JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

New Insights into the Structural Characteristics of the Arabinogalactan−Protein (AGP) Fraction of Gum Arabic

T. MAHENDRAN,[†] P. A. WILLIAMS,*[,†] G. O. PHILLIPS,[†] S. AL-ASSAF,[†] AND T. C. BALDWIN[§]

Glyndwr University, Plas Coch, Mold Road, Wrexham LL11 2AW, United Kingdom, and School of Applied Sciences, University of Wolverhampton, Wolverhampton, Wulfruna St WV1 1SB, United Kingdom

The structural characteristics of the gum exudate of *Acacia senegal* (gum arabic) have been investigated by monitoring the composition and physicochemical properties before and after treatment with proteolytic enzyme and various alkaline systems. Molecular mass (M_w) and radius of gyration (R_g) measurements were performed using gel permeation chromatography (GPC) coupled to refractive index, UV absorbance, and multiangle light scattering detectors and indicated that the macromolecules present have a compact structure. It was found that treatment with proteolytic enzyme caused the arabinogalactan−protein component (AGP) with average molecular mass $\sim 2 \times 10^6$ Da to degrade, yielding material of molecular mass $\sim 4 \times 10^5$ Da, whereas the bulk of the material corresponding to the protein-deficient arabinogalactan component (AG) with molecular mass 4×10^5 remained unaffected. Barium hydroxide was found to hydrolyze the polysaccharide component (AG) itself in addition to the proteinaceous component as demonstrated in control experiments using dextran. However, sodium borohydride/sodium hydroxide treatments were unable to hydrolyze dextran and were assumed to hydrolyze only the proteinaceous component of gum arabic. The AGP component was completely degraded, yielding material of molecular mass $\sim 4.5 \times 10^4$ Da. It has been concluded, therefore, that the enzyme did not fully hydrolyze all of the protein present and that the AGP component of gum arabic consists of carbohydrate blocks of $\sim 4.5 \times 10^4$ Da linked to a polypeptide chain consistent with the wattle blossom structure. Because the AGP was degraded to differing extents using a mild and more severe sodium borohydride/sodium hydroxide treatment, it was concluded that the polysaccharide moieties were linked through both *O*-serine and *O*-hydroxyproline residues. The gum arabic sample was deglycosylated by treatment with anhydrous hydrogen fluoride and revealed the presence of two putative core proteins of $\sim 3 \times 10^4$ and $\sim 5 \times 10^3$ Da, respectively, which correspond to proteins of approximately 250 and 45 amino acids in length. A new model for the structure of the AGP component has been proposed.

KEYWORDS: Gum arabic structure; *Acacia senegal*; arabinogalactan protein; gel permeation chromatography; multiangle light scattering; wattle blossom model; alkaline hydrolysis; deglycosylation

INTRODUCTION

Gum arabic occurs as a sticky liquid that exudes from the stems and branches of acacia trees (*Acacia senegal* and *Acacia seyal*), which grow across the Sahelian belt of Africa, principally Sudan (*1*−*4*). It is produced when the trees are subjected to stress conditions such as heat, drought, or wounding. The gums from *A. senegal* and *A. seyal* consist of galactose, arabinose, rhamnose, and glucuronic acid and have a very highly branched structure (*1*−*11*). They also contain a small proportion of proteinaceous material (*12*−*15*). The molecular mass of gum arabic varies with the source, but a typical average value for gum from *A. senegal* is ∼380,000, whereas the gum from *A. seyal* has an average molecular mass of ∼850,000 (*6*−*8, 16, 17*). Most work on the chemical and physicochemical characterization of gum arabic has concentrated on *A. senegal* gum. Anderson et al. (*5*) subjected the gum to a series of Smith degradation procedures and concluded that the core was a highly branched unit of β 1,3-linked galactose residues with extensive branching at the C6 position. The branches were shown to consist of galactose, arabinose, rhamnose, and glucuronic acid. Churms et al. (*18*−*20*) also carried out Smith degradation studies and concluded that the galactose core consisted of 13 β 1,3

* Corresponding author (e-mail williamspa@newi.ac.uk).
[†] Glyndwr University.
[§] University of Wolverhampton.

