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

Zhan Dong Zhao, Li Tan, Allan Marshall Showalter, Derek Thomas Anthony Lamport and Marcia Jane Kieliszewski

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

Received 12 February 2002; revised 9 April 2002; 23 April 2002.

*For correspondence (fax +1 740 597 1772; e-mail kielisz@helios.phy.ohiou.edu).

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, glycosylphosphatidylinositol 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 glycosubstituents 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 pI ranging from 4 to 7, and a minor AGP fraction with a pI of 2 (data not shown). LeAGP-1 occurred in the pI 4–7 fraction, judging by immunoassays of the recovered fractions.

Further fractionation of the focused AGPs (pI 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 immuno-reactivity, 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).

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
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\textsuperscript{3}EGFP-LeAGP-1, pUC-SS\textsuperscript{3}EGFP-LeAGP-1\textsuperscript{K}, and pUC-SS\textsuperscript{3}EGFP-LeAGP-1\textsuperscript{GPI}. The chymotrypsin-labile Tyr residue engineered into the fusion glycoprotein to allow removal of EGFP from LeAGP-1 is encoded in the BsrGI 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.
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 Bsrl 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 sequence, TGQTOAAXVGAKATTOOAAO (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).
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.

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

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Tobacco LeAGP-1αK</th>
<th>Tobacco LeAGP-1αGPI</th>
<th>Tobacco LeAGP-1</th>
<th>Tomato LeAGP-1</th>
<th>Tomato LeAGP-1c DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyp</td>
<td>24</td>
<td>25</td>
<td>24</td>
<td>29</td>
<td>–</td>
</tr>
<tr>
<td>Asx</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>Thr</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8.8</td>
</tr>
<tr>
<td>Ser</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>11.7</td>
</tr>
<tr>
<td>Gix</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>Pro</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>27.5</td>
</tr>
<tr>
<td>Gly</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3.5</td>
</tr>
<tr>
<td>Ala</td>
<td>24</td>
<td>22</td>
<td>26</td>
<td>21</td>
<td>23.4</td>
</tr>
<tr>
<td>Val</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8.2</td>
</tr>
<tr>
<td>Met</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1.7</td>
</tr>
<tr>
<td>Ile</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>Leu</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>Tyr</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Phe</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>His</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>Lys</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6.4</td>
</tr>
<tr>
<td>Arg</td>
<td>1</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Trp</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The composition of transgenic LeAGP-1 shows less Lys than predicted, while LeAGP-1αK 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-1αK showed no reaction. EGFP was removed from the fusion glycoproteins before analysis.

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

<table>
<thead>
<tr>
<th>Glycosyl residue</th>
<th>LeAGP-1</th>
<th>LeAGP-1αK</th>
<th>LeAGP-1αGPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rha</td>
<td>8</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Ara</td>
<td>39</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>Gal</td>
<td>37</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>GlcUA</td>
<td>12</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Man</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

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).
The percentage of Hyp-PS predicted is the maximum assuming that each non-contiguous Hyp has a polysaccharide adduct.

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 arabinoxylan polysaccharide. A 100% addition of arabinoxylan 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 arabinoxylan polysaccharide addition (60% of the total Hyp) according to the Hyp contiguity hypoth-

---

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.

<table>
<thead>
<tr>
<th>Hyp glycoside</th>
<th>Percentage of total hydroxyproline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted</td>
</tr>
<tr>
<td>Hyp-PS</td>
<td>58-60</td>
</tr>
<tr>
<td>Hyp-Ara 4</td>
<td>3</td>
</tr>
<tr>
<td>Hyp-Ara 3</td>
<td>11</td>
</tr>
<tr>
<td>Hyp-Ara 2</td>
<td>28</td>
</tr>
<tr>
<td>Hyp-Ara</td>
<td>8</td>
</tr>
<tr>
<td>NG-Hyp</td>
<td>12-14</td>
</tr>
</tbody>
</table>

*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 et al., 1999; J. Xu, Z.D.Z. and M.J.K., unpublished results). Transformed cells exhibited a uniform peripheral green fluorescence when turgid; on plasmolysis the fluorescence remained largely with the retracted protoplast, indicating fluorescence when turgid; on plasmolysis the fluorescence and Showalter, 2000; Li and Showalter, 1996a).

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 arabinoxylan-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 specificities.

