Plant Physiol. (1991) 96, 848–855
0032-0889/91/96/0848/08/$01.00/0

Received for publication September 27, 1990
Accepted February 19, 1991

Gum Arabic Glycoprotein Is a Twisted Hairy Rope[1, 2]

A New Model Based on *O*-Galactosylhydroxyproline as the Polysaccharide Attachment Site

Wu Qi, Cynthia Fong, and Derek T. A. Lamport*

Agrobiological College, Beijing Agricultural University, Beijing, P. R. China, 100094 (W.Q.), Halfhollow Hill High School East, Dix Hills, New York 11746 (C.F.), and Michigan State University-Department of Energy, Plant Research Laboratory, and Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824 (D.T.A.L.)

ABSTRACT

Separation of the wound exudate from *Acacia senegal* (L.) Willd., "gum arabic," on a preparative Superose-6 column gave two major fractions: a high molecular weight gum arabic glycoprotein (GAGP) containing about 90% carbohydrate and a lower molecular weight heterogenous gum arabic polysaccharide fraction. Hydrogen fluoride-deglycosylation of GAGP gave a large (~400 residue) hydroxyproline-rich polypeptide backbone (dGAGP). Alkaline hydrolysis of GAGP showed that most of the carbohydrate was attached to the polypeptide backbone as small (~30 residue) hydroxyproline (Hyp)-polysaccharide substituents. After partial acid hydrolysis of the Hyp-polysaccharide fraction we identified *O*-galactosylhydroxyproline as the glycopeptide linkage, identical with that of hydroxyproline-rich arabinogalactan-proteins (AGPs). However, unlike the acidic alanine-rich AGPs, GAGP is basic and notably deficient in alanine. Thus, while the GAGP polypeptide backbone more closely resembles that of the Hyp-rich cell wall protein extensin, the GAGP polysaccharide sidechains resemble AGPs. Possibly all three proteins comprise a phylogenetically related extensin superfamily of extended rodlike macromolecules. The "wattle-blossom" model for AGP and gum arabic predicts a few large polysaccharide substituents along the polypeptide backbone of a spheroidal macromolecule. On the contrary, our data imply a rodlike molecule with numerous small polysaccharide substituents (attached to 24% of the Hyp residues), regularly arranged along a highly periodic polypeptide backbone based, hypothetically, on a 10 to 12 residue repetitive peptide motif. Thus, a simple statistical model of the gum arabic glycoprotein predicts a repeating polysaccharide-peptide subunit of about 7 kilodaltons. The small polysaccharide substituents will maximize intramolecular hydrogen bonding if aligned along the long axis of the molecule, forming in effect a twisted hairy rope. Electron micrographs of rotary shadowed GAGP molecules support that prediction and may also explain how such apparently large molecules can exit the cell by endwise reptation through the small pores of the primary cell wall.

Gums are of considerable commercial importance and, as products of the specific wound response, gummosis, also of biological interest (2, 5). Hence, their structural elucidation will enhance our understanding of both their biological and utilitarian roles. These goals are far from simple; for example, gum arabic research extends more than 50 years or so (29)! As an additional complicating factor, gums are not exclusively polysaccharide, invariably containing small but nevertheless significant amounts of protein rich in hydroxyproline. Quite possibly this exists as a structural HRGP[3]; thus, gums may be related to other HRGPs such as the AGPs as suggested by Akiyama *et al.* (1) or the basic extensins of the primary cell wall as suggested here.

Our long-standing interest in cell wall HRGPs (16), combined with recent suggestions of a possible amino acid-polysaccharide (glycopeptide) linkage in gum arabic (1), prompted us to reexamine the problem with techniques used successfully for structural characterization of extensin HRGPs.

We report here that fractionation via preparative Superose-6 fplc resolves crude gum arabic into two major components: a somewhat heterogeneous GAP (a glucuronorhamnoarabinogalactan containing little or no protein) and a unique high mol wt GAGP corresponding to the "arabinogalactan-protein complex" described recently (25, 31).

Deglycosylation of the GAGP by anhydrous HF-solvolysis (22, 23) gave about 90% weight loss; subsequent gel filtration on Superose-6 and cleanup on a polySulfoethylaspartamide HPLC column, yielded a pure polypeptide (dGAGP), considerably smaller than the native GAGP, indicating a very highly (~90%) glycosylated glycoprotein. The amino acid composition of dGAGP resembled the typical serine-rich extensins rather than AGPs. In particular GAGP's low content of acidic amino acids and alanine was quite different from typical alanine-rich acidic AGPs (11). On the other hand, GAGP's high serine and histidine content was similar to that of typical extensins (28).

Further chemical degradation identified the linkage between protein and carbohydrate: alkaline hydrolysis of the

[1] Supported by a grant from Pepsico, Inc., and DOE grant No. DE-AC02-76ERO-1338.

[2] Dedicated to the memory of Dr. Michael A. Jermyn.

[3] Abbreviations: HRGP, hydroxyproline-rich glycoprotein; AGP, arabinogalactan-protein; dGAGP, deglycosylated gum arabic glycoprotein; fplc, fast protein liquid chromatography; GAGP, gum arabic glycoprotein; GAP, gum arabic polysaccharide; HF, hydrogen fluoride; Hyp, hydroxyproline; TEM, transmission electron microscopy.

intact highly glycosylated gum arabic glycoprotein released a series of small hydroxyproline-polysaccharides which, on further partial acid hydrolysis, gave *O*-galactosylhydroxyproline, confirming that glycopeptide linkage as the major polysaccharide attachment site. Remarkably, the *O*-galactosylhydroxyproline linkage is similar to that identified in sycamore-maple AGP, as previously reported from this laboratory (24). Thus, gum arabic glycoprotein shares some characteristics of AGPs and some of extensin, raising the possibility that these three glycoproteins comprise a phylogenetically related HRGP-extensin superfamily of rodlike macromolecules.