10.1021/jf800849a CCC: $40.75
Published on Web 09/11/2008

galactose residues having two branch points giving a repeating subunit with a molecular mass of 8000 Da. Anderson and Stoddart fractionated the gum using sodium sulfate, and subsequent analysis showed that a significant amount of protein was present associated with a high molecular mass fraction and that most of the gum was found to contain very little protein (*21*). Akiyama et al. (*12*) showed that both hydroxyproline and serine residues were present and that a proportion was covalently attached to carbohydrate moieties. They and others have also found that the gum precipitated with β-glucosyl Yariv reagent, confirming that it is a type of arabinogalactan−protein complex (AGP) (*12, 15, 22, 23*). AGPs commonly occur at the cell surface in plants and have a role in plant growth and development (*24*−*26*). They generally consist of a hydroxyproline-rich protein core, which is decorated by galactose- and arabinose-rich polysaccharide units O-linked through hydroxyproline and possibly serine and threonine (*26*−*28*). Some AGPs also contain N-linked polysaccharides. Isolated hydroxyproline residues appear to be the point of attachment for polysaccharide chains, whereas clusters of hydroxyproline are the sites of attachment for oligoarabinose chains containing four to six residues (*26, 28*).

GPC analysis has shown that the gum consists of three main molecular mass fractions, which have varying protein contents (*6, 8, 16, 17, 29*). The main component, which we will refer to as the arabinogalactan fraction (AG), represents ~90% of the gum and contains <1% protein. It has a molecular mass of $\sim 2.5 \times 10^5$ with $R_g \sim 10$ nm and $R_h \sim 15$ nm determined by GPC/MALLs and dynamic light scattering, respectively. Sanchez et al. (*11*) have recently shown by small-angle neutron scattering experiments, together with transmission electron microscopy and atomic force microscopy, that this fraction has a disk-like morphology with a diameter of 20 nm and a thickness below 2 nm. A second component, which we will refer to as the arabinogalactan−protein fraction (AGP), represents ~10% of the total, contains ~10% protein, has a molecular mass of $1-2 \times 10^6$, $R_g \sim 30-35$ nm, and $R_h \sim 45$ nm. Because the molecular mass of this fraction was found to decrease on treatment with proteolytic enzyme from $\sim 1.5 \times 10^6$ to 2.5×10^5, it has been suggested that it has a wattle blossom type structure typical of AGPs (*12*) with five/six large carbohydrate blocks linked to a polypeptide chain (*30*). The third component, which we will refer to as glycoprotein (GP), represents only ~1% of the total, contains up to 50% protein, and is not degraded by proteolytic enzyme, which is typical for many AGPs (*31*). The GP fraction has a molecular mass of $\sim 2 \times 10^5$, $R_g \sim 10$ nm, and $R_h \sim 10$ nm. The neutral sugar contents of the three components have been reported to be similar (*31*); however, the protein-rich fractions were found to contain significantly less glucuronic acid (*31*). Renard et al. (*10*) undertook circular dichroism studies on the various fractions and concluded that the AGP and GP components have polyproline II, β-sheet, and random coil secondary structures, whereas there was no secondary structure proposed for the AG component. The highly branched nature of the carbohydrate component for each of the fractions was confirmed by NMR and methylation analysis (*31*). Qi et al. (*32*) isolated the AGP fraction by preparative GPC and subjected it to deglycosylation using hydrofluoric acid. They found that the remaining protein consisted of ~400 amino acids containing ~130 hydroxyproline residues with a 10−12 residue repetitive motif. Further work by Goodrum et al. (*33*) suggested a 19 residue consensus motif containing six Hyp-arabinosides and two Hyp-polysaccharides. Qi et al. (*32*) performed transmission electron microscopy experiments on the polypeptide, which indicated that it was ~150 nm long and was rod-like. Following alkaline hydrolysis experiments using barium hydroxide, they concluded that the carbohydrate was attached as an O-linked substituent to the peptide through hydroxyproline. Whereas 12.1% of the hydroxyprolines were nonglycosylated, 63.5% were linked to short-chain oligomers (~3 sugar units) and 24.3% to carbohydrate blocks consisting of about 30 sugar residues corresponding to a molecular mass of 4.44×10^3 Da. It was concluded that the molecules resembled a "twisted hairy rope". Akiyama et al. (*12*) reported similar percentages of nonglycosylated hydroxyproline and hydroxyproline linked to arabinose chains and larger carbohydrate blocks (*12*) but also reported the presence of serine−carbohydrate links. Qi et al. (*32*) estimated the molecular mass of the high molecular mass AGP fraction to be 2.2×10^5, which is significantly less than values reported for the whole gum itself (*6, 8, 16, 17*). Furthermore, the concept that the molecules are rod-like is inconsistent with light scattering and viscometric data, which indicate the molecules adopt a compact conformation (*6*−*8, 16, 17*). The linear polypeptide chains observed by transmission electron microscopy are likely to adopt a very different conformation in solution. Transmission electron microscope studies on purified AGPs from other sources have also indicated a compact shape (*34, 35*).

