined Hyp glycosylation profiles of a known AGP.

Excluding its signal sequences, LeAGP-1 DNA encodes 47 Pro residues of which 45 are hydroxylated to yield Hyp, judging by the amino acid compositions shown in Table 1. As Gly-Pro and Lys-Pro are not hydroxylated in known HRGP sequences (Goodrum *et al.*, 2000; Kieliszewski and Lamport, 1994), we designated these as the two Pro residues in the sequence shown in Figure 6. LeAGP-1 contains 27 non-contiguous Hyp residues, hence 27 candidates for arabinogalactan polysaccharide addition (60% of the total Hyp) according to the Hyp contiguity hypoth-
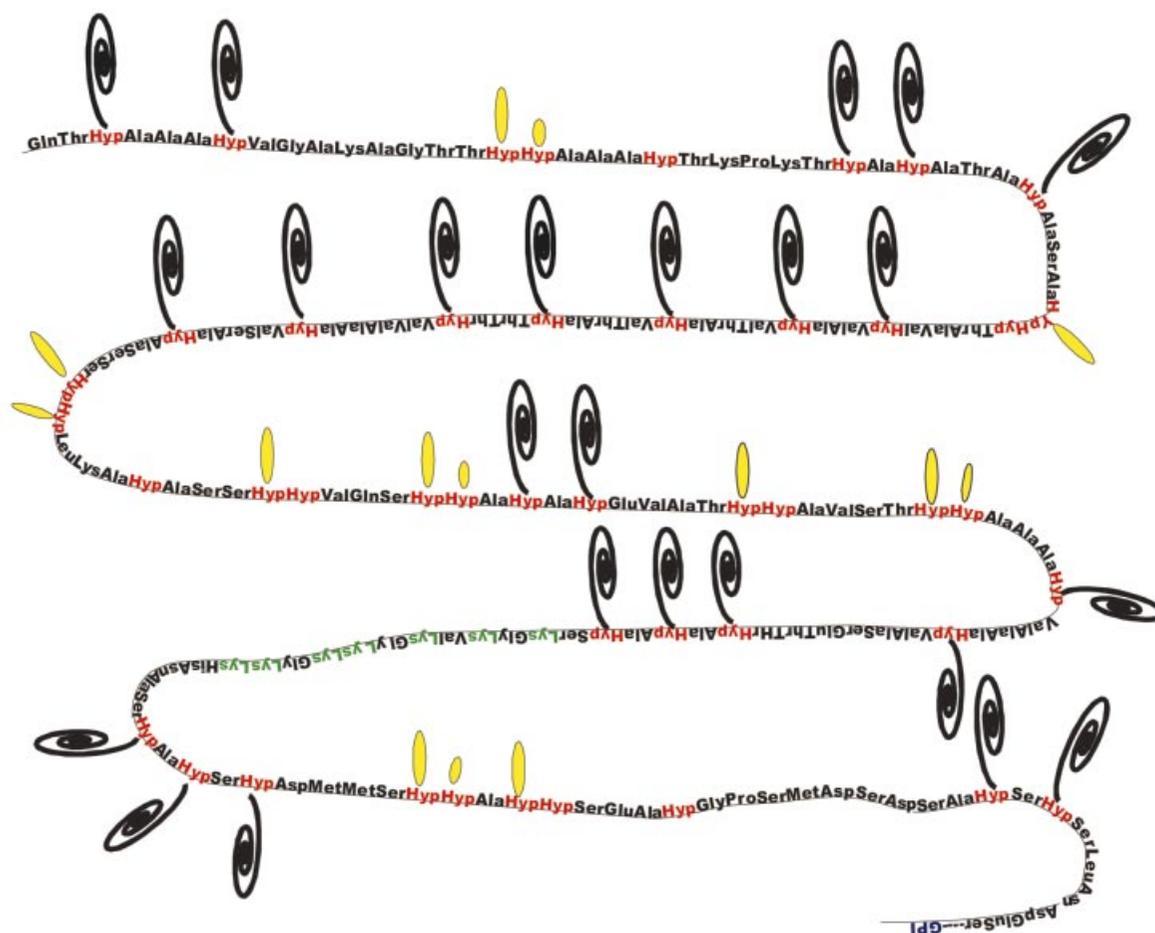
Figure 6. Proposed model of LeAGP-1 glycoprotein *O*-Hyp glycosylation.

