Structure of a Hydroxyproline (Hyp)-Arabinogalactan Polysaccharide from Repetitive Ala-Hyp Expressed in Transgenic Nicotiana tabacum*

Received for publication, October 29, 2003, and in revised form, January 12, 2004
Published, JBC Papers in Press, January 14, 2004, DOI 10.1074/jbc.M311864200

Li Tan‡, Feng Qiü, Derek T. A. Lamport¶, and Marcia J. Kieliszewski‡

From the ‡Department of Chemistry and Biochemistry, Ohio University, Athens, Ohio 45701, the ¶Department of Chemistry, Oklahoma State University, Stillwater, Oklahoma 74078, and §§School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom

A synthetic gene encoding the fusion protein (Ala-Hyp)51-enhanced green fluorescent protein expressed in Nicotiana tabacum cells produced a fusion glycoprotein with all proline residues hydroxylated and substituted with an arabinogalactan polysaccharide. Alkaline hydrolysis of the fusion glycoprotein yielded a population of hydroxypoline (Hyp)-arabinogalactan polysaccharides ranging in size from 13 to 26 saccharide residues/Hyp, with a median size of 15–17 residues. We isolated a 15-residue Hyp-arabinogalactan for structure determination by sugar analyses and one- and two-dimensional nuclear magnetic resonance techniques that provided the assignment of proton and carbon signals of a small polysaccharide structure having a highly branched backbone of 3-, 6-, and 3,6-linked D-Gal residues and also a 1,6-linked β-D-Galp “kink.” The backbone had two side chains of Galp substituted at position 3 with an arabinose di- or trisaccharide and at position 6 with glucuronic acid or rhamnosyl glucuronic acid. Energy-minimized space-filling molecular models showed hydrogen bonding within polysaccharides associated to repetitive Ala-Hyp and also between polysaccharides and the peptide backbone. Polysaccharides distorted the peptide Ramachandran angles consistent with the circular dichroic spectra of isolated (Ala-Hyp)51 and its reversion to a polypolyproline II-like helix after deglycosylation. This first complete structure of a Hyp-arabinogalactan polysaccharide shows that computer-based molecular modeling of Hyp-rich glycoproteins is now feasible and supports the suggestion that small repetitive subunits comprise larger arabinogalactan polysaccharides.

Arabinogalactan proteins (AGPs)1 expressed at the plant cell surface comprise a multigene family of hydroxyproline-rich glycoproteins (HRGPs) broadly implicated in all aspects of plant growth and development from fertilization to apoptosis (1, 2).

Compared with other HRGPs like the extensins that contain highly repetitive motifs based on contiguous Hyp residues, AGPs show a lower peptide periodicity based on clustered non-contiguous Hyp residues. Although short O-linked arabinoside substituents decorate contiguous Hyp residues of both extensins and AGPs, the larger arabinogalactans attached to clustered non-contiguous Hyp define the hyperglycosylated AGPs (90–95% carbohydrate) (3, 4). These differences in glycosylation reflect different networks, cellular locations, and biological functions. Thus cross-linked extensins form covalent networks in muro, whereas noncovalently associated AGPs (5, 6) occur primarily at the periplasmic interface where a glycosylphosphatidylinositol lipid initially anchors them to the plasma membrane (7–10).

The precise role of AGPs, both soluble and anchored, remains unclear. However, AGPs readily associate (5, 11); furthermore, β-glycosyl Yariv compounds interact specifically with AGPs and also inhibit cell expansion (12, 13). These observations imply that multiple weak interactions between the polysaccharide components are essential to the role of AGPs (8–10). The arabinogalactan substituents occur in small clusters along the polypeptide backbone forming glycomodules that are highly conserved (14–16) and hence of functional significance, no doubt involving specific interactions. Thus, it is of paramount interest to elucidate the size, structure, arrangement, shape, and binding partners of these glycomodules in more detail.

AGPs show a lower peptide periodicity based on clustered non-contiguous Hyp residues. Although short O-linked arabinoside substituents decorate contiguous Hyp residues of both extensins and AGPs, the larger arabinogalactans attached to clustered non-contiguous Hyp define the hyperglycosylated AGPs (90–95% carbohydrate) (3, 4). These differences in glycosylation reflect different networks, cellular locations, and biological functions. Thus cross-linked extensins form covalent networks in muro, whereas noncovalently associated AGPs (5, 6) occur primarily at the periplasmic interface where a glycosylphosphatidylinositol lipid initially anchors them to the plasma membrane (7–10).

The precise role of AGPs, both soluble and anchored, remains unclear. However, AGPs readily associate (5, 11); furthermore, β-glycosyl Yariv compounds interact specifically with AGPs and also inhibit cell expansion (12, 13). These observations imply that multiple weak interactions between the polysaccharide components are essential to the role of AGPs (8–10). The arabinogalactan substituents occur in small clusters along the polypeptide backbone forming glycomodules that are highly conserved (14–16) and hence of functional significance, no doubt involving specific interactions. Thus, it is of paramount interest to elucidate the size, structure, arrangement, shape, and binding partners of these glycomodules in more detail.

AGP arabinogalactan polysaccharides typically consist of short side chains containing arabinose and often rhamnose (Rha) and glucuronic acid (GlcUA) attached to a β-(1→3)-linked galactan backbone core (11). Based on 13C NMR spectra of crude gum arabic degradation products, Defaye and Wong (17) deduced a complete linkage analysis of a short pentameric side chain containing Ara, galactose (Gal), Rha, and GlcUA, and they proposed that the side chain is a major component of gum arabic arabinogalactan polysaccharides. Likewise, Akiyama and Kato (18) suggested a similar structure occurred in the AGPs of Nicotiana tabacum suspension-cultured cells. More recently, Gane applied one-dimensional and two-dimensional homonuclear and heteronuclear NMR techniques to characterize a mixture of AGPs from Nicotiana alata and described a structure having a highly branched backbone of 3-, 6-, and 3,6-linked Galp with terminal Galp and Ara side chains (11). However, complete structural elucidation of an entire AGP polysaccharide remained a formidable problem.