MATERIALS AND METHODS

Nodules of authentic gum arabic (*Acacia senegal*) were a gift from Pepsico Inc. We powdered the nodules in a Tekmar A-10 mill for about 2 min to yield a fine white flour. (Note: Where appropriate we include numbered references to our detailed laboratory protocols available on request.)

Acid Hydrolysis and Amino Acid Analysis (p20584 & p11078)

Dried samples were hydrolyzed with 0.2 mL of 5.5 N constant boiling HCl in a sealed tube at 110°C for 18 h.

We used a Pickering High Speed Na^+ cation exchange column (3 mm i.d. × 150 mm) in series with a BX-8 cation exchange column (3.7 × 70 mm, Benson Co.) eluted by Pickering buffers A, B, and C. Postcolumn fluorometric detection involved NaOCl oxidation and OPA coupling which allowed Hyp and Pro detection as previously described (15) and data capture via PE Nelson Turbochrom software.

Sugar Analysis via Trimethylsilylation of the O-Methyl Glycosides (p19979 & p10880)

We used a modified trimethylsilyl derivatization procedure based on that of Bhatti *et al.* (6) briefly as follows: Dried samples containing 50 to 100 μg sugar together with 100 nmol mannitol as internal standard were methanolysed in 1.5 N methanolic HCl for 90 min at 95°C in a sealed vial under nitrogen. After N2 blow down the dried (3 h *in vacuo*) methanolysate was reacted with 50 μL pyridine/trimethylchlorosilane/hexamethyldisilazane for 30 min at room temperature. We injected 2 to 5 μL onto a 3% SP-2100 gas chromatography column (12 feet × ⅛ i.d.) with He as carrier gas (40 mL/min) and temperature programmed from 120 to 225°C at 4°C/min. Data capture was via the PE Nelson Turbochrom system.

Alkaline Hydrolysis (p04479)

Samples were hydrolyzed in 0.2 M $Ba(OH)_2$ for 18 h at 105°C, neutralized with 1 N H_2SO_4, centrifuged, and freeze-dried as previously described (19).

Hydroxyproline Glycoside Chromatography on Chromobeads C2 (p12383)

We dissolved the neutralized alkaline hydrolysate of either the crude gum arabic or purified GAGP (containing about 100 μg Hyp) in 0.5 mL water and applied it to a 75 × 0.6 cm column (H^+ form) of Technicon chromobeads C-2 resin eluted with a 0 to 0.5 N HCl gradient and monitored with a specific postcolumn hydroxyproline assay at 560 nm (19).

Hydroxyproline Assay (p17176)

Hydroxyproline determination involved a highly specific automated assay using alkaline hypobromite oxidation and subsequent coupling with acidic Ehrlich's reagent monitored at 560 nm as previously described (19).

Purification of Gum Arabic Glycoprotein via Superose-6 Gel Filtration (p20685)

We purified the gum arabic glycoprotein by gel filtration of 100 mg/mL aliquots injected onto a Pharmacia Superose-6 preparative column (1.6 × 50 cm) eluted with 0.2 M (pH 7) sodium phosphate buffer at 1 mL/min and monitored at 220 nm, with data capture via PE Nelson Turbochrom software. An analytical Superose-6 column (1 × 30 cm) eluted with the same buffer was run at 530 μL/min.

Uronic Acid Estimation (p35375)

We used the specific colorimetric assay based on reaction with *m*-hydroxydiphenyl (7).

Arabinogalactan Protein Assay via Yariv Antigen (p19080)

Yariv antigen was a kind gift from the late Dr. Michael Jermyn. The assay involved addition of 100 μg β-glucosyl Yariv compound to 200 μL sample, incubation for 1 h at room temperature, precipitation in 2% NaCl, washing, redissolving in 20 mM NaOH, and measuring the A_{420} nm.

HF Deglycosylation (p14984)

We deglycosylated 30 to 100 mg Superose-6 purified GAGP in a microapparatus (27) containing 2 mL anhydrous HF and 200 μL anhydrous methanol for 1 h, 2 h at 0°C, and 3 h at room temperature. The reaction was quenched by pouring into 20 mL water at 0°C, then dialyzed for 48 h at 4°C and freeze-dried.

HPLC on PolySULFOETHYL Aspartamide SCX

A 4.6 × 100 mm column (from The Nest Group) was equilibrated in buffer A: 10 mM (pH 3) sodium phosphate buffer containing 10% (v/v) acetonitrile. After sample injection we eluted the column at 0.75 mL/min using a 0 to 10% buffer B gradient in 60 min, where buffer B is buffer A containing in addition 1 M NaCl.

Preparation of Hyp-polysaccharide via Sephadex (p18887)

Neutralized alkaline hydrolysates [0.2 M $Ba(OH)_2$] of purified gum arabic were fractionated on a Sephadex G-25 superfine column (30 × 1.2 cm) eluted with 0.1 M acetic acid into

Table I. *Amino Acid Composition (mol %) of Gum Arabic Fractions*

The simplest empirical formula for a small repetitive motif calculated from the below data is the 11-mer: Hyp_4 Ser_2 Thr Pro Gly Leu His.

Amino Acid	Crude Gum Arabic	GAGP	dGAGP	GAP	Contaminant
Hyp	32.7	36.9	30.4	13.3	7.1
Asp	3.9	1.6	3.4	9.5	4.5
Thr	7.0	8.8	8.8	4.8	2.8
Ser	16.3	19.4	21.0	10.6	26.4
Glu	2.9	1.9	4.0	9.7	11.5
Pro	7.6	6.8	6.2	7.6	3.6
Gly	6.9	6.4	7.5	9.5	14.4
Ala	1.9	1.3	2.2	5.7	6.1
Val	2.3	0.8	1.0	7.4	2.6
Ilu	1.2	0.4	0.7	3.1	1.5
Leu	6.8	6.4	5.4	6.9	2.4
Tyr	0.6	0.3	0.7	1.8	0.9
Phe	1.8	0.9	1.0	4.7	1.1
Lys	2.7	1.0	1.2	0.0	8.6
His	5.8	7.1	6.4	2.7	4.5
Arg	0.0	0.0	0.3	2.7	0.5
Protein[a] (% w/w)	4.0%	7.0%			

[a] % w/w from weight loss on HF treatment.