Hyp residues are highlighted in red; black spirals represent arabinogalactan polysaccharide substituents; yellow ellipses represent arabinosides. The highly basic, lysine-rich module of LeAGP-1 is highlighted in green. Although we estimate there are about 27 non-contiguous Hyp residues in LeAGP-1 (≈60% of the total Hyp), not all have arabinogalactan-polysaccharide attached, judging by the Hyp-glycoside profiles shown in Table 3. We placed the arabinogalactan polysaccharide substituents on those non-contiguous Hyp residues that were most clustered. This figure features 24 Hyp residues with arabinogalactan polysaccharide substituents (≈53% of the total Hyp); 14 arabinosylated Hyp residues (≈31% of the total Hyp); and seven non-glycosylated Hyp residues (≈16% of the total Hyp in LeAGP-1). Needless to say, many closely related glycoforms of LeAGP-1 exist, as evidenced by the variations in the Hyp-glycoside profiles of the fusion glycoproteins shown in Table 3. Furthermore, carbohydrate may also occur on the Ser or Thr residues, additions that would be directed by different *O*-glycosylation codes.

esis that also suggests clustered non-contiguous Hyp residues are the primary candidates. This corresponds to 23–25 Hyp polysaccharides (shown as black spirals in Figure 6) and accounts for 52–56% of the total Hyp, consistent with the experimental data (Table 3). The Hyp contiguity hypothesis also predicts that a maximum of 18 LeAGP-1 Hyp residues will be arabinosylated because they are contiguous; this accounts for 40–41% of the total Hyp (maximum). However, earlier results with HRGPs containing dipeptidyl Hyp repeats (Kieliszewski *et al.*, 1995) showed that while the first of the two Hyp residues in dipeptidyl Hyp is always arabinosylated (this would account for nine Hyp residues in the LeAGP-1 fusion glycoproteins, or 20% of the total Hyp), the second Hyp in the block is incompletely arabinosylated (Shpak *et al.*,

2001). The actual figures of 30–33% Hyp arabinosylated (14–15 Hyp residues) with 13–17% remaining non-glycosylated (six to eight Hyp residues; Table 3) are therefore consistent with these earlier results.

Thus, starting from a DNA sequence, we can now tentatively predict the likely positions and types of *O*-Hyp glycosubstituents in the primary sequence of any appropriately targeted HRGP; this includes HRGP-based natural products, such as novel plant gums that can be designed through a synthetic gene approach. The Hyp contiguity hypothesis is therefore a useful predictive tool in the functional genomics toolbox, as it will help to define more clearly questions about the biosynthesis and degradation of AGPs. For example, does the initiation and further growth of a Hyp polysaccharide involve *en bloc* transfer of

a preformed oligosaccharide subunit from a lipid carrier to Hyp residues? There are some suggestive data in the literature (Bolwell, 1986; Bowles and Northcote, 1972; Hayashi and MacLachlan, 1984), including the *N*-glycosylation precedent and the occurrence of dolichyl-activated *O*-mannosylation of Ser/Thr residues in yeast (Gemmill and Trimble, 1999; Strahl-Bolsinger *et al.*, 1999). There is also evidence of AGP degradation or processing in some tissues (Labarca and Loewus, 1972; Labarca and Loewus, 1973; Wu *et al.*, 1995) that can now be quantified and compared with the intact AGP purified from an appropriate expression system. Related to this is the fact that not all AGP arabinogalactan polysaccharides share the same fine structure. This raises questions about tissue and species-specific differences in polysaccharide structure, and how such differences influence function. For example, the AGPs from *Nicotiana alata* stigmas and styles essentially lack uronic acid residues and rhamnose (Gane *et al.*, 1995) in contrast to the arabinogalactan polysaccharides we observed in the AGPs from BY2 cells.

If 'structure is function' (Liljas, 1999), then we need to consider how the predominating saccharide and polysaccharide components of AGPs contribute to the overall molecular surface and properties ultimately responsible for their various biological roles that are, as yet, formulated in only general terms. Two modes of *O*-Hyp glycosylation – arabinosylation and galactosylation – lead to the addition of short, neutral oligoarabinosides and larger, often acidic arabinogalactan polysaccharides, respectively. They form distinctive glycopeptide motifs (or glycomodules) with very different properties. For example, the neutral homo-oligoarabinosides of contiguous Hyp, particularly the tetra-arabinosylated Ser-Hyp$_4$ glycomodule, stabilize the polypeptide as an extended polyproline II-like structure (van Holst and Varner, 1984; Shpak *et al.*, 2001). The role of the often acidic Hyp heteropolysaccharides is less apparent and, judging from circular dichroism spectra (Figure 5; Shpak *et al.*, 2001), they do not stabilize the polypeptide in any particular conformation, although it is probably extended (Qi *et al.*, 1991) and relatively flexible compared to the extensins. However, small clusters of Hyp polysaccharide in LeAGP-1 peptide motifs such as TOAOATAO and VOVAOVTAO-VTAOTTO (Figure 6) form glycomodule clusters that seem designed for water retention through the short 1,6-linked mobile side chains (Rees, 1977) of the galactan backbone.