Recently, we introduced the use of synthetic genes to express single repetitive glycopeptide motifs of AGPs and extensins which effectively amplifies glycomodules of interest and simplifies their isolation (14, 15, 19). AGPs often have short Ala-Hyp repeats. Synthetic gene constructs of these repeats gave an (Ala-Hyp)51 expression product with all the Hyp residues glycosylated by an arabinogalactan polysaccharide (19). This
novel AGP was therefore suitable for isolating Hyp-arabinogalactan polysaccharides, first described in sycamore-maple (20) and tomato cells (21). Here we report the first complete structural analysis of an arabinogalactan polysaccharide, including a space-filling model that depicts possible glycan interactions.

EXPERIMENTAL PROCEDURES

Isolation of the [Ala-Hyp]51-EGFP Fusion Protein and ([Ala-Hyp]51)2—The ([Ala-Hyp]51)-EGFP fusion glycoprotein was isolated from spent medium of 20-day transformed N. tabacum BY-2 suspension cultured cells by a combination of hydrophobic interaction and reversed phase chromatography, as described earlier (19). ([Ala-Hyp]51) was isolated from tryptic digests of ([Ala-Hyp]51)-EGFP by gel permeation chromatography, also described earlier (19).

Anhydrous Hydrogen Fluoride (HF) Deglycosylation—We deglycosylated 3–5 mg of ([Ala-Hyp]51) for 1 h at 4 °C as described earlier (3).

Base Hydrolysis of ([Ala-Hyp]51)-EGFP—Sixty-six milligrams of ([Ala-Hyp]51)-EGFP was dissolved in 9 ml of 0.44 N NaOH and hydrolyzed at 105 °C for 18 h. The hydrolysate was chilled on ice and neutralized with 1 N HCI to pH 7.6, followed by freeze-drying.

Hyp-Arabinogalactan Glycoamino Acid Purification—The neutralized, dry base hydrolysate of ([Ala-Hyp]51)-EGFP was dissolved in deionized, distilled H2O and applied to a 50 × 6.8-cm column (H+ form) of Technicon Chromobeads C washed and eluted with deionized, distilled H2O. The water eluate contained the Hyp-polysaccharides that were pooled and freeze-dried before further purification on an analytical Superdex-peptide gel filtration column (Amersham Biosciences). The Superdex-peptide column was equilibrated and isocratically eluted in 20% acetonitrile (aqueous) at a flow rate of 0.3 ml/min. One hundred microliters of each fraction (fraction size was 0.9 ml) were assayed for Hyp as described earlier (20). Superdex fractions representing the median arabinogalactan size were combined and refractionated on the Superdex-peptide column using the same conditions described above, except the fraction volumes were smaller (450 μl/1.5 min). Fractions containing the Hyp-glycoamino acids were individually freeze-dried. We chose the fraction with the most material, designated AHP-1, for further analyses.

Sugar Composition and Linkage Analyses—The monosaccharide composition of AHP-1 was analyzed as alditol acetate derivatives by gas chromatography (22) or by gas chromatography-mass spectrometry of the per-O-trimethylsilyl derivatives (23, 24). Uronic acids were also estimated colorimetrically (25). For linkage analysis, samples were permethylated, depolymerized, reduced, and acetylated resulting in partially methylated alditol acetates that were analyzed by gas chromatography-mass spectrometry as described earlier (24).

Glycan Size Estimation—The glycan size was determined by dividing the masses of monosaccharide in a known weight of AHP-1 or ([Ala-Hyp]51)-EGFP by the moles of Hyp present. Hyp amount was determined colorimetrically (20) and the monosaccharide content by gas chromatography as described above.

Circular Dichroism—CD spectra of standard poly-l-hydroxyproline (5–20 kDa), Sigma) and purified ([Ala-Hyp]51) and deglycosylated ([Ala-Hyp]51) were recorded on a Jasco-715 spectropolarimeter (Jasco Inc., Easton, MD). Spectra were averaged over two scans with a bandwidth of 1 nm and step resolution of 0.5 nm. All spectra were reported in terms of mean residue ellipticity with the 190–250-nm region using a 1-mm path length. Samples were dissolved in water at a concentration of 18 μg.

NMR Experiments—All one- and two-dimensional NMR experiments were carried out on AHP-1 (~2.5 mg AHP-1/ml D2O) with Varian Unity INOVA 600-MHz spectrometer operating at 1H and 13C frequencies of 598.621 and 150.536 MHz, respectively. Analyses were conducted at 60 °C, as at this temperature the residual H2O signal overlapped with the fewest resonances and the line widths had narrowed. The only exception was a NOESY experiment that was carried out at 30 °C to slow down molecular motion and increase the number of cross-peaks. Quadrature detection in the indirect dimension was achieved by the hypercomplex method (26) for both homonuclear and heteronuclear experiments. HMBC experiments were typically conducted with spectral width of 8 kHz for 1H and 30 kHz for 13C, respectively, a JCH set to 150 Hz, a relaxation delay of 1.5 s, the delay for minimizing signals from 1H-H cross peaks to 0.5 s, and WURST decoupling (27) on the 13C channel. Data were collected as an array of 2 × 128K point, which after linear prediction in t1 dimension and zero filling in both dimensions produced a 2 × 1K data matrix. For HMBC experiments, the acquisition conditions were similar to those used for HMQCC spectra but with no decoupling in the 13C channel. Data set consisting of 4 × 1K data points was acquired with multibounds JCH set to 8 Hz. DQCOFSY spectra were collected with spectral width of 8 kHz in both dimensions. A relaxation delay of 2 s was used. The TOCSY mixing time was 90 ms. The NOESY spectra were recorded with mixing time of 250 ms. Data were processed by using NMRPipe (28) and analyzed by using nmrview (29).