1 mL fractions. We freeze-dried the void for further fractionation on Sephadex G-50.

TEM Sample Preparation

We prepared samples of GAGP for TEM as previously described (14) by adsorption onto freshly cleaved mica chips and rotary shadowing with Pt/C followed by backing with carbon and then examination of the replicas in a JEOL 100 CX II transmission electron microscope.

RESULTS

Compositional Analysis of Crude Gum Arabic

Amino acid analyses (Table I) and sugar analyses (Table II) from representative gum arabic nodules were similar to those previously reported by other workers (1, 3). Our powdered samples of individual gum arabic nodules contained approximately 2 to 4% protein in which Hyp was a prominent component accounting for more than 30 mol% of the total amino acid residues.

Table II. *Sugar Compositions of Gum Arabic (% weight)*

Sugar	Crude Gum Arabic	GAGP	GAP	dGAGP		
				1 h[a]	2 h[a]	1 + 1 h[b]
Arabinose	26.1	28.3	21.2	3.1	2.0	0.5
Rhamnose	9.9	8.2	9.3	1.8	1.0	0.3
Galactose	40.1	36.3	35.6	5.0	3.1	1.0
Glucuronic	9.4	6.9	8.5	2.1	0.9	0.4
Total % weight	85.5	79.7	74.5	12.0	7.0	2.2

[a] Deglycosylation time. [b] Deglycosylated twice for 1 h each time.

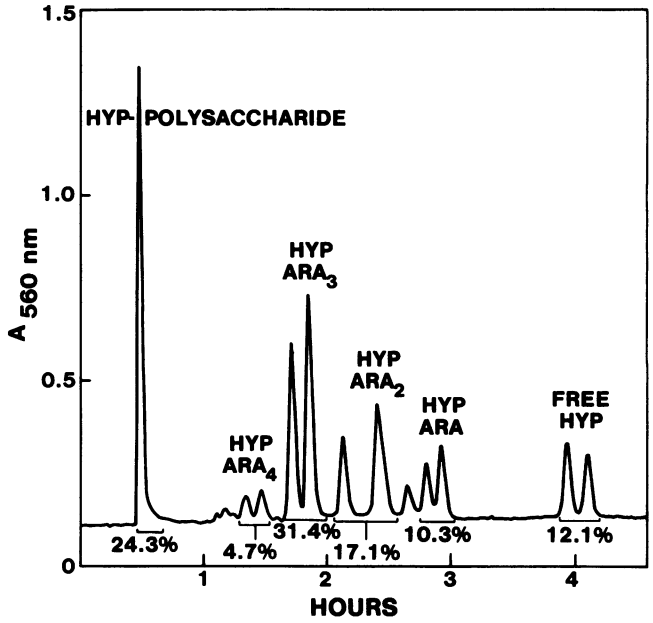

Figure 1. Hyp glycoside profile of crude gum arabic on chromobeads C-2. A neutralized alkaline hydrolysate of crude gum arabic (15 mg containing about 100 μg Hyp) was dissolved in 0.5 ml H_2O and chromatographed as described in "Materials and Methods."

Alkaline hydrolysates of the crude gum arabic gave a series of Hyp glycosides; based on Hyp content, 76 mol% of these resolved on a Technicon chromobeads cation exchange column (Fig. 1) as typical small Hyp glycosides while the other 24 mol% ("Hyp-polysaccharide") voided the column indicating much larger attached *O*-substituents; these fractionated further on Sephadex G-50 to yield a heterogeneous series of small Hyp-polysaccharides (see below).

Purification of GAGP

Gel filtration of crude gum arabic by fplc on a Superose-6 preparative column typically gave two major fractions when monitored by UV absorbancy at 220 nm and a specific uronic acid assay (Fig. 2); the first fraction (about 10% of the total mass) corresponded to the GAGP and contained about 10% hydroxyproline-rich protein (Table I) and about 90% carbohydrate consisting mainly of arabinose and galactose with less rhamnose and glucuronic acid (Table II). The other heterogeneous fraction (about 90% of the total mass), which we term GAP, accounted for the bulk of the glucuronic acid content of crude gum arabic (Fig 2; Table II) but contained only a small (0.3–1%) amount of protein relatively poor in Hyp (Table I). Both GAGP and GAP reacted strongly with the Yariv reagent (Table III). In our assay 100 μg gave the following absorbancies at 520 nm: GAGP 1.509; GAP (Table I; Fig. 2) 1.448; sycamore-maple AGP 0.532. However, contrary to a previous report (1) dGAGP did not react with the Yariv reagent.

HF-Deglycosylation of GAGP

After HF-deglycosylation of Superose-purified GAGP, the peak retention time increased significantly (Fig. 3; Table IV)

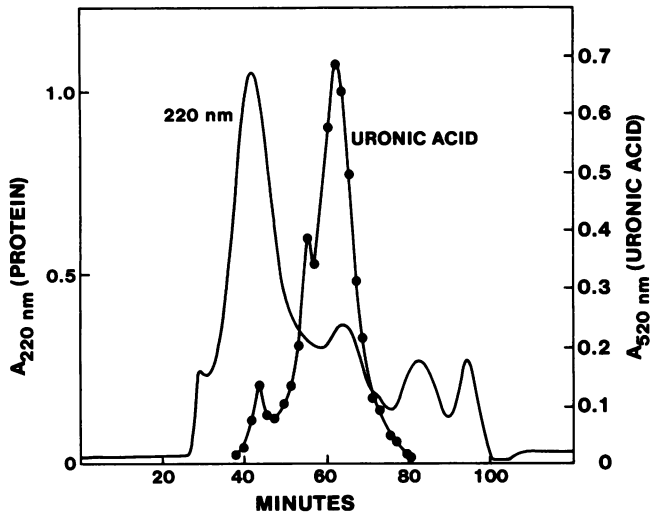

Figure 2. Purification of GAGP on a Superose-6 preparative column. GAGP corresponds to the fraction eluting between 38 and 48 min, and GAP 52 and 72 min.

consistent with a ~90% weight loss and formation of the deglycosylated product dGAGP. HF-deglycosylation at 0°C either for 1 h or two consecutive 1 h treatments (each followed by a cold water quench and dialysis) yielded a narrow symmetrical deglycosylated product peak on the Superose-6 analytical column (Fig. 3B) containing, respectively, 12 and 2% residual sugar (Table II) and an amino acid composition that remained essentially unchanged from the initial gum arabic to the final clean up of dGAGP on Sulfoethylaspartamide (Fig. 3D; Table I). More vigorous HF-deglycosylation (3 h at 23°C) gave a broad double peak product on Superose-6 (Fig. 3C; Table IV) indicating some peptide bond nicking under the more severe solvolysis conditions.