This raises questions about the amount, distribution and molecular orientation of AGPs at the cell surface. Despite early evidence for AGPs at the protoplast surface (Larkin, 1978), attempts to quantify AGP plasma membrane loading have been fraught with the uncertainties of membrane isolation and assay of AGPs. Thus quantitative descriptions of membrane AGPs have been limited to 'abundant', with two important exceptions: tobacco (Zhu *et al.*, 1993)

and rose cultures (Komalavilas *et al.*, 1991) containing, respectively, 160 and 66 µg AGP mg^{-1} membrane protein. A recent microscopic description of AGPs as polyhedral arrays forming a reticulum on the plasma membrane of isolated protoplasts (Gens *et al.*, 2000) also supports the idea of AGPs as major components (Sherrier *et al.*, 1999) of the plasma membrane. Such a glycocalyx resulting from AGP association (Du *et al.*, 1996; Gane *et al.*, 1995) via multiple weak homophilic interactions between the polysaccharide side chains might act as an assembly layer guiding the self-assembly of a cell-wall microcomposite. As β-glycosyl Yariv reagents precipitate AGPs (Jermyn and Yeow, 1975), the Yariv reagent may mediate disruption of AGP networks; this may account for inhibition of multiple processes including pollen tube extension (Roy *et al.*, 1998); root growth (Willats and Knox, 1996); root hair initiation (Ding and Zhu, 1997); cell suspension culture growth (Serpe and Nothnagel, 1994); cellulose deposition (Vissenberg *et al.*, 2001); and somatic embryogenesis (Chapman *et al.*, 2000). Variation in the walls that characterize different tissues may rationalize the increasingly large repertoire of known AGPs.

A strong correlation between cell extension and freely soluble AGPs (Willats and Knox, 1996) suggests that AGPs act as wall-loosening agents via intussusception *in muro* of 'polysaccharide expansins' regulated by the phospholipase-mediated release of their membrane-bound precursor AGPs. This implies a role for some AGPs more akin to lubricant than adhesive. For example, isolation of pipettable cell-suspension cultures from callus cultures depends on the friability of the callus (Lamport, 1964) and this, in turn (judging from the AGP content of the spent culture medium) may depend on abundant secretion of AGPs to ensure cell separation. The recognition that AGPs are highly modular proteins consisting largely, but not exclusively, of repetitive glycomodules now enables a more rigorous approach to their functional analysis. Thus, using the tools described here, we can design new or modified AGPs and, using appropriate expression systems, test for loss or gain of function.

Experimental procedures

Crude tomato AGPs isolated from tomato culture medium

We grew tomato suspension-cultured cells (Bonnie Best) on Shenk–Hildebrandt medium at 27°C (5% packed cell volume inoculum) for 8 days (Shpak *et al.*, 2001). Concentration of the culture filtrates by rotary evaporation was followed by precipitation of the AGPs with β-glucosyl Yariv reagent and its subsequent removal by reduction with sodium dithionite (Gao *et al.*, 1999) to yield a crude AGP preparation.

Co-precipitation with Yariv reagent

We co-precipitated AGPs with the Yariv reagent as described earlier (Serpe and Nothnagel, 1994), using tobacco crude AGPs as standards (Shpak *et al.*, 1999).

Tomato AGP fractionation by isoelectric focusing

Yariv-precipitated tomato AGPs (50 mg) were fractionated on a 110 ml LKB isoelectric focusing column (LKB Instruments Inc., Rockville, MD) in a pH 2–6 ampholyte range stabilized by a glycerol gradient at 10°C for 72 h at 300 V and a final current of 1–2 mA. After assaying aliquots by Yariv precipitation, we dialyzed and freeze-dried fractions for immunoassay and further fractionation.

Tomato AGPs fractionated by gel-permeation chromatography

We applied either crude or isoelectric focusing-fractionated tomato AGPs (100 mg) to a semi-preparative Superose-12 gel-permeation column (16 mm id × 500 mm, Amersham Pharmacia Biotech, Piscataway, NJ) equilibrated in 200 mM sodium phosphate buffer (pH 7) and eluted at a flow rate of 1 ml min^{-1}. Fractions (2 ml) were monitored at 220 nm.

Tomato AGPs fractionation by anion-exchange chromatography

Tomato AGPs (100 mg) were applied to a DEAE-Sepharose column (16 mm id × 700 mm; Amersham Pharmacia Biotech) equilibrated in 20 mM sodium acetate buffer (pH 6.0), and eluted with 300 ml 0–2 M sodium chloride gradient in acetate buffer at a flow rate of 1.2 ml min^{-1}. Fractions (2 ml) were monitored at 220 nm.

Tomato AGPs fractionated by reversed-phase liquid chromatography

Partially purified AGPs (after isoelectric focusing, gel permeation and anion-exchange fractionation) were loaded onto a Hamilton semi-preparative polymeric reversed-phase column (10 μm PRP-1, 7 × 305 mm; Hamilton Co., Reno, NV) equilibrated with buffer A (0.1% trifluoroacetic acid, TFA). Proteins were eluted for 72 min with a linear gradient of 0–12% buffer B (0.1% TFA/80% aqueous acetonitrile) at a flow rate of 0.75 ml min^{-1}. The fractions containing LeAGP-1 were determined by immunoassay (see below), then pooled, freeze-dried and refractionated on the PRP-1 column using the same conditions.