Molecular Modeling Software—We used HyperChem version 7.0 running on a fast dual-processor Pentium machine (plus 1GB RAM) to build and energy-minimize the ([Ala-Pro]51)-glycosylated and non-glycosylated peptide via the MM+ force field and steepest descent algorithm. The HyperChem Sugar Builder module allowed the facile construction of small oligosaccharides that were then energy-minimized and ligated to form the small polysaccharide AHP-1 followed by final energy-minimization again using the Amber3 force field. (We corrected a software error in the anomeric configuration of 5-sugars by inserting the α- for β-anomer and vice versa in order to follow IUPAC definitions.) Three AHP-1 polysaccharides were then attached glycosidically to ([Ala-Pro]51) to form a tight cluster of three contiguous arabinogalactans on non-contiguous Hyp residues and then subjected to final energy minimization using the Amber3 force field. We found it helpful to define the approximate bond lengths for the Gal-Hyp glycosidic link by using the “set bond length” command when attaching the polysaccharide to the polypeptide.

RESULTS AND DISCUSSION

Isolation of Hyp-Arabinogalactan Polysaccharide (AHP-1) from (Ala-Hyp)51-EGFP

Earlier work showed that ([Ala-Hyp]51)-EGFP, isolated as described earlier (19), contained 51-O-Hyp-linked arabinogalactan polysaccharides containing Gal, Ara, GlcUA, and Rha. Base hydrolysis of ([Ala-Hyp]51)-EGFP released Hyp-arabinogalactans that voided a Chromobeads cation exchange column (Fig. 1A). Further fractionation on a Superdex-peptide gel filtration...
column yielded a major peak containing small Hyp-arabinogalactan polysaccharides ranging in size from about 13–26 monosaccharide residues judging by the molar ratio of monosaccharide to Hyp in each fraction (Fig. 1B). Fractions 11 and 12 from the column contained the most material (2.8 mg), judging by the recovered weights, and therefore were chosen for refractionation on the Superdex column. Refractionation of 11 and 12 produced a single major peak (Fig. 1C) the tip of which we collected for structural analyses. The peak was designated Ala-Hyp polysaccharide-1 (AHP-1). Monosaccharide analyses of isolated AHP-1 indicated it was rich in Gal and Ara (50 and 33 mol %, respectively) with lesser amounts of Rha and GlcUA (4 and 14 mol %, respectively), yielding the following molar ratios: 7 Gal:5 Ara:2 GlcUA:0.5 Rha. This was consistent with the estimated size and composition of the isolated oligosaccharide and with the NMR data below, albeit with an underestimate of Rha. The isolated AHP-1 glycan size was also consistent with that deduced from the sugar composition of intact (Ala-Hyp)51-EGFP, which also showed a ratio of 15.1 sugar residues per Hyp. This indicates that Hyp-polysaccharides (Ala-Hyp)51-EGFP, which also showed a ratio of 15.1 sugar residues per Hyp. This indicates that Hyp-polysaccharides were released, but not degraded, by alkaline hydrolysis.

Consistent with the sugar composition, the linkage analysis of AHP-1 (Table I) showed mainly 3,6-linked Gal with lesser amounts of terminal, 3- and 6-linked Gal residues were either terminal or 4-linked. However, the linkage analyses (repeated three times) were not strictly consistent with the estimated linkage ratios. Alkaline hydrolysis of AHP-1 by 1H NMR spectroscopy. Spectra are shown in Figs. 2—4, and the assignments are reported in Tables II and III. The 1H assignments were obtained from a one-dimensional proton spectrum (Fig. 2), from homonuclear COSY, TOCSY, ROESY (Fig. 4), and NOESY two-dimensional spectra and a heteronuclear HMQC two-dimensional spectrum (not shown). The 13C assignments were established from a combination of two-dimensional HMQC (not shown) and HMBC heteronuclear spectra (Fig. 3) and proton decoupled one-dimensional and DEPT 13C spectra.

**AHP-1 Sugar Molar Ratios and Anomeric Configuration**

A one-dimensional 1H NMR spectrum determined the number of saccharide anemic protons with aid from the glycosyl composition analyses. The spectrum (Fig. 2) showed resonances at 5.25, 5.09, 4.79, 4.77, 4.70, 4.57, and −4.50 ppm in a ratio of 4:1:1:4:1:4:1:4 determined by integrating areas in the one-dimensional 1H NMR spectrum. We assigned the resonances based on known chemical shifts (11, 30–33) and the AHP-1 composition as follows: signals at 5.25 and 5.09 ppm corresponded to H-1 of four and one α-L-Ara residues, respectively; peak C to H-1 of a single α-L-Rha residue; peak D to H-4 of the AHP-1 Hyp residue; peak E to H-1 of four β-D-Gal residues (the galactan main chain); peak F to H-1 of a single β-D-Gal residue (G, the Gal linked to Hyp); and peak G to H-1 of two β-D-Gal residues and two β-D-GlcUA residues.

**Table I**

| Deduced glycosidic linkage | Mol | Estimated molar amounts |
|---------------------------|-----|------------------------|
| Rha                       | Terminal 3 | 0.5 |
| GlcUA                     | Terminal 5 | 2.0 |
| Ara                        | 4-1 8 | |
| Ara                        | Terminal 7 | 2.9 |
| Gal                        | 3- 9 | |
| GlcUA                     | 5- 3 | |
| Gal                        | Terminal 4 | 10.0 |
| GlcUA                     | 3- 14 | |
| GlcUA                     | 4- 4 | |
| GlcUA                     | 3,6- 45 | |

* Estimated assuming there are 2 mol of GlcUA/AHP-1.
A second set of weak cross-peaks in the HMQC spectrum occurred indicating the Hyp residues were a mixture of L-Hyp and allo-Hyp isomers formed during base hydrolysis of the polypeptide backbone (35, 36) (Table III, set 2).