This paper further investigates the structural characteristics of gum arabic by subjecting it to both enzymatic and alkaline hydrolysis and HF-deglycosylation and analyzing the products by GPC/MALLS and SDS-PAGE.

Table 1. Amino Acid (AA) Content of Gum Arabic (GA) and Gum Arabic Treated (GAT) with 0.2 M $NaBH_4$/0.5 M NaOH

AA	GA (nmol/mg)	GAT (nmol/mg)
hydroxyproline	54.200	43.300
aspartic acid	10.600	6.740
threonine	15.900	9.490
serine	28.700 (Hyp:Ser = 1.88:1)	15.600 (Hyp:Ser = 2.77:1)
glutamic acid	8.290	6.510
proline	15.600	11.100
glycine	10.600	8.120
alanine	5.070 (Hyp:Ala = 10.69:1)	5.180 (Hyp:Ala = 8.35:1)
cysteine	0.000	0.000
valine	7.290	5.500
methinine	0.110	0.220
isoleucine	2.380	1.450
leucine	15.100	10.800
tyrosine	2.300	1.530
phenylalanine	6.330	4.440
histidine	10.700	6.850
tryptophan	0.000	0.000
lysine	5.130	3.310
arginine	2.120	1.170
total	201.00	141.00
% of AA in total composition	2.15	1.51

MATERIALS AND METHODS

Materials. A kibbled gum arabic sample originating from *A. senegal* species was obtained from Agrisales Ltd., London, U.K. The sugar composition was determined by HPLC as described previously (*16*), and the sample was found to contain 40% arabinose, 34% galactose, and 12% rhamnose, which is consistent with gum from *A. senegal*. The sample was also subjected to amino acid analysis (Alta Bioscience, University of Birmingham), and the results are reported in **Table 1** (column 2). As has been found previously (*3*) the main amino acids were found to be hydroxyproline, serine, proline, threonine, leucine, glycine, and histidine. The protein content was calculated to be 2.15%.

Table 2. Elemental Analysis of Gum Arabic

	Al	Ba	Ca	K	Mg	Mn	P	S	Sr
ppm	4.11	58.05	6642.5	7230.5	2196.5	8.375	3.2	18	38.55

Elemental analysis by ICP-OES was undertaken, and the results reported in **Table 2** show the presence of a number of metal cations, notably potassium, calcium, and magnesium as has been reported by others (*3*). Trace amounts of phosphorus and sulfur were also detected. Yavdav et al. (*36*) also reported similar levels of phosphorus and attributed them to the presence of glycosylphosphatidylinositol (GPI)-anchored proteins, which occur in other AGPs (*36*).

The dextran was obtained from Sigma (lot 22k1194 produced by *Leuconostoc mesenteroides*) and was quoted to have a molecular mass of 4.021×10^4.

Methods. *Molecular Mass Distribution.* The molecular mass distribution of gum arabic was determined by GPC using a Waters (Division of Millipore) solvent delivery system model 600A or P-500 dual piston syringe pump (Pharmacia Biotech), a Rheodyne series 7125 injector with a 100 μL loop, and a column packed with Superose 6HR 10/30 (with bead diameter of 11−15 μm). A value of 0.141 cm^3/g was used for the refractive index increment (d*n*/d*c*) (*37*). The Dawn DSP laser light scattering photometer equipped with a 632.8 nm He−Ne laser (Wyatt Technology Corp.) with 15 detectors was used in conjunction with a Wyatt Optilab DSP interferometric refractometer operated at 632.8 nm equipped with a 10 mm P100 cell (Wyatt Technology Corp.) and a UV−visible spectrophotometer (Agilent) at 214 nm. Data accumulation for detectors used Wyatt Technology ASTRA 4.5 software. Measurements were carried out at room temperature, and samples were prepared in 0.2 M NaCl. Accurately weighed 2.0 mg/mL samples were filtered through 0.45 μm nylon filters and injected and analyzed.

Enzyme Hydrolysis. Kibbled gum arabic sample (0.0226 g) was dissolved in 9.0 mL of $NaHPO_4/Na_2HPO_4$ buffer at pH 7.5 and was treated with protease (Type XIV from *Streptomyces griseus* protease EC 1.0 cm^3 of 1.6 mg/mL, Sigma). The samples were incubated at 37 °C in a stoppered test tube. Aliquots (2.0 cm^3) of the protease-treated gum were removed at intervals (21, 48, 72, 105, and 130 h). The resulting solutions were analyzed by GPC/MALLS as described above.