PRP-1-fractionated tomato LeAGP-1 was loaded onto a Hamilton analytical reversed-phase column (5 μm PRP-1, 4.6 × 150 mm, Hamilton Co.) equilibrated in buffer A. Proteins were eluted with a gradient of 0–12% buffer B in 72 min at a flow rate of 0.5 ml min^{-1}.

Deglycosylated LeAGP-1 fractions (see below) were separated on the analytical PRP-1 column equilibrated with buffer A and eluted with a gradient of 0–40% buffer B for 100 min at a flow rate of 0.5 ml min^{-1}. Column effluents were monitored for absorption at 220 nm.

Immunoassays

Samples dotted onto nitrocellulose filters were immunodetected using rabbit antibodies raised against the lysine-rich peptide of LeAGP-1 or anti-EGFP antibodies as primary antibodies; secondary antibodies were goat anti-rabbit IgG alkaline phosphatase-conjugated secondary antibodies (Gao and Showalter, 2000; Gao *et al.*, 1999).

Isolation of the LeAGP1 signal sequence and creation of plasmid pUC-SS^{tom}-EGFP

We amplified the tomato LeAGP-1 signal sequence using the cDNA *LeAGP-1c* (Li and Showalter, 1996a) and primers designed to introduce *Bam*HI and *Xma*I restriction sites to the amplified fragment. (The *LeAGP-1c* sequence is available in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X99148.) All primers were designed using PRIMER PREMIER software (Biosoft International, Palo Alto, CA) and synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). The signal sequence sense primer introduced the *Bam*HI site (underlined): 5′-CTC TTT TTC TCT GGA TCC GGT CTA TAT TTT CTT TTA GC-3′ and the antisense primer introduced an *Xma*I site (underlined): 5′-C GGG TGC TGC CCG GGT TGT CTG ACC CGT GAC ACT TGC-3′ (Figure 2). The isolated fragment was subcloned into plasmid *pUC-SS^{tob}-EGFP* (Shpak *et al.*, 1999) as a *Bam*HI/*Xma*I fragment in place of the tobacco extensin signal sequence (SS^{tob}). We designated this new *pUC19*-derived plasmid *pUC-SS^{tom}-EGFP*.

Creation of plasmids pUC-SS^{tom}-EGFP-LeAGP-1, pUC-SS^{tom}-EGFP-LeAGP-1ΔK and pUC-SS^{tom}-EGFP-LeAGP-1ΔGPI

pUC-SS^{tom}-EGFP-LeAGP-1

A *Bsr*GI restriction site was introduced to the 5′-end of *LeAGP-1c* (Li and Showalter, 1996a) immediately after the signal sequence and an *Eag*I restriction site introduced to the 3′-end of *LeAGP-1c* via PCR amplification (Figure 2). The 5′ sense primer containing the *Bsr*GI site (underlined) was:

5′-GCA ATG TAC ACG GGT CAG ACA CCT GCC GCA GCA CCC GTT G-3′

The 3′-antisense primer contained the *Eag*I site (underlined):

5′-GAC GTC ACG GCC GCT TTA GAA CAA GAG CCA GCT CAT CAC AGC-3′

The amplified *LeAGP-1* fragment was subcloned into *pUC-SS^{tom}-EGFP* immediately following *EGFP* as a *Bsr*GI/*Eag*I fragment.

pUC-SS^{tom}-EGFP-LeAGP-1ΔK

First, *pUC-SS^{tom}-EGFP-LeAGP-1* was digested with *Bsr*GI and *Nsi*I to remove the 5′-end of *LeAGP-1* through the lysine-rich region. We discarded the smaller *Bsr*GI/*Nsi*I fragment encoding the N-terminus of LeAGP-1 and the lysine-rich region, but kept the larger *pUC-SS^{tom}-EGFP* plasmid fragment which still retained the nucleotides encoding the C-terminal AGP portion of LeAGP-1, the GPI-anchor addition signal sequence, the signal sequence and EGFP.

Next, PCR amplification of a section of *LeAGP-1c* sandwiched between the signal sequence and the region encoding the lysine-rich region introduced a *Bsr*GI site at the 5′-end of the amplified fragment and an *Nsi*I site at the 3′-end. The sense primer for

amplification was the 5'-sense primer with a *Bsr*GI site (described above) for construction of *pUC-SS^tom-EGFP-LeAGP-1*. The anti-sense primer contained an *Nsi*I restriction site (underlined below):

5'-AC TTT TC<u>A TGC AT</u>T AGG AGC CGG AGC TGG AGT TGT CTC-3'

The resulting *Bsr*GI/*Nsi*I fragment was subcloned into the isolated *Bsr*GI/*Nsi*I *pUC-SS^tom-EGFP* plasmid fragment described above, generating the plasmid *pUC-SS^tom-EGFP-LeAGP1ΔK* where ΔK denotes a *LeAGP-1* clone lacking the lysine-rich region.