A quicker method of preparing dGAGP by direct deglycosylation of crude gum arabic followed by Superose-6 fractionation gave a product with an amino acid composition similar to that of dGAGP prepared from Superose-purified GAGP (Table I).

Glycopeptide Linkages of GAGP

Alkaline hydrolysates of the Superose-purified glycoprotein (GAGP) gave void (Hyp-polysaccharide) and retarded gel filtration fractions on Sephadex G-25 (Fig. 4); both reacted positively, in the highly specific Hyp assay, without further hydrolysis, thus indicating nonpeptidyl Hyp. Partial acid hydrolysis (0.05 N H_2SO_4 for 1 h at 100°C) of the G-25 void (*i.e.* the Hyp-polysaccharide), followed by cation exchange chromatography of the products on Chromobeads gave a characteristic Hyp glycoside profile (Fig. 5) which included free Hyp and peaks 3 and 4 whose composition (Table V) showed a 1:1 molar ratio of galactose and Hyp. Note that this profile is quite different from that obtained with the native crude gum arabic glycoprotein in Figure 1.

We also fractionated alkaline hydrolysates of GAGP directly on Sephadex G-50 (Fig. 6) to yield a somewhat heterogeneous high mol wt fraction (equivalent to the G-25 void Hyp-polysaccharide) and a low mol wt fraction; both reacted as though they contained free, *i.e.* nonpeptidyl Hyp. Table VI shows that sugar/Hyp molar ratios of the high mol wt subfractions in Figure 6, range from about 18 to >40 sugar residues/Hyp, corresponding to an average Hyp-polysaccharide of about 30 sugar residues. Figure 7 shows a statistical model for the GAGP (see "Discussion").

TEM of GAGP

Figure 8 shows that rotary shadowed GAGP molecules are filamentous rodlike molecules similar to monomers of extensin but about twice as long (~150 nm) and somewhat thicker. Judging from a few nearly circular images (Fig. 8A) GAGP macromolecules possess a higher degree of conformational freedom, *i.e.* GAGP is more flexible than the relatively stiff extensin monomers (14). However, flow-polarized GAGP (Fig. 8C), like extensin (14), shows no obviously kinked regions, although there is evidence of molecular asymmetry in the form of a thinner tail region (Fig. 8C).

DISCUSSION

Although extensively investigated for many years (29) complete structural elucidation of *Acacia senegal* gum exudates ("gum arabic") remains a formidable challenge. The presence of protein, known for some time (4) may be a complicating factor but more likely the essential clue. Recently, Randall *et al.* (25) showed component heterogeneity with most of the protein occurring in the highest mol wt fraction of the crude gum arabic. We confirm their observation here and we have identified the high mol wt gum arabic fraction as a specific glycoprotein (GAGP), *i.e.* a Hyp-rich glucuronorhamnoarabinogalactan. GAGP, like typical AGPs, reacts positively with Yariv's reagent, but differs very significantly from typical (*i.e.* alanine-rich) AGPs by its much lower content of alanine and acidic amino acids (Table I). On the other hand, alkaline hydrolysis of crude gum arabic (Fig. 1) and the purified GAGP (Fig. 4) yielded a mixture of Hyp glycosides similar to that first observed in this laboratory for an AGP isolated from sycamore maple (24); characteristically, these glycosides occurred in two distinct sizes: low mol wt (entirely retarded on a Sephadex G-50 column; Fig. 6) mainly Hyp arabinosides (17); and small Hyp-polysaccharides which voided a Sephadex G-25 column (Fig. 4) but showing considerable size heterogeneity on Sephadex G-50 (Fig. 6). Table VI indicates that the largest Hyp-polysaccharides have a Hyp:sugar molar ratio of about 1:40 or approximately half that of the sycamore AGP Hyp-polysaccharide which was retarded on G-75 but voided a G-50 column (24).

Table III. *Reaction with Yariv Reagent*

Sample Weight	Sycamore		
	AGP	GAGP	GAP
	Absorbancies at 520 nm		
20	0.092	0.434	
50	0.259	0.915	
100	0.532	1.509	1.448