Alkaline Hydrolysis Using Barium Hydroxide. One gram of gum arabic was treated with 0.22 M $Ba(OH)_2$ for 18 h at 50 °C. Excess $Ba(OH)_2$ was neutralized using 0.5 M H_2SO_4 in an ice bath (*9, 32, 33*), and the resulting mixtures were centrifuged at 2500 rpm for 2 h at 25 °C. The supernatant was recovered and dialyzed for 72 h and freeze-dried. The resulting product (2 mg/mL) was prepared in 0.2 M NaCl and analyzed by GPC/MALLS. A control experiment was carried out using dextran.

Alkaline Hydrolysis Using NaOH/$NaBH_4$. Two procedures were followed on treating gum arabic with sodium borohydride and sodium hydroxide, using 0.2 M $NaBH_4$/0.5 M NaOH (*12*) and 2 M $NaBH_4$/1 M NaOH (*38*−*40*). The former cleaves *O*-glycans at polysaccharide−polypetide linkages involving Ser or Thr but not those involving Hyp (nor hydrolysine, which is not present in gum arabic). The latter also does not cleave glycan−Hyp linkages but is strong enough to cleave peptide bonds and thus release glycans that still have a Hyp at the reducing terminus.

Treatment of Gum Arabic with 0.5 M NaOH/0.2 M $NaBH_4$. One gram of gum arabic was treated with 100 mL of 0.2 M $NaBH_4$ and 0.5 M NaOH for 18 h at 50 °C. Excess $NaBH_4$ was neutralized using 0.5 M H_2SO_4 in an ice bath. The resulting mixture was centrifuged at 2500 rpm for 1 h at 25 °C, and the supernatant was dialyzed for 24 h and freeze-dried. The resulting product (2.0 mg/mL) was prepared in 0.2 M NaCl and analyzed by GPC/MALLS. A control experiment was performed with dextran.

Treatment of Gum Arabic with 1 M NaOH/2 M $NaBH_4$. One gram of gum arabic was treated with 100 mL of 2 M $NaBH_4$ and 1.0 M NaOH for 6 h at 100 °C. Excess $NaBH_4$ was neutralized using 1 M acetic acid in 5 volumes of methanol in an ice bath. The resulting solid sample was washed three times with methanol, and the product was dried at 40 °C for 24 h. The resulting product (2.0 mg/mL) was prepared in 0.2 M NaCl and analyzed by GPC/MALLS. A control experiment was performed with dextran.

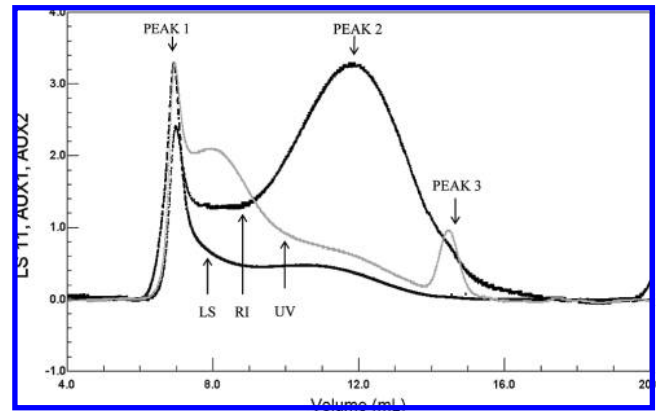

Figure 1. GPC elution profiles of gum arabic obtained using UV (214 nm), RI, and light scattering detectors.

Determination of the Molecular Mass of the Polypeptide Chain Backbone. Gum arabic was deglycosylated by Professor A. Mort, University of Oklahoma, by treatment with hydrofluoric acid (HF) (*41*), and the resulting polypeptide was loaded onto a 15% SDS-PAGE gel that was silver-stained according to the method of Ansorge (*42*).