PUC-SS^tom-EGFP-LeAGP-1ΔGPI

We constructed *LeAGP-1ΔGPI*, which denotes the clone *LeAGP-1* lacking the putative GPI anchor-signal sequence, by PCR amplification of *LeAGP-1c* using the 5' sense primer with a *Bsr*GI site (described above) for construction of *pUC-SS^tom-EGFP-LeAGP-1* and a 3' antisense primer having an *Eag*I restriction site (underlined):

5'-CAA TTT G<u>CG GCC G</u>CT TTA CTC ATC GTT AAG AGA TGG GCT GGG A-3'

The resulting PCR-amplified fragment was inserted into *pUC-SS^tom-EGFP* as a *Bsr*GI/*Eag*I fragment. Constructs were sequenced at the Ohio University sequencing facility. After sequencing, the constructs were subcloned into the plant vector pBl121 (Clontech Laboratories, Palo Alto, CA) as *Bam*HI–*Sst*I fragments in place of the glucuronidase reporter gene. All constructs were under control of the 35S cauliflower mosaic virus promoter, and can be obtained from the Kieliszewski Laboratory.

Agrobacterium *and tobacco cell transformation and selection of cell lines*

The pBl121-based plasmids containing the *EGFP-LeAGP-1* constructs were delivered into *Agrobacterium tumefaciens* strain LBA4404 by the freeze–thaw method (An *et al.*, 1988), then suspension-cultured tobacco cells (*Nicotiana tabacum* BY-2) were transformed with the *Agrobacterium* as described earlier (McCormick *et al.*, 1986). Transformed tobacco cell lines (three lines of each construction) were selected and maintained as described earlier (Shpak *et al.*, 2001). Callus lines exhibiting the brightest green fluorescence when examined via fluorescence microscopy were chosen for propagation in liquid culture, and lines expressing the highest amount of fusion glycoprotein in the culture medium were further characterized biochemically. The number of transgenes present in each transformed cell was not determined. Transformed cells are available from the Kieliszewski laboratory.

Microscopy

EGFP-fusion protein expression in transformed tobacco cell lines was visualized using a confocal laser-scanning fluorescence microscope equipped with a fluorescein isothiocyanate filter set, as described earlier (Shpak *et al.*, 1999).

Medium collection from transgenic BY2 tobacco cells

Culture medium was concentrated by rotary evaporation, then dialyzed against distilled water and concentrated again by rotary evaporation.

Transgenic tobacco cell medium: fractionation via hydrophobic interaction chromatography

Concentrated tobacco medium in 2 M sodium chloride was fractionated on a Phenyl-Sepharose 6 Fast Flow column (16 × 700 mm, Amersham Pharmacia Biotech) equilibrated in 2 M sodium chloride and eluted stepwise in 1 M sodium chloride and then water, as described earlier (Shpak *et al.*, 2001).

Purification of EGFP-LeAGP-1, EGFP-LeAGP-1ΔK and EGFP-LeAGP-1ΔGPI by reversed-phase chromatography

The fluorescent hydrophobic interaction chromatography fractions were injected onto a Hamilton semi-preparative polymeric reversed-phase column (see above) and the fusion proteins were eluted with a linear gradient of 0–30% buffer B in 10 min, then 30–70% buffer B for 70 min. Deglycosylated fusion proteins were isolated on the analytical PRP-1 with a linear gradient of 0–80% buffer B in 100 min at a flow rate of 0.5 ml min^{-1}.

Anhydrous hydrogen fluoride deglycosylation

Samples were deglycosylated with hydrogen fluoride (HF) as described earlier (Sanger and Lamport, 1983).

Monosaccharide and glycosyl linkage analysis

Neutral sugars were determined by gas chromatography after alditol acetate derivatization (Albersheim *et al.*, 1967), as described earlier (Shpak *et al.*, 1999). Linkage analysis was determined after reduction and esterification of the uronic acid residues (Kim and Carpita, 1992), followed by sample methylation and GC–MS analysis as described earlier (Sims and Bacic, 1995)

Hyp glycoside profiles

Arabinogalactan-proteins (5–10 mg) were hydrolyzed in 0.44 N sodium hydroxide (105°C, 18 h), neutralized, and fractionated as described earlier (Shpak *et al.*, 1999; Shpak *et al.*, 2001).