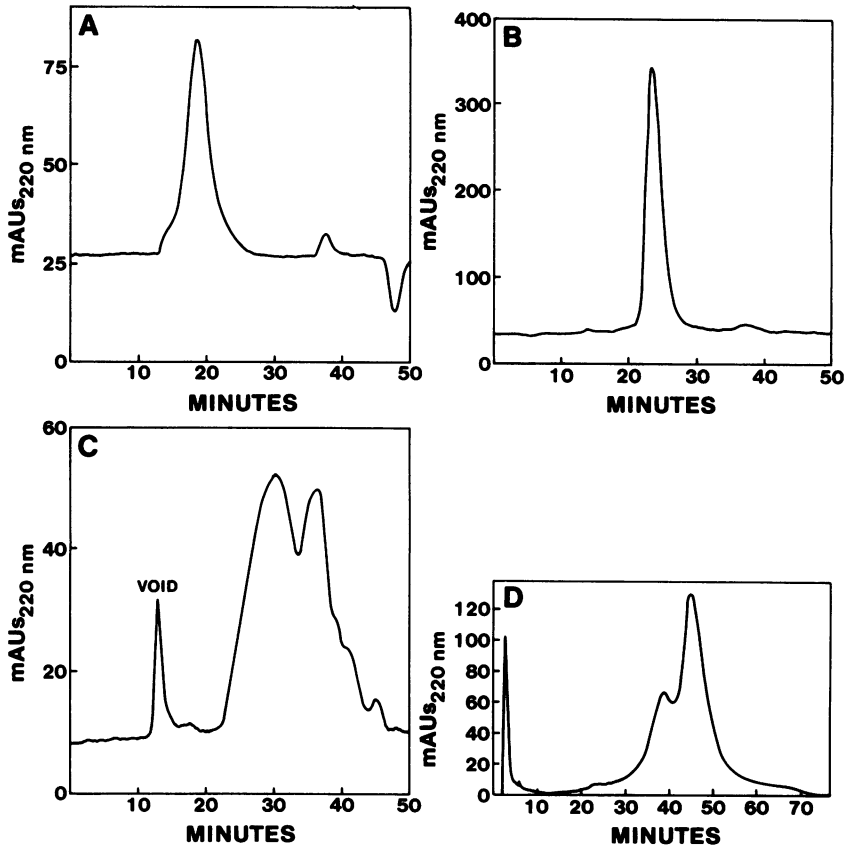

Figure 3. GAGP and dGAGP profiles via Superose-6 (analytical) and polySULFOETHYL Aspartamide CSX. A, Intact GAGP via Superose-6; B, dGAGP (after 2 h HF treatment at 0°C) via Superose-6; C, dGAGP (after 3 h at RT) via Superose-6; D, dGAGP (2 h HF at 0°C) clean up on polySULFOETHYL Aspartamide CSX.

An outstanding similarity between the gum arabic glycoprotein and AGPs appeared on partial acid hydrolysis of the Hyp-polysaccharides (obtained by alkaline hydrolysis of GAGP) which yielded *O*-galactosylhydroxyproline exclusively as the smallest glycosylated product (Fig. 5; Table V). This indicates direct attachment of the galactan backbone to Hyp residues of the polypeptide backbone. Interestingly, *O*-galactosylhydroxyproline *per se* also occurs in alkaline hydrolysates of Chlamydomonas cell walls (21) and as methylglucuronosylgalactosylhydroxyproline isolated from *Acacia erioloba* gum (12) although not as a polysaccharide attachment site.

Now the question arises: how much of the GAGP polysaccharide is attached as an *O*-linked substituent to Hyp residues? Figures 1 and 4 show that, after alkaline hydrolysis, Hyp occurred in three size classes with a percent Hyp distribution: (a) nonglycosylated: 12.1%; (b) oligoglycosylated: 63.5%; and (c) polyglycosylated: 24.3%. From the data in Figures 1 and 6 and Table VI, we calculated the average molecular masses for the Hyp-saccharide size classes: (b) low molecular mass Hyp glycosides 440 D; and (c) Hyp-polysaccharide 4.44 kD (using an average 30 sugars/Hyp with an average residue weight of 148 D—see "Results"). Hence, we could then calculate the relative contribution of Hyp-saccharides (b) and (c) to a hypothetical gum arabic polypeptide containing 100 Hyp residues, as follows: (b) 28 kD (440 × 64%) plus (c) 107 kD (4440 × 24%) equals 135 kD. Thus, we estimate that if the GAGP polypeptide contained 100 Hyp residues (corresponding to 300 amino acid residues total or 30 kD) the gum arabic glycoprotein *in toto* would weigh in at about 165 kD. Hence, the next question—what is the actual size of the GAGP polypeptide backbone?

After HF-deglycosylation and subsequent gel filtration, dGAGP coeluted with glycosylated extensin P1 of tomato (Fig. 3; Table IV). Therefore, the polypeptide backbone of

Table IV. *Retention Times of GAGP, dGAGP, and P1 Extensin on Superose-6 (analytical)*
Void RT = 12.9 min at 530 μL/min (Fig. 3C).

RT	Sample						
	GAGP	dGAGP				Tomato P1	Tomato P1 + dGAGP[c]
		1 h[a]	2 h[a]	1 + 1 h[b]	3 h[a]		
(min)	18.5	23.8	23.6	24.0	30.1	24.9	24.6

[a] Deglycosylation time. [b] Deglycosylation twice for 1 h each time. [c] dGAGP (1 h) and tomato P1 coeluted as a single peak.

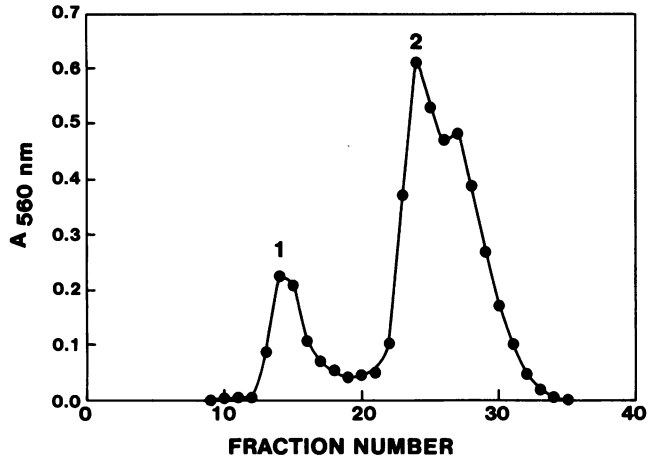

Figure 4. Preparation of Hyp-polysaccharide from an alkaline hydrolysate of GAGP by gel filtration on Sephadex G-25. A neutralized alkaline hydrolysate of GAGP, 15 mg in 150 μL 0.1 M acetic acid, was injected onto a Sephadex superfine G-25 column eluted and assayed as described in "Materials and Methods." Peak 1 is the void which also corresponds to the Hyp-polysaccharide described in "Results." Peak 2 corresponds to the low mol wt Hyp-glycosides.