A cross-peak in the HMBC spectrum (Fig. 3, cross-peak I) indicated that the GlcUA residues were linked to position 6 of each main chain Gal residue, G1 and G0 (GlcUA H-1/Ga/b C-6). The two cross-peaks G indicated Gal residues G4 and G5 were linked to position 6 of main chain Gal residues, G1 and G0 (G4, G5 H-1/Ga/b C-6). The two cross-peaks H and three J identified the Gal-(1→3)-Gal linkages, as they arose from correlations between Gal H-1 (Ga/Gb) and Gal C-3 (Ga/Gb, C-3 83.0 ppm) and between Gal C-1 (Ga/Gb) and Gal H-3 (Ga/Gb, H-3 at 3.88 and 3.82 ppm), and Gal-Ga (H-1 at 4.71 ppm/C-3 at 83.0 ppm; C-1 at 105 ppm/H-3 at 3.86 and 3.82 ppm). Finally, cross-peak I identified the Gal G0-(1→4)-O-Hyp linkage.
spectrum, we estimated AHP-1 had two terminal Ara residues, one 5-linked Ara residue and two 3-linked Ara residues.

**There Are Two Ara Side Chains and They Are 3-Linked to Gal**—The HMBC spectrum contained two cross-peaks corresponding to C-1 of Ara (110.3 ppm) and H-3 of Gal (3.70 and 3.68 ppm), thus identifying two α-L-Ara-(1→3)-Gal glycosidic linkages and the likelihood of two Ara side chains (Fig. 3, cross-peak D). This conclusion was supported by the occurrence of two distinct terminal Ara residues, discussed below.

The Ara residue designated A1 (Table II) displayed chemical shifts typical of terminal Ara residues (11, 17, 36, 37). Residue A1 is linked to O-5 of Ara A2, judging by 1HCH correlations between the H-1 signals of A1 (5.09 ppm) and the 13C signal at 81.5 ppm corresponding to C-3 of Ara A2 or C1 of Table II (11, 17, 30), the shift at 81.5 ppm, labeled C in Fig. 3, corresponded to C-3 of Gal (11, 17, 30, 32) having a substituent at O-3 corroborating the α-L-Ara-(1→3)-Gal linkages identified above (Gc and Gg in Table II); the signal at 85.3 ppm, corresponding to C-4 of Ara (Table II, A2, A3, and C1) (11, 17, 30). The remaining signal at 83.4 ppm, labeled D in Fig. 3, corresponded to C-3 of Ara having a substituent at O-3 (32, 33, 36) and indicated the presence of α-L-Ara-(1→3)-α-L-Ara linkages. These results considered together with the one-dimensional 1H NMR spectrum (Fig. 3) and the glycosyl linkage composition of the Ara residues (Table I) suggested the following structural units: one α-L-Ara-(1→5)-α-L-Araf, two α-L-Ara-(1→3)-α-L-Araf, and two α-L-Araf-(1→3)-β-D-Galp.

The occurrence of two α-L-Ara-(1→3)-Gal linkages and hence two oligoarabinosyl chains were evident in the HMBC spectrum where cross-peaks (labeled D in Fig. 3) indicated Ara C1 to Galp H-3 correlations. We concluded the five Ara residues form two units linked to O-3 of Gal, probably α-L-Ara-(1→5)-α-L-Araf-(1→3)-α-L-Araf-(1→3)-α-L-Araf-(1→3)-Gal (A1, A2, A3, and C1) in Fig. 5) and α-L-Araf-(1→3)-α-L-Araf-(1→3)-α-L-Araf-(1→3)-Gal (C1, C2, and C3 in Fig. 5). However, the possibility that the two chains have the sequences: α-L-Araf-(1→5)-α-L-Araf-(1→3)-Gal and α-L-Araf-(1→3)-α-L-Araf-(1→3)-α-L-Araf-(1→3)-Gal or α-L-Araf-(1→5)-α-L-Araf-(1→3)-α-L-Araf-(1→3)-α-L-Araf-(1→3)-Gal and α-L-Araf-(1→3)-Gal cannot be ruled out but seem less likely given the similarity of the chemical environments of the two chains.

**Identification of α-L-Rhap-(1→4)-β-D-GlcUA-(1→6)-β-D-Galp and β-D-GlcUA-(1→6)-β-D-Galp—TOCSY, COSY, and HMBC spectra** provided a full assignment of the 1H signals of the α-L-Rhap residue and the β-D-GlepUA residues in AHP-1 (Table II). The corresponding 13C chemical shifts (Table II), assigned with the aid of the one-dimensional 13C and HMBC spectra, were similar to those reported by Defaye and Wong (17) except we did not find a signal at 175.6 ppm corresponding to C-6 of β-D-GlepUA. Cross-peaks in the HMBC spectra arising from H-1 of Rhap and C-4 of GlepUA and from C-1 of Rhap and H-4 of GlepUA indicated that Rhap was linked to GlepUA through