Uronic acid assay

The uronic acid content of each fusion glycoprotein (100 μg) involved a colorimetric assay (A$_{520}$ nm) using m-hydroxydiphenyl (Sigma Chemical Co., St Louis, MO) (Blumenkrantz and Asboe-Hansen, 1973) and a D-glucuronic acid standard.

Hydroxyproline assay

The hydroxyproline content of samples was assayed as described earlier (Kivirikko and Liesmaa, 1959; Shpak *et al.*, 1999).

Removal of EGFP from the fusion glycoproteins by chymotryptic digestion

Each fusion glycoprotein (≈10 mg, 5 mg ml^{-1} aqueous) was heat-denatured in boiling water for 2 min, cooled, then incubated in an equal volume of freshly prepared 2% (w/v) ammonium bicarbonate containing 5 mM calcium chloride and chymotrypsin (substrate : enzyme ratio 100 : 1 w/w). After overnight digestion at

room temperature, fractionation on a PRP-1 column gave the LeAGP-1 glycoproteins without EGFP for amino acid and sequence analysis, and also for circular dichroism. The chymotrypsin-labile Tyr residue introduced through the *Bsr*GI site is shown in Figure 2.

Amino acid composition and protein sequence analyses

Amino acid compositions and protein sequences were determined at the Michigan State University Macromolecular Facility, Department of Biochemistry, Michigan State University, East Lansing, MI.

Circular dichroism

We determined the circular dichroism spectra of poly L-hydroxyproline (5–20 kDa, Sigma) and the isolated LeAGP-1 glycoprotein modules before and after deglycosylation on a Jasco-715 spectropolarimeter (Jasco Inc., Easton, MD). Spectra were averaged over two scans with a bandwidth of 1 nm, and step resolution was 0.1 nm. All spectra are reported in terms of mean residue ellipticity with the 180–250 nm region using a 1 mm path length. The modules and poly L-hydroxyproline (6.4 μM of each) were dissolved in water.

References

Albersheim, P., Nevins, D.J., English, P.D. and Karr, A. (1967) A method for the analysis of sugars in plant cell-wall polysaccharides by gas liquid chromatography. *Carbohydrate Res.* 5, 340–345.

An, G., Ebert, P.R., Mitra, A. and Ha, S.B. (1988) Binary vectors. In *Plant Molecular Biology Manual.* Dordrecht, the Netherlands: Martinus Nijhoff, pp. 1–19.

Apweiler, R., Hermjakob, H. and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta*, 1473, 4–8.

Bacic, A., Du, H., Stone, B.A. and Clarke, A.E. (1996) Arabinogalactan proteins: a family of cell-surface and extracellular matrix plant proteoglycans. *Essays Biochem.* 31, 91–101.

Blumenkrantz, N. and Asboe-Hansen, G. (1973) New methods for quantitative determination of uronic acids. *Anal. Biochem.* 54, 484–489.

Bolwell, P.G. (1986) Microsomal arabinosylation of polysaccharide and elicitor-induced carbohydrate-binding glycoprotein in French bean. *Phytochemistry*, 25, 1807–1813.

Bowles, D.J. and Northcote, D.H. (1972) The sites of synthesis and transport of extracellular polysaccharides in the root tissues of maize. *Biochem. J.* 130, 1133–1145.

Buer, C.S., Weathers, P.J. and Swartzlander, G.A., Jr (2000) Changes in Hechtian strands in cold-hardened cells measured by optical microsurgery. *Plant Physiol.* 122, 1365–1377.

Camolezi, F.L., Pizauro, J.M., Leone, F.A. and Ciancaglini, P. (1999) A simple laboratory experiment to demonstrate the interaction of proteins bearing glycosylphosphatidylinositol anchors with liposomes. *Biochem. Educ.* 27, 37–40.

Chapman, A., Blervacq, A.S., Vasseur, J. and Hilbert, J.L. (2000) Arabinogalactan proteins in *Cichorium* somatic embryogenesis: effect of beta-glucosyl Yariv reagent and epitope localisation during embryo development. *Planta*, 211, 305–314.

Choi, S.-B., Wang, C., Muench, D.G., Ozawa, K., Franceschi, V.R., Wu, Y. and Okita, T.W. (2000) Messenger RNA targeting of rice seed storage proteins to specific ER subdomains. *Nature*, 405, 765–767.

Cleary, A.L. (2001) Plasma membrane–cell wall connections: roles in mitosis and cytokinesis revealed by plasmolysis of *Tradescantia virginiana* leaf epidermal cells. *Protoplasma*, 215, 21–34.

Ding, L. and Zhu, J. (1997) A role for arabinogalactan-proteins in root epidermal cell expansion. *Planta*, 203, 289–294.

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.

Ferris, P.J., Woessner, J.P., Waffenschmidt, S., Kilz, S., Dress, J. and Goodenough, U.W. (2001) Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. *Biochemistry*, 40, 2978–2987.

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.