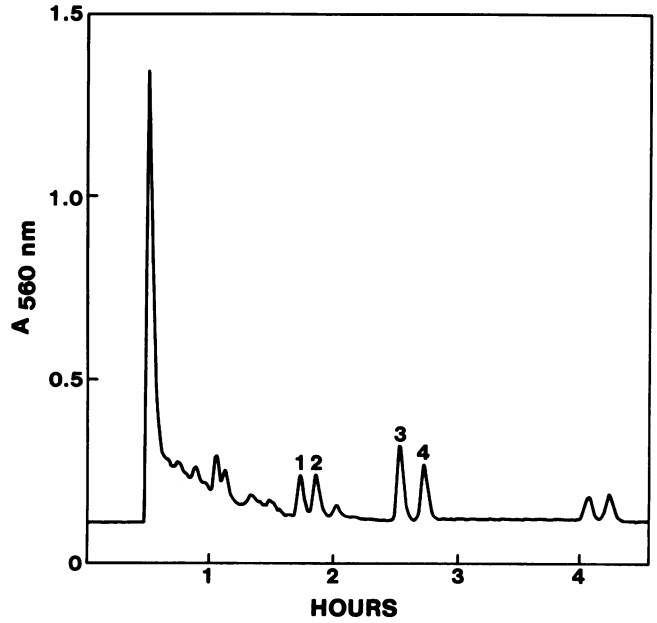

Figure 5. Partial acid hydrolysate of Hyp-polysaccharide on Chromobeads C-2. Hyp-polysaccharide (Sephadex G-25 void) partially hydrolyzed and run on Chromobeads C-2 as described in "Materials and Methods" and "Results." Table V shows analytical data for the numbered peaks where 3 and 4 are identified as *O*-galactosylhydroxyproline and the double peaks at 4 h are free Hyp.

Table V. *GAGP Hyp-Polysaccharide*

Composition of partial acid hydrolysis products after cation exchange chromatography on Chromobeads C-2.

Peak No. (Fig. 5)	Composition Expressed as Sugar/Hyp Molar Ratios			
	Ara/Hyp	Rha/Hyp	Gal/Hyp	GlcU/Hyp
1	0.44	0.45	2.37	0.28
2	0.15	0.11	1.9	0.04
3	0.04	0.03	0.92	0.07
4	0.04	0.02	0.93	0.03

dGAGP was greater than 300 residues, probably >400 total amino acid residues, corresponding to ~130 Hyp residues calculated from the ~33 mol% Hyp content of GAGP (Table I). This size estimate indicates that Hyp-saccharides would contribute about 180 kD to a 400-residue (40 kD) dGAGP polypeptide backbone to yield a total molecular mass of ~220 kD for the intact gum arabic glycoprotein. This estimate is significantly lower than previous estimates of 560 kD based on gel filtration (9) or 340 kD based on sedimentation (9) of an assumed spheroidal macromolecule. But, the Hyp/Pro content implies a rod rather than a spheroid; such highly asymmetric rodlike molecules behave as if they were much larger, giving anomalously small elution volumes on gel filtration possibly due to their limited endwise insertion into the gel matrix (14).

Our calculation of a 180 kD carbohydrate contribution based on Hyp-saccharide analyses indicates GAGP is 82% carbohydrate, compared with 93% inferred from weight loss on exhaustive HF-solvolysis. The difference between these two figures may reflect experimental error (sugar analyses or an underestimate of the polypeptide size) or quite possibly some degradation during the base-catalyzed hydrolysis of GAGP leading to a low size-estimate of Hyp-polysaccharide. Nevertheless, the data account for most, if not all, the glucuronoarabinogalactan as *O*-linked Hyp-polysaccharide, hence, glycosylated serine or other residues would only make a minor contribution, if any.

So far we have implicitly assumed, on the basis of gel filtration peak symmetry, that dGAGP is a single polypeptide backbone. Our strategy based on "purification to constant

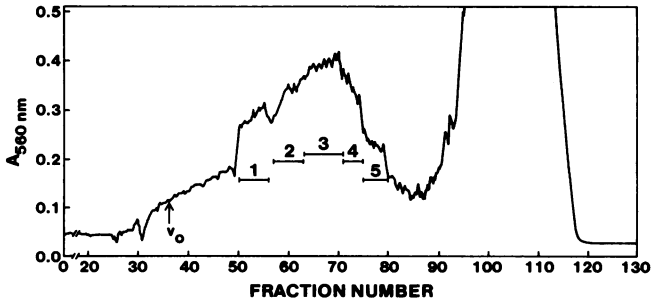

Figure 6. Fractionation of GAGP Hyp-polysaccharide on Sephadex G-50. A neutralized alkaline hydrolysate of 70 mg GAGP was chromatographed on a Sephadex G-50 column and assayed as described in "Materials and Methods."

Table VI. *GAGP Alkaline Hydrolysate via Sephadex G-50*

Hyp-polysaccharide fractions 1 to 5 (Fig. 6); compositions expressed as moles sugar/mole Hyp.

Sugar	Fraction No.				
	1	2	3	4	5
Arabinose	16	11	10	10	9
Rhamnose	5	3	3	3	2
Galactose	17	11	10	9	6
Glucuronic	4	2	2	2	1
Sugars/Hyp	42	27	25	24	18

composition" (Table I) confirms that assumption, although we cannot rule out some microheterogeneity in the carbohydrate sidechains. However, further fractionation of dGAGP on Superose-6 followed by Sulfoethylaspartamide did remove a minor serine-rich, Hyp-poor contaminant (Fig. 3D; Table I). Thus, taking into account the reasonable agreement between the quantitative Hyp-saccharide data and the weight loss on HF-deglycosylation, Superose-purified GAGP itself consists of a single hydroxyproline-rich glycoprotein with a minor serine-rich contaminant.