---

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

**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 arabinosid es. 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-poly saccharide 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.
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 stigma 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-Hyp₄ glycomodule, stabilize the polysaccharide 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 VOAOVTAO-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 glycalyx 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 (Jermy 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 Shenh-Hildebrandt medium at 27°C 15% 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 diazylated 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 × 700 mm; Amersham Pharmacia Biotech, Piscataway, NJ) equilibrated in 200 mM sodium phosphate buffer (pH 7) and eluted at a flow rate of 1.2 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, 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. Oligosacrylated 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.

Immuoassays

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™-EGFP

We amplified the tomato LeAGP-1 signal sequence using the cDNA LeAGP-1c (Li and Showalter, 1996a) and primers designed to introduce BamHI and Xmal 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 BamHI site (underlined): 5’-CTC TTT TTC TCT GGA TCC GTT CTA TAT TTT CT TTA GC-3’ and the antisense primer introduced an Xmal site (underlined): 5’-C GG GTC TGC CCC GGG TGC TGT CTT ACC GTG GAC ACT TGC-3’ (Figure 2). The isolated fragment was subcloned into plasmid pUC-SS™-EGFP (Shpak et al., 1999) as a BamHI/Xmal fragment in place of the tomato extensin signal sequence (SS™). We designated this new pUC19-derived plasmid pUC-SS™-EGFP.

Creation of plasmids pUC-SS™-EGFP-LeAGP-1, pUC-SS™-EGFP-LeAGP-1-1K and pUC-SS™-EGFP-LeAGP-1-1GI

pUC-SS™-EGFP-LeAGP-1

A BsrGI restriction site was introduced to the 5'-end of LeAGP-1c (Li and Showalter, 1996a) immediately after the signal sequence and an EagI restriction site introduced to the 3'-end of LeAGP-1c via PCR amplification (Figure 2). The 5’ sense primer containing the BsrGI 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 EagI 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™-EGFP immediately following EGFP as a BsrGI/EagI fragment.

pUC-SS™-EGFP-LeAGP-1-1K

First, pUC-SS™-EGFP-LeAGP-1 was digested with BsrGI and NsiI to remove the 5’-end of LeAGP-1 through the lysine-rich region. We discarded the smaller BsrGI/NsiI fragment encoding the N-terminus of LeAGP-1 and the lysine-rich region, but kept the larger pUC-SS™-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 BsrGI site at the 5’-end of the amplified fragment and an NsiI site at the 3’-end. The sense primer for
amplification was the 5'-sense primer with a BsrGI site (described above) for construction of pUC-SS<sup>com</sup>-EGFP-LeAGP-1. The anti-sense primer contained an NsiI restriction site (underlined below):

5'-AC TTT TCA TGC ATT AGG AGC CGG AGC TGG AGT TGT CTC-3'

The resulting BsrGI/NsiI fragment was subcloned into the isolated BsrGI/NsiI pUC-SS<sup>com</sup>-EGFP plasmid fragment described above, generating the plasmid pUC-SS<sup>com</sup>-EGFP-LeAGP1JK where ΔK denotes a LeAGP-1 clone lacking the lysine-rich region.

**PUC-SS<sup>com</sup>-EGFP-LeAGP-1ΔK**

We constructed LeAGP-1ΔKGI, 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 BsrGI site (described above) for construction of pUC-SS<sup>com</sup>-EGFP-LeAGP-1 and a 3' antisense primer having an EagI restriction site (underlined):

5'-CAA TTT GCG GCC GCT TTA CTC ATC GTT AAG AGA TGG GCT GGG A-3'

The resulting PCR-amplified fragment was inserted into pUC-SS<sup>com</sup>-EGFP as a BsrGI/EagI 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 BamHI-SstI fragments in place of the glucononidase 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-1ΔKGI 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<sup>-1</sup>.

**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<sub>520</sub> 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<sup>-1</sup> 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 BsrGl 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.

Acknowledgements

This work was supported by grants from the National Science Foundation (IPB-9727757, IBN-0110413, MCB-9874744), the Ohio Department of Biochemistry, Michigan State University, East Lansing, MI.

References


Gilson, P., Gaspar, Y.M., Oxley, D., Youl, J.J. and Bacic, A. (2001)
NaAGP4 is an arabinoxylan protein whose expression is suppressed by wounding and fungal infection in *Nicotiana alata*. *Protoplasma*, 215, 128–139.