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 alata. Carbohydrate Res.* 277, 67–85.

Gao, M. and Showalter, A.M. (2000) Immunolocalization of LeAGP-1, a modular arabinogalactan-protein, reveals its developmentally regulated expression in tomato. *Planta*, 210, 865–874.

Gao, M., Kieliszewski, M.J., Lamport, 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.

Garcia-Gomez, B.L., Campos, F., Hernandez, M. and Covarrubias, A.A. (2000) Two bean cell wall proteins more abundant during water deficit are high in proline and interact with a plasma membrane protein. *Plant J.* 22, 277–288.

Gaspar, Y.M., Johnson, K.L., McKenna, J.A., Bacic, A. and Schultz, C.J. (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding. *Plant Mol. Biol.* 47, 161–176.

Gemmill, T.R. and Trimble, R.B. (1999) Overview of *N*- and *O*-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta*, 1426, 227–237.

Gens, J.S., Fujiki, M. and Pickard, B.G. (2000) Arabinogalactan protein and wall-associated kinase in a plasmalemmal reticulum with specialized vertices. *Protoplasma*, 212, 115–134.

Gilson, P., Gaspar, Y.M., Oxley, D., Youl, J.J. and Bacic, A. (2001)

NaAGP4 is an arabinogalactan protein whose expression is suppressed by wounding and fungal infection in *Nicotiana alata*. *Protoplasma*, **215**, 128–139.

Goodrum, L.J., Patel, A., Leykam, J.F. and Kieliszewksi, M.J. (2000) Gum arabic glycoprotein contains glycomodules of both extensin and arabinogalactan-glycoproteins. *Phytochem.* **54**, 99–106.

Hayashi, T. and Maclachlan, G. (1984) Biosynthesis of pentosyl lipids by pea membranes. *Biochem. J.* **217**, 719–803.

Hecht, K. (1912) Studien uber den Vorgang der Plasmophysiologie der Grenzschichten lebender Pflanzenzellen. *Cohns Beitrage Zur Biologie Pflanzen*, **11**, 137–189.

van Holst, G.-J. and Varner, J.E. (1984) Reinforced polyproline II conformation in a hydroxyproline-rich glycoprotein from carrot root. *Plant Physiol.* **74**, 247–251.

Homer, R.B. and Roberts, K. (1979) Glycoprotein conformation in plant cell walls: circular dichroism reveals a polyproline II structure. *Planta*, **146**, 217–222.

Jermyn, M.A. and Yeow, Y.M. (1975) A class of lectins present in the tissues of seed plants. *Aust. J. Plant Physiol.* **2**, 501–531.

Kieliszewski, M., de Zacks, R., Leykam, J.F. and Lamport, D.T.A. (1992) A repetitive proline-rich protein from the gymnosperm Douglas Fir is a hydroxyproline-rich glycoprotein. *Plant Physiol.* **98**, 919–926.

Kieliszewski, M.J. and Lamport, D.T.A. (1994) Extensin: repetitive motifs, functional sites, post-translational codes and phylogeny. *Plant J.* **5**, 157–172.

Kieliszewski, M.J., O'Neill, M., Leykam, J. and Orlando, R. (1995) Tandem mass spectrometry and structural elucidation of glycopeptides from a hydroxyproline-rich plant cell wall glycoprotein indicate that contiguous hydroxyproline residues are the major sites of hydroxyproline-*O*-arabinosylation. *J. Biol. Chem.* **270**, 2541–2549.

Kim, J.-B. and Carpita, N.C. (1992) Changes in esterification of the uronic acid groups of cell wall polysaccharides during elongation of maize coleoptile. *Plant Physiol.* **98**, 646–653.

Kivirikko, K.I. and Liesmaa, M. (1959) A colorimetric method for determination of hydroxyproline in tissue hydrolysates. *Scand. J. Clin. Lab.* **11**, 128–131.

Komalavilas, P., Zhu, J. and Nothnagel, E.A. (1991) Arabinogalactan-proteins from the suspension culture medium and plasma membrane of rose cells. *J. Biol. Chem.* **266**, 15956–15965.

Labarca, C. and Loewus, F.A. (1972) The nutritional role of pistil exudate in pollen tube wall formation in *Lilium longiflorum*. I. Utilization of injected stigma exudate. *Plant Physiol.* **50**, 7–14.

Labarca, C. and Loewus, F.A. (1973) The nutritional role of pistil exudate in pollen tube wall formation in *Lilium longiflorum*. II. Production and utilization of exudate from the stigma and style canal. *Plant Physiol.* **52**, 87–92.

Lamport, D.T.A. (1964) Cell suspension cultures of higher plants, isolation and growth energetics. *Exp. Cell Res.* **33**, 195–206.

Lamport, D.T.A. (1965) The protein component of primary cell walls. *Adv. Bot. Res.* **2**, 151–218.