RESULTS AND DISCUSSION

Molecular Mass Distribution. The GPC elution profiles of the gum arabic sample using light scattering (LS), RI, and UV (at 214 nm) is shown in **Figure 1**. RI is a sensitive measure of concentration, whereas UV absorbance is sensitive also to the chemical nature of the eluting species, particularly the proteinaceous component. Light scattering is sensitive to concentration and molecular mass. As shown in **Figure 1** the RI elution profile consists of two main peaks, the major peak (peak 2), which corresponds to about 90% of the gum (the AG fraction), and a higher molecular mass peak (peak 1), which corresponds to about 10% (the AGP fraction). This fraction elutes at the void volume of the column and is not fully resolved. The elution profiles obtained by UV are very different from the refractive index profiles. Two of the peaks correspond to peaks 1 and 2 observed by RI, although the relative intensity of each is different due to differences in protein content as has been reported previously (*15, 16*). The peak at an elution volume of ~8 mL also corresponds to the AGP fraction (*16*). In addition there is a further peak, peak 3, which corresponds to the GP fraction. **Figure 2** shows the M_w and R_g profiles as a function of elution volume. The average M_w and R_g values for the AGP were determined using the Astra software and found to be $\sim 2 \times 10^6$ and ~45 nm, respectively. The AG was found to have a molecular mass of $\sim 4 \times 10^5$ and the GP a molecular mass of $\sim 2 \times 10^5$. The results are consistent with previous findings for other samples and confirm that the molecules have a compact structure (*1*).

Enzyme Hydrolysis. Although enzyme degradation studies have been reported in the literature previously (*15*), the studies have generally been undertaken over a relatively short period of time, and it is possible that degradation was incomplete. The present study investigates degradation over a period of 130 h. The effect of protease on the molecular mass distribution of the gum sample is illustrated in **Figures 3** and **4**. The RI elution profiles (**Figure 3**) after treatment with protease show a systematic decrease in the intensity of peak 1, which corresponds to the high molecular mass AGP component. The peak corresponding to the AG fraction (peak 2) shows a small decrease in intensity. The reduction in peak intensities is accompanied

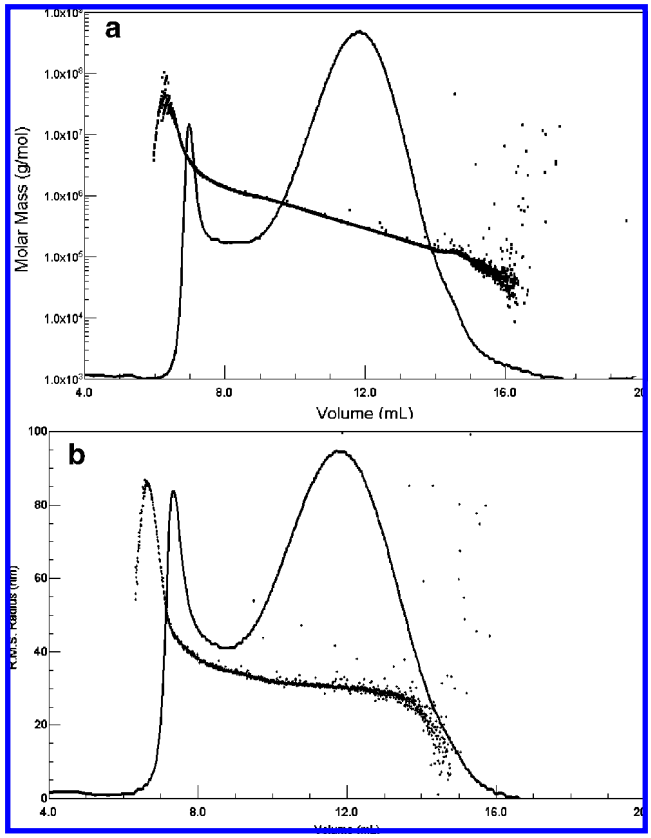

Figure 2. (**a**) GPC RI and M_w elution profiles for gum arabic. (**b**) GPC RI and R_g elution profiles for gum arabic.

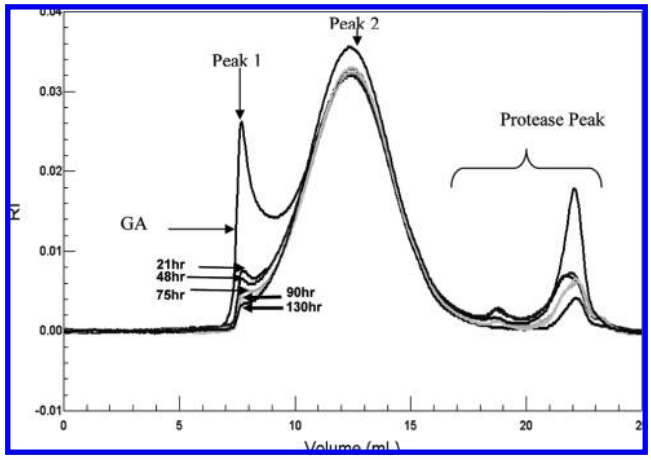

Figure 3. GPC RI elution profile of gum arabic on treatment with proteolytic enzyme for various periods of time.