---

**Table I. Summary of 13C and 1H chemical shifts for AHP-1**

| Glycosidic linkage | Residue<sup>a</sup> | Chemical shifts  |
|-------------------|-----------------|-----------------|
|                   | Residue<sup>b</sup> | C-1/H-1 | C-2/H-2 | C-3/H-3 | C-4/H-4 | C-5/H-5a,b | C-6/H-6a,b |
| Terminal α-L-Ara  | A1              | 108.7/5.09   | 82.6/4.23 | 77.8/9.55 | 85.2/4.12 | 62.9/3.71 |
| 5-α-L-Ara         | A2              | 110.3/5.25   | 82.4/4.22 | 78.0/9.34 | 85.3/4.13 | 68.0/3.88 |
| 3-α-L-Ara         | A3              | 110.3/5.25   | 82.4/4.22 | 83.4/4.24 | 85.4/3.09 | 62.5/3.70 |
| Terminal α-L-Ara  | C1              | 110.3/5.25   | 82.4/4.22 | 77.8/9.55 | 85.2/4.13 | 62.5/3.71 |
| 3-α-L-Ara         | C2              | 110.3/5.25   | 82.4/4.22 | 83.4/4.24 | 85.4/3.09 | 62.5/3.70 |
| Terminal α-L-Rha  | B1              | 101.9/4.79   | 71.2/3.95 | 71.3/7.58 | 73.2/2.42 | 70.2/4.01 |
| 4-β-D-GlcUA       | B2              | 103.8/4.52   | 74.3/3.46 | 77.8/7.43 | 80.1/2.42 | 75.5/3.57 |
| Terminal β-D-Gal  | D1              | 103.8/4.51   | 75.0/3.54 | 78.9/7.33 | 72.0/5.55 | 76.5/3.53 |
| 3-β-D-Gal         | G1              | 104.5/5.48   | 71.6/3.66 | 81.5/8.08 | 69.8/4.12 | 74.8/3.82 |
| 3-β-D-Gal         | G2              | 104.7/5.46   | 71.6/3.66 | 81.5/8.08 | 69.8/4.12 | 74.8/3.82 |
| 3-β-D-Gal         | G3              | 102.4/4.57   | 71.6/3.68 | 83.1/8.06 | 69.8/4.22 | 74.8/3.78 |
| 6-β-D-Gal         | G4              | 105.0/4.68   | 72.8/3.57 | 83.0/2.64 | 70.5/3.91 | 74.9/3.82 |
| 3-β-D-Gal         | G5              | 105.0/4.71   | 72.8/3.57 | 83.0/8.08 | 69.8/4.22 | 74.8/3.82 |
| Terminal β-D-Gal  | G6              | 105.0/4.71   | 72.8/3.57 | 74.3/3.70 | 70.5/3.90 | 72.7/3.73 |

<sup>a</sup> Residue designations (A<sub>1</sub>, A<sub>2</sub>, etc.) correspond to those in Figs. 5 and 6.

---

**Table III. Summary of 13C and 1H chemical shifts of the Hyp residue in AHP-1**

| Chemical shifts  | C-2/H-2 | C-3/H-3 | C-4/H-4 | C-5/H-5 |
|------------------|---------|---------|---------|---------|
|                  | ppm     | ppm     | ppm     | ppm     |
| 4-Linked Hyp      | Set 1   | 61.0/4.32 | 36.3/2.19 | 2.62 |
|                  | Set 2   | 61.3/4.21 | 36.0/2.42 | 2.56 |

<sup>b</sup> Chemical shifts are in ppm relative to TMS.
an α-(1→4) linkage. This was consistent with the composition and linkage analyses of AHP-1 (Table II). The one-dimensional 1H NMR spectrum, the HMBC spectrum, the linkage analyses of AHP-1, and the assignments shown in Table II indicated AHP-1 had a terminal β-D-GlcUA residue, and a terminal α-L-Rhap residue 1→3-linked to a second GlcUA residue.

Earlier work showed that a substitution at O-6 of Gal produces a downfield shift in the 13C signal from 62 ppm (unsubstituted) to 71 ppm (substituted) (11, 17). Likewise, the signals for H-6 shift from a single signal at 3.77 ppm (unsubstituted) to a split signal at 3.90/4.03 ppm (substituted) (11). Similar shifts reported here indicated the presence in AHP-1 of Gal residues having adducts at O-6 (i.e., as in 3,6-linked and 6-linked Gal, Ga,Gb,G0,G1 in Table II) all of which is consistent with the glycosyl linkage composition (Table I). Furthermore, the signals assigned to H-1 of GlcUA and C-6 of Galp yielded cross-peaks in the HMBC spectrum (labeled F in Fig. 3); therefore, we concluded the two GlcUA residues in AHP-1 were linked to O-6 of two Gal residues, which can only be Ga and Gb, judging by their unique C-6 resonances at 70.2 ppm. The H-1 signals of Ga and Gb were also correlated with unique Gal C-6 signals at 70.5 and 70.8 ppm that arose from two other Gal residues, designated G0 and G1, in Table II. Thus Ga and Gb, bearing Ara adducts at O-3 and GlcUA adducts at O-6, are themselves linked to the O-6 of two other Gal residues in AHP-1. Gal residue G0, which is linked to Hyp, has already been identified as a 3,6-linked residue and, judging by the chemical shifts assigned to G1,G2,G3,G4,G5,G6,G7 in Table II, is a 3,6-linked Gal residue as well (Table II). Thus Ga,Gb,G0,G1 were identified as 3,6-linked Gal residues and account for 4 of the 7 Gal residues in AHP-1.

In summary, AHP-1 contains two oligosaccharide side chains consisting of Ga and Gb substituted at position 3 with the α-L-Ara units and at position 6 with α-1-Rhap-(1→4)-β-D-GlcUA or β-D-GlcUA units.

**Fig. 5. Proposed primary structure of AHP-1.** Each residue is labeled in a manner corresponding to those featured in Table II, including the anomeric configurations and linkage positions. Hyp is in the lower right-hand corner of the structure. A NOESY spectrum collected at 30°C (not shown) indicated that the Ara side chain containing the 1→5-linked Ara residue was nearest the nonreducing end of the glycan; however, the spectrum did not indicate which side chain contained the Rha residue. Here the Rha, B1, is featured in the top center above, on the side chain nearest the non-reducing end.
The Galactan Backbone

Four Gal residues identified in the one-dimensional $^1$H NMR spectrum (Fig. 2) had H-1 shifts at ~4.7 ppm (specifically, 4.68 and 4.71 ppm, G1-G4 in Table II) and corresponding C-1 chemical shifts at 105.0 ppm, assigned through the HMQC spectrum. We assigned the shifts at 4.68 and 4.71 ppm, respectively, to four Gal residues within the galactan backbone, assignments not previously made because these chemical shifts are typically obscured by the water peak in earlier spectra collected at 30 °C (11, 31) but evident in spectra such as those reported here collected at a higher temperatures (60 °C) in which the water peak shifted upward.

There are two Gal residues with H-1 signals at 4.68 ppm and two Gal residues with H-1 signals at 4.71 ppm, judging by the presence in the TOCSY spectrum of four ring systems associated with these H-1 signals (Table II). We determined that G1 is linked to O-3 of G0, that G2 is linked to O-3 of G1, that G3 is linked to O-6 of G2, and that G4 is linked to O-3 of G3 based on the following evidence.

The HMBC spectrum (Fig. 3) showed three cross-peaks arising from G4-C-1 signals at 105.0 ppm and H-Gal signals at 3.82, 3.86, and 3.88 ppm (Fig. 3, cross-peaks labeled J) indicating that three of the four Gal residues discussed above were involved in β-D-Galp-(1→3)-Gal linkages, as Gal that is unsubstituted at position 3 exhibits H-3 chemical shifts much further upfield (i.e. 3.66 ppm (11)). None of these three Gal residues is G0 or G3 judging by the H-3 chemical shifts of G1 and G4 (Table II). Thus there is a Gal residue, designated G5, in a 1→3-linkage to G6 (G0 has a H-3 signal at 3.86 ppm). The second C-1/H-3 HMBC cross-peak indicated G2 is linked to O-3 of G1 (G1 has a H-3 signal at 3.82 ppm), and the third cross-peak indicated that there is a Gal residue linked to O-3 of another Gal residue possessing a H-3 signal at 3.88 ppm, designated G4 and G3, respectively. The G4→G3 and the G4→G2 linkages were corroborated by another set of cross-peaks in the HMBC spectrum arising from H-1 of G1 and C-3 of G0 and from H-1 of G4 and C-3 of G0 (Fig. 3, cross-peak H). Residue G4 is the terminal Gal in AHP-1, judged by the HMBC spectrum which showed a correlation between signals at 62.7 ppm (C-6 of G4, Table II) and 3.90 and 3.70 ppm (H-G4-A2(1→4)-Hyp) (G2, G3, and G0 in Fig. 5) and β-D-Galp-(1→3)-β-D-Galp (G4 and G3 in Fig. 5).

The NOESY spectrum determined at 60 °C (not shown) corroborated the results from the HMBC spectrum in that it showed three unique correlations between H-1 of G1 and H-3 of G0, between H-1 of G2 and H-3 of G1, and between H-1 of G4 and H-3 of G3.

We considered the possibility that G3 and G2 are the same Gal residue, i.e. the structure β-D-Galp-(1→3)-β-D-Galp-(1→3)-β-D-Galp-(1→3)-β-D-Galp (G2, G3, and G0 in Fig. 5) and β-D-Galp-(1→3)-β-D-Galp (G4 and G3 in Fig. 5). The NOESY spectrum determined at 60 °C (not shown) showed a correlation between an H-1 signal at 4.71 ppm (H-1 of G3 or G4) and signals at 3.94 and 4.04 ppm arising from the H-6 protons of a neighboring Gal (Table II). Although G4 like G3 has an H-1 signal at 4.71 ppm, G4 is a terminal residue and linked to O-3 of G3 (discussed above). Thus, it is G3 that is linked to O-6 of G2 (Fig. 5). A NOESY spectrum collected at 30 °C (not shown) corroborated that G3 is spatially near G2 as the spectrum showed correlations between the signal for H-1 of G3 (4.71 ppm) and the H-6 signals of another Gal (3.94 and 4.04 ppm) and between the signal for H-1 of G3 (4.71 ppm) and a signal at 3.91 ppm (H-4 of G3).

Thus, AHP-1 contains the 5-residue galactan backbone: β-D-Galp-(1→3)-β-D-Galp-(1→3)-β-D-Galp-(1→3)-β-D-Galp-G1-(1→4)-Hyp, and it has two branched side chains attached to the backbone such as α-L-Araf C1-(1→3)-α-L-Araf C1-(1→3)-β-D-GlcUA D1-(1→6)-β-D-Galp G3-(1→6)-β-D-Galp-(1→6)-α-L-Araf A1-(1→5)-α-L-Araf A2-(1→3)-α-L-Araf A3-(1→3)-α-L-Rhap B1-(1→4)-β-D-GlcUA B2-(1→6)-β-D-Galp G4-(1→6)-β-D-Galp-(1→6).

The final task was to determine the placement of the side chains along the galactan backbone. As discussed above, only G0 and G1 in the main chain galactan had ring systems with resonances characteristic of 3,6-linked Gal (Table II) and therefore are the sites for side chain attachment at O-6 (Fig. 5), but which side chain occurs on G0 and which on G1?

A NOESY spectrum carried out at 30 °C (not shown) had a signal at 4.71 ppm (H-1 of G3 or G4) that was correlated with another at 5.09 ppm (H-1 of A0 in Table II and Fig. 5). This suggested that Ara A1 in the Ara side chain, A1-(1→5)-A2-(1→3)-A3-(1→3)-Gal-(1→4)-Hyp, and it has two branched side chains attached to the backbone such as α-L-Araf C1-(1→3)-α-L-Araf C1-(1→3)-β-D-GlcUA D1-(1→6)-β-D-Galp G3-(1→6)-β-D-Galp-(1→6)-α-L-Araf A1-(1→5)-α-L-Araf A2-(1→3)-α-L-Araf A3-(1→3)-α-L-Rhap B1-(1→4)-β-D-GlcUA B2-(1→6)-β-D-Galp G4-(1→6)-β-D-Galp-(1→6).

Galactan Backbone Subunits

The AHP-1 structure may help resolve a long-standing question of “kinks” in the arabinogalactan backbone, which presumably arise from periodic occurrences of either (1→5)-α-L-Araf or (1→6)-β-D-Galp linkages. These residues were identified by their sensitivity to periodate oxidation and the subsequent release of apparently small repetitive arabinogalactan subunits (38, 39). However, as shown here, 5-linked Araf terminates a side chain. Thus, the single 6-linked Gal residue, if part of a truncated main chain repeat unit, could be the missing kink in the arabinogalactan main chain. Characterization of larger Hyp-arabinogalactans should help resolve the issue.

A Molecular Model of AHP-1

By using the structure of O-Hyp glyco-substituents and their location predicted from the Hyp contiguity hypothesis, it is now feasible to approach the molecular modeling of HRGPs. Figs. 5 and 6A show the proposed AHP-1 structure inferred from NMR and carbohydrate analyses: a 15-residue glycan based on a (1→3)-β-D-Galp main chain with bifurcated tetra- and pentasaccharide side chains containing Ara, Rha, and GlcUA linked through a side chain Gal to the main chain. As 10 residues define the maximum size of an oligosaccharide (40), AHP-1 is a 15-residue borderline arabinogalactan polysaccharide, small enough for molecular modeling to produce an energy-minimized polysaccharide conformation and orientation for comparison and possible corroboration by the NMR structure. It also enabled a search for possible interactions between closely clustered polysaccharides and a test of H-bonding be-
tween polysaccharides and the polypeptide backbone hypothesized earlier (41).

We adopted a strategy of constructing a small molecule containing O-Hyp-linked AHP-1 arabinogalactans, designated (Ala-Pro)₆ (C). Nitrogen atoms are shown in dark blue; the oxygen atoms are red; hydrogen atoms are gray; and carbon atoms are turquoise blue. 

**Fig. 6.** Space-filling CPK Models of AHP-1 (A), a side view of glycosylated (Ala-Pro)₆ (B), and an end-on view of glycosylated (Ala-Pro)₆ (C). Nitrogen atoms are shown in dark blue; the oxygen atoms are red; hydrogen atoms are gray; and carbon atoms are turquoise blue.

Bar, 1 nm. A, Hyp arabinogalactan AHP-1. This energy-minimized structure is stabilized by three H-bonds. Two stabilize the G₉ side chain: one between Ara C₂ and Ara C₁ (the C-2 OH donates to the C-6 O) and one between Gal G₉ and Gal G₄ (the C-2 OH donates to the C-4 O). A third H-bond stabilizes the G₉ side chain between Gal G₄ and Ara A₉ (the C-6 OH donates to the C-2 O). Dimensions of AHP-1 are as follows: x = 3.5 nm; y = 1.6 nm; z = 2.6 nm. Ara and Gal residue labels correspond to those of Fig. 5. B, Glycosylated (Ala-Pro)₆. Side view of a polysaccharide cluster. Three AHP-1 glycans labeled A-C are O-linked to the Hyp residues of the glycosylated (Ala-Pro)₆ model (residues 4, 6, and 8 of the peptide). The protein backbone lies across the figure with the N terminus at the far left (nitrogen atoms are in dark blue). Note the close proximity of polysaccharide B (green) to the polypeptide backbone where the arabinose disaccharide residues C₁ and C₂ on the G₉ side chain form three H-bonds as follows: the hydroxymethyl (C-5) of Ara residue C₁ to both the carbonyl of Hyp residue 4 and to the peptide N of Ala residue 5; and the C-2 hydroxyl of Ara residue C₂ to the NH of Ala residue 6. In contrast, the arabinose trisaccharide residues at the tip of each polysaccharide form peripheral hook-like projections; these may result in multiple weak interactions (“molecular Velcro”) with the Yariv reagent (2) which specifically interacts with the arabinogalactans on AGPs. C, glycopeptide (Ala-Pro)₆. End-on view of a polysaccharide cluster. Reorienting the polypeptide so it is perpendicular to the plane of the paper shows a syndiotactic propeller-like arrangement of the arabinogalactan polysaccharides around the polypeptide, providing surfaces for interactions and interdigitation with other matrix molecules.

We adopted a strategy of constructing a small molecule containing O-Hyp-linked AHP-1 arabinogalactans, designated (Ala-Pro)₆ (Fig. 6, B and C); the arabinogalactans can be considered as bulky side chains of the polypeptide. After building and energy minimizing the polysaccharides, we glycosidically linked three of them to C-4 of the internal Pro residues in the 12-residue non-glycosylated peptide (Ala-Pro)₆ to form a typical AGP tight cluster consisting of three consecutive arabinogalactan glycomodules in the resulting glycopeptide: Ala-Pro-Ala-Hyp-Ala-Hyp-Ala-Hyp-Ala-Pro-Ala-Pro in which each Hyp had an AHP-1 substituent (underlined). The (Ala-Pro)₆ nonglycosylated backbone initially showing a polyproline II conformation (PPII) with Ramachandran angles: φ -75° and ψ +145° (42) was re-energy minimized after each successive polysaccharide addition. A single AHP-1 polysaccharide contained three internal H-bonds (Fig. 6A) when energy-minimized in vacuo; energy minimization of AHP-1 in a 51-Å³ periodic box containing 5833 water molecules yielded essentially the same conformation and H-bonding patterns. Polysaccharides attached to the peptide showed similar internal H-bonding but none between adjacent polysaccharides. Most interesting, however, the central polysaccharide of the cluster (polysaccharide B in Fig. 6, B and C) also showed H-bonding with the polypeptide backbone, although the earlier suggestion of extensive interaction between arabinogalactan polysaccharides and polypeptide backbone (41) is too simplistic. Indeed, rather than following
the contours of the polypeptide, the polysaccharides tend to splay out in the alternating syndiotactic manner with dimensions ranging from 3.5 to 5 nm across two polysaccharides orthogonal to the peptide backbone; the peptide ϕ/ψ angles deviated significantly from the PPII conformation, ϕ fluctuating from −43 to −132 and ψ from +98 to +153, indicating that the polysaccharides impose steric constraints deforming the polypeptide backbone, confirmed by the "random coil" CD spectrum of (Ala-Hyp)51 and the partial restoration of 3-fold PPII helicity observed after its deglycosylation with anhydrous HF (Fig. 7). The low temperature NOESY spectrum indicated that the side chain Rha residue is close to the main chain Gal residues G1 or G2 and therefore favors rhamnose attachment to the G2 side chain of G1; the molecular model is consistent with this as it shows Rha closer to G1 than G2. The model also shows two shallow pockets in AHP-1, the first lined by Gal, GlcUA, and Rha and the second by side chain Gal residue G2 and main chain Gal residues G1, G2, and G3 each have a 1,6-linked side chain substituent in close proximity to one another; this results in steric hindrance to their rotation around the 1,6-linkage, further enhanced by the H-bonding, and thus may also stabilize these shallow pockets as putative water-binding sites that may account for the nonfreezing water content identified by differential scanning calorimetry of gum arabic arabinogalactan polysaccharides (43).

Glycomodule Subunits

Judging from the similarities between gum arabic side chains (17) and diverse species ranging from rice (32) to grape berries (37), the AHP-1 arabinogalactan is a small consensus subunit that represents many AGP glycomodules. This confirms the insightful observation of Churms et al. (44) that "any arabinogalactan of this general type, irrespective of its molecular complexity, can be regarded as being composed of uniform subunits" and raises the distinct possibility that AGP biosynthesis involves their en bloc transfer from a lipid intermediate (16, 45). "Snapshots" of small structural units in other arabinogalactans are also similar to AHP-1 albeit with some variations, possibly including fucose (6-deoxy-1-galactose) (1, 46) as a conservative replacement for Rha (6-deoxy-1-mannose). Other significant variants may include an alternative β-anomeric configuration for the terminal α-L-Araf (32), the presence of xylose (1), and fewer rhamnose and uronic acid residues (47, 48). These putative variants need confirmation by their isolation and characterization as unique Hyp-polysaccharides or glycopeptides.

On the other hand, AGPs from diverse species like Acacia senegal (Leguminosae) (17), N. tabacum (18), and Lycopersicon esculentum (Solanaceae) (4) contain repetitive structures similar to AHP-1. This structural conservation points to common functions of cell-surface HRGPs that originally evolved in the green algae (49) where self-assembly of the wall involves modular extensin-like HRGPs to a lesser (50) or greater extent (51, 52), with AGPs contributing to mucilage around cells (53). Higher plants have relegated AGP mucilage to the secondary role of wound-induced protective gum exudates (41), whereas primary roles remain less than obvious. However, concurrent work2 suggests a dual role for AGPs based on the dimensions of the anionic polysaccharide structures described here and their quantitative distribution. Thus, monomeric AGPs may act as plasticizers in muro (54), whereas the classical AGPs at the membrane/wall interface, such as LeAGP-1 (4), may form a periplasmic polymer cushion that stabilizes the plasma membrane protecting it from the inherently high hydrostatic pressures in turgid plant cells.

Acknowledgments—The glycosyl linkage analyses and some of the composition analyses were performed at the Complex Carbohydrate Research Center, the University of Georgia. The Complex Carbohydrate Research Center was the recipient of National Institutes of Health Grant 2-P41-RR05351-06. We thank Oklahoma State University and Temple University for use of their NMR facilities, and the University of Sussex for their laboratory facilities.

REFERENCES

1. Gaspar, Y., Johnson, K. L., McKenna, J. A., Bacic, A., and Schultz, C. J. (2001) Plant Mol. Biol. 47, 161–176
2. Nothnagel, E. A. (1997) Int. Rev. Cytol. 174, 195–291
3. Kieliszewski, M., O’Neill, M., Leykam, J., and Orlando, R. (1995) J. Biol. Chem. 270, 2541–2549
4. Zhao, Z., Tan, L., Showalter, A. M., Lampert, D. T. A., and Kieliszewski, M. J. (2002) Plant J. 31, 431–444
5. Du, H., Clarke, A., and Bacic, A. (1996) Trends Cell Biol. 6, 411–414
6. Gens, J. S., Fujiki, M., and Pickard, B. G. (2000) Prototransla 212, 115–134
7. Youl, J. J., Bacic, A., and Oxley, D. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 7921–7926
8. Svetek, J., Yadav, M. P., and Nothnagel, E. A. (1999) J. Biol. Chem. 274, 14724–14733
9. Sherrier, D. J., Prime, T. A., and Dupree, P. (1999) Electrophoresis 20, 2027–2035
10. Oxley, D., and Bacic, A. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 14246–14251

2D. T. A. Lampert, A. M. Showalter, and M. J. Kieliszewski, unpublished data.