Lamport, D.T.A. (1967) Hydroxyproline-*O*-glycosidic linkage of the plant cell wall glycoprotein extensin. *Nature*, **216**, 1322–1324.

Lamport, D.T.A. (1970) Cell wall metabolism. *Ann. Rev. Plant Physiol.* **21**, 235–270.

Lamport, D.T.A. (1984) Glycosylated hydroxyproline. *Meth. Enzymol.* **106**, 523–528.

Larkin, P.J. (1978) Plant protoplast agglutination by artificial carbohydrate antigens. *J. Cell Sci.* **30**, 283–292.

Li, S. and Showalter, A.M. (1996a) Cloning and developmental/

stress regulated expression of a gene encoding a tomato arabinogalactan protein. *Plant Mol. Biol.* **32**, 641–652.

Li, S. and Showalter, A.M. (1996b) Immunolocalization of extensin and potato tuber lectin in carrot, tomato and potato. *Physiol. Plant.* **97**, 708–718.

Liljas, A. (1999) Structure is function. *Science*, **285**, 2077–2078.

McCormick, S., Niedermeyer, J., Fry, J., Barnason, A., Horsch, R. and Fraley, R. (1986) Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*. *Plant Cell Rep.* **5**, 81–84.

Nam, J., Mysore, K.S., Zheng, C., Knue, M.K., Matthysse, 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.

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.

Pope, D.G. (1977) Relationships between hydroxyproline-containing proteins secreted into the cell wall and medium by suspension-cultured *Acer pseudoplatanus* cells. *Plant Physiol.* **59**, 894–900.

Qi, W., Fong, C. and Lamport, D.T.A. (1991) Gum arabic glycoprotein is a twisted hairy rope: a new model based on *O*-galactosylhydroxyproline as the polysaccharide attachment site. *Plant Physiol.* **96**, 848–855.

Rees, D.A. (1977) *Polysaccharide Shapes*. London, UK: Chapman & Hall, pp. 41–62.

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.

Ruggiero, F., Exposito, J.Y., Bournat, P., Gruber, V., Perret, S., Comte, J., Olagnier, B., Garrone, R. and Theisen, M. (2000) Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. *FEBS Lett.* **469**, 132–136.

Sanger, M.P. and Lamport, D.T.A. (1983) A micro-apparatus for liquid hydrogen fluoride solvolysis: sugar and amino sugar composition of *Erysiphe graminis* and *Triticum aestivum* cell walls. *Anal. Biochem.* **128**, 66–70.

Schultz, C.J., Johnson, K.L., Currie, G. and Bacic, A. (2000) The classical arabinogalactan protein gene family of Arabidopsis. *Plant Cell*, **12**, 1751–1767.

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) Glyco-sylphosphatidylinositol-anchored cell-surface proteins from Arabidopsis. *Electrophoresis*, **20**, 2027–2035.

Shpak, E., Leykam, J.F. and Kieliszewski, M.J. (1999) Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. *Proc. Natl Acad. Sci. USA*, **96**, 14736–14741.

Shpak, E., Barbar, E., Leykam, J.F. and Kieliszewski, M.J. (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. *J. Biol. Chem.* **276**, 11272–11278.

Sims, I.M. and Bacic, A. (1995) Extracellular polysaccharides from suspension cultures of *Nicotiana plumbaginifolia*. *Phytochem.* **38**, 1397–1405.

Strahl-Bolsinger, S., Gentzsch, M. and Tanner, W. (1999) Protein *O*-mannosylation. *Biochim. Biophys. Acta*, **1426**, 297–307.

Svetek, J., Yadav, M.P. and Nothnagel, E.A. (1999) Presence of a

glycosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. *J. Biol. Chem.* **274**, 14724–14733.

Takos, A.M., Dry, I.B. and Soole, K.L. (2000) Glycosyl-phosphatidylinositol-anchor addition signals are processed in *Nicotiana tabacum*. *Plant J.* **21**, 43–52.

Vissenberg, K., Feijo, J.A., Weisenseel, M.H. and Verbelen, J.P. (2001) Ion fluxes, auxin and the induction of elongation growth in *Nicotiana tabacum* cells. *J. Exp. Bot.* **52**, 2161–2167.

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 β-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. *Plant J.* **9**, 919–925.

Wu, H., Wang, H. and Cheung, A.Y. (1995) A pollen tube growth stimulatory glycoprotein is deglycosylated by pollen tubes and displays a glycosylation gradient in the flower. *Cell*, **82**, 395–403.

Yariv, J., Rapport, 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 alata* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. *Proc. Natl Acad. Sci. USA*, **95**, 7921–7926.

Zhu, J.-K., Bressan, R.A. and Hasegawa, P.M. (1993) Loss of arabinogalactan-proteins from the plasma membrane of NaCl-adapted tobacco cells. *Planta*, **190**, 221–226.