While a double treatment with HF enhanced deglycosylation (Table II), it did not alter the retention time or the peak width of dGAGP, nor did it yield any minor peaks; this demonstrated stability of peptide bonds to HF at 0°C is noteworthy. However, HF-deglycosylation at room temperature (23°C) significantly decreased molecular size with a concomitant increase in heterogeneity (Fig. 3C); this gel filtration profile showed distinct polypeptide size classes which indicate a possible nonrandom cleavage of peptide bonds caused by the more severe room temperature HF-treatment.

Enzymic degradation of other HRGPs invariably produces very simple peptide maps indicative of a highly repetitive periodic structure (15, 20, 30). Seven amino acids account for 92 mol% of dGAGP (Table I). Thus, assuming from a characteristically simple amino acid composition that GAGP is like other HRGPs, even its somewhat larger peptide backbone

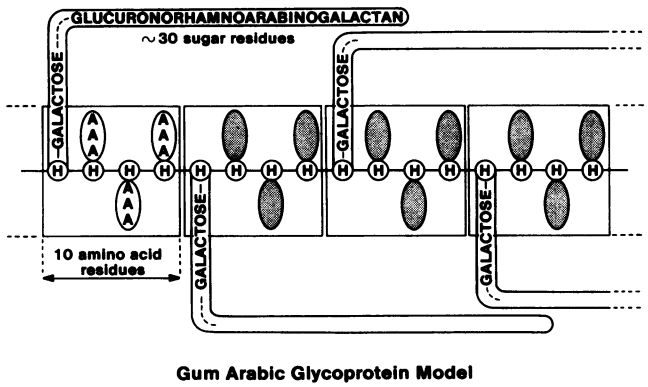

Figure 7. Hypothetical statistical model of the GAGP. A hypothetical repetitive block size of 7 kD contains 10 amino acid residues: 1 kD; 30 sugar residues: 4.44 kD; 3 Hyp-triarabinosides: 1.32 kD. The glucuronoarabinogalactan is probably a galactan backbone with glucuronic acid, rhamnose, and arabinose sidebranches.

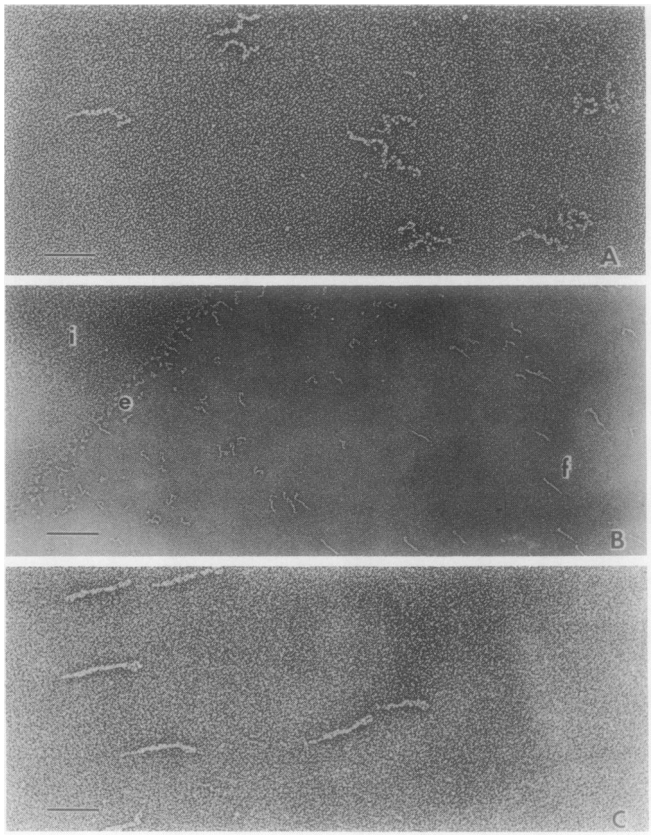

Figure 8. Transmission electron micrographs of Superose-purified GAGP after rotary shadowing. A, Native molecules showing random conformation; bar = 100 nm; B, droplet edge; bar = 300 nm; i: drop interior; e: drop edge; f: flow-polarized region; C, flow-polarized molecules; bar = 100 nm.

is probably a highly regular structure consisting of repetitive units. Based on the simple empirical formula: Hyp_4 Ser_2 Thr Pro Gly Leu His (Table I), together with the data (Fig. 1; Table I) showing every twelfth amino acid residue is polyglycosylated Hyp, we predict a 10 to 12 residue peptide motif as the basic repetitive unit of the GAGP polypeptide backbone. This rules out the repetitive Ala-Hyp domains which characterize AGPs (13), and shows that the GAGP polypeptide backbone is much more like a serine-rich basic extensin than an alanine-rich AGP.

However, in one respect, notably the Hyp-polysaccharide posttranslational modification, GAGP resembles AGPs rather than extensin! The polysaccharide substituent has an average of 30 sugar residues, corresponding to a polysaccharide sidechain with a 12-residue galactan backbone (with short glucuronic acid, rhamnose, and arabinose sidebranches) (Table VI), agreeing quite well with the 13-galactose residue backbone subunit postulated by Churms *et al.* (9).

The overall peptide-saccharide arrangement, represented as a statistical model in Figure 7, with a fundamental 7 kD subunit, would organize the polysaccharide sidechains in a highly orderly fashion, probably aligned along the long axis in the form of a twisted hairy rope. This close interaction between a polypeptide and its saccharide sidechains would

maximize intramolecular H-bonding, as suggested earlier (18) for the much smaller hydroxyproline arabinosides, but the longer (β1–3) galactan backbone, as a member of the "hollow helix family" (26) would be ideally suited to twine around a polyproline-II or similar extended helical backbone. A novel double helix!

Based on the above argument together with our very recent electron microscopy (Fig. 8) we conclude that GAGP contains a semiflexible (Hyp/Pro-rich) polypeptide backbone of >400 residues (~150 nm long, ~5 nm diameter; axial ratio: ~30:1) and is definitely rodlike (Fig. 8) rather than the spheroidal molecule postulated by the "wattle blossom" model for both AGPs (11) and the gum arabic protein complex (10). Overall the GAGP would therefore exhibit a very high tacticity or stereoregularity that might explain some of its interesting properties. Not the least of these is how large macromolecules, like gums and AGPs, migrate through a primary cell wall with a nominal molecular mass cutoff of 68 kD for globular proteins (8). On the other hand, the 4 to 5 nm porosity (8) would permit permeation by reptating ropelike rods!