by the development of a small peak at an elution volume of 19.0 mL and by an increase in the intensity of the peak at 22 mL, which corresponds to the salt elution peak. It is noted that the enzyme itself also elutes between 17 and 22 mL, and this is more evident by comparison of the UV chromatographs (**Figure 4**) for the gum arabic alone and with enzyme added. The UV elution profiles clearly show that the treatment with protease gives rise to a sudden decrease in the intensity of peak 1 and a gradual increase in the intensities of peaks 2 and 3. It is very interesting that peak 3, which is a highly protein-rich component, shows a marked resistance to hydrolytic degradation by protease and remains intact after the enzyme treatment. This may be due to the fact that this fraction has a very different amino acid composition from the AGP and AG fractions, with aspartic acid, phenylalanine, serine, glutamic acid, glycine, valine, and leucine

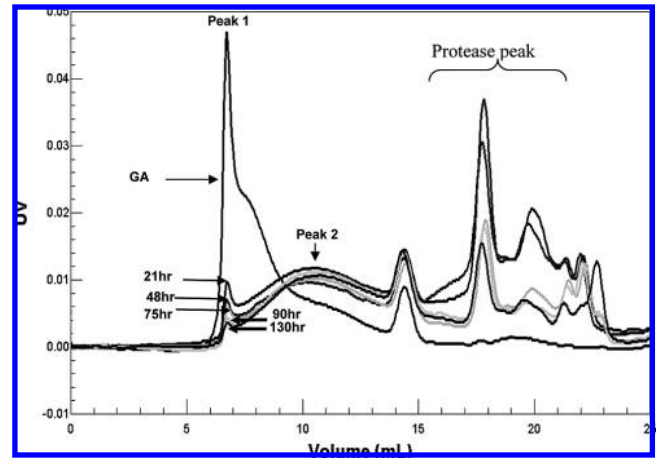

Figure 4. GPC UV (214 nm) elution profile of gum arabic on treatment with proteolytic enzyme for various periods of time.

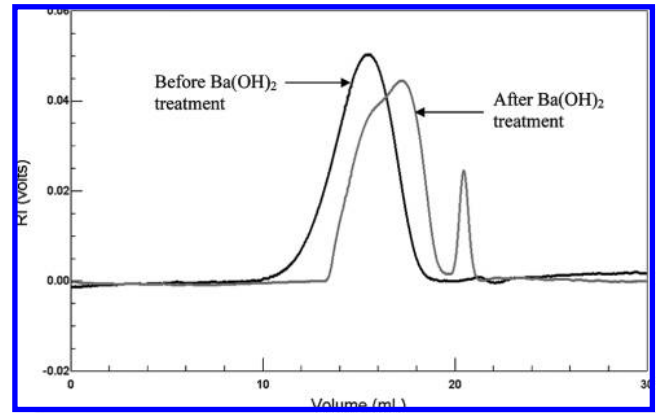

Figure 5. GPC RI elution profiles for dextran before and after treatment with 0.22 M $Ba(OH)_2$.

being the main components (*14*). Similar results have been obtained by Osman et al. (*15*) on gum arabic fractions, although their studies were over a shorter period of up to 48 h.

The results demonstrate that the high molecular mass AGP component is hydrolyzed by the enzyme and its molecular mass decreases from $\sim 2 \times 10^6$ to $\sim 4 \times 10^5$ (peak 2), indicating that large carbohydrate blocks are linked to a polypeptide chain in accordance with the "wattle blossom" type structure (*3*). The specific rotation of the solution was found to remain constant after the degradation, and therefore the chemical structure of the gum must be unaltered.

Alkaline Hydrolysis Using Barium Hydroxide. Figure 5 shows the elution profiles of dextran before and after treatment with 0.22 M $Ba(OH)_2$ as used by Qi et al. in their studies on gum arabic (*32*). The profiles clearly show that hydrolysis of the polysaccharide chains occurs and that the molecular mass of dextran is significantly decreased by the treatment. It is possible that depolymerization occurs due to alkali-catalyzed β-elimination. The RI and UV elution profiles for gum arabic before and after treatment with 0.22 M $Ba(OH)_2$ are shown in panels **a** and **b** of **Figure