ACKNOWLEDGMENTS

We wish to thank Dr. John Heckman for the electron microscopy, Dr. Marcia Kieliszewski for HPLC and amino acid analyses, and Mahmoud Pegah for computer systems. We have especially appreciated the interest and advice of Dr. Peter Wolf (Pepsico Inc.) who made this project possible.

LITERATURE CITED

1. **Akiyama Y, Eda S, Kato K** (1984) Gum arabic is a kind of arabinogalactan-protein. Agric Biol Chem **48:** 235–237
2. **Anderson DMW, Dea ICM** (1969) Chemotaxonomic aspects of the chemistry of *Acacia* gum exudates. Phytochemistry **8:** 167–176
3. **Anderson DMW, McDougall FJ** (1987) Degradative studies of gum arabic (*Acacia senegal* (L.) Willd.) with special reference to the fate of the amino acids present. Food Addit Contam **4:** 247–255
4. **Anderson DMW, Stoddart JF** (1966) Studies on uronic acid materials. Carbohydr Res **2:** 104–114
5. **Aspinall GO** (1970) III. Plant gums *In* W Pigman, D Horton, eds, The Carbohydrates, 2B. Academic Press, New York, pp 522–536
6. **Bhatti T, Chambers RE, Clamp JR** (1970) The gas chromatographic properties of biologically important *N*-acetylglucosamine derivatives, monosaccharides, disaccharides, trisaccharides, tetrasaccharides, and pentasaccharides. Biochim Biophys Acta **222:** 339–347
7. **Blumenkrantz N, Asboe-Hansen G** (1973) New method for quantitative determination of uronic acids. Anal Biochem **54:** 484–489
8. **Carpita N** (1982) Limiting diameters of pores and the surface structure of plant cell walls. Science **218:** 813–814
9. **Churms SC, Merrifield EH, Stephen AM** (1983) Some new aspects of the molecular structure of *Acacia senegal* gum (gum arabic). Carbohydr Res **123:** 267–279
10. **Connolly S, Fenyo JC, Vandevelde MC** (1988) Effect of a proteinase on the macromolecular distribution of *Acacia senegal* gum. Carbohydr Polym **8:** 23–32
11. **Fincher GB, Stone BA, Clarke AE** (1983) Arabinogalactan-proteins: structure, biosynthesis, and function. Annu Rev Plant Physiol **34:** 47–70
12. **Gammon DW, Stephen AM** (1986) Glycosylated hydroxyproline derivatives from *Acacia erioloba* exudate. Carbohydr Res **154:** 289–295
13. **Gleeson PA, McNamara M, Wettenhall REH, Stone BA, Fincher GB** (1989) Characterization of the hydroxyproline-rich protein core of an arabinogalactan-protein secreted from suspension-cultured *Lolium multiflorum* (Italian ryegrass) endosperm cells. Biochem J **264:** 857–862
14. **Heckman JW, Terhune BT, Lamport DTA** (1988) Characterization of native and modified extensin monomers and oligomers by electron microscopy and gel filtration. Plant Physiol **86:** 848–856
15. **Kieliszewski MJ, Leykam JF, Lamport DTA** (1990) Structure of the threonine-rich extensin from *Zea mays*. Plant Physiol **92:** 316–326
16. **Lamport DTA** (1965) The protein component of the primary cell wall. Adv Bot Res **2:** 151–218
17. **Lamport DTA** (1967) Hydroxyproline-O-glycosidic linkage of the plant cell wall glycoprotein extensin. Nature **216:** 1322–1324
18. **Lamport DTA** (1980) Structure and function of plant glycoproteins. Biochem Plants **3:** 501–541
19. **Lamport DTA, Miller DH** (1971) Hydroxyproline arabinosides in the plant kingdom. Plant Physiol **48:** 454–456
20. **Li XB, Kieliszewski MJ, Lamport DTA** (1990) A Chenopod extensin lacks repetitive tetrahydroxyproline blocks. Plant Physiol **92:** 327–333
21. **Miller DH, Lamport DTA, Miller M** (1972) Hydroxyproline heterooligosaccharides in *Chlamydomonas*. Science **176:** 918–920
22. **Mort AJ, Lamport DTA** (1977) Anhydrous hydrogen fluoride deglycosylates glycoproteins. Anal Biochem **82:** 289–309
23. **Mort AJ, Komalavilas P, Rorrer GL, Lamport DTA** (1989) Anhydrous hydrogen fluoride and cell wall analysis. Mod Meth Plant Anal **10:** 39–69
24. **Pope DG** (1977) Relationships between hydroxyproline-containing proteins secreted into the cell wall and medium by suspension-cultured *Acer pseudoplatanus* cells. Plant Physiol **59:** 894–900
25. **Randall RC, Phillips GO, Williams PA** (1988) The role of the proteinaceous component on the emulsifying properties of gum arabic. Food Hydrocolloids **2:** 131–140
26. **Rees DA** (1977) Polysaccharide Shapes. Chapman & Hall, London pp 1–80
27. **Sanger MP, Lamport DTA** (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
28. **Showalter A, Varner JE** (1989) Plant hydroxyproline-rich glycoproteins. Biochem Plants **15:** 485–520
29. **Smith F** (1939) The constitution of arabic acid. Part I. The isolation of 3-d-galactosido-1-arabinose. J Chem Soc 744–753
30. **Smith JJ, Muldoon EP, Willard JJ, Lamport DTA** (1986) Tomato extensin precursors P1 and P2 are highly periodic structures. Phytochemistry **25:** 1021–1030
31. **Vandevelde MC, Fenyo JC** (1985) Macromolecular distribution of *Acacia senegal* gum (gum arabic) by size-exclusion chromatography. Carbohydr Polym **5:** 251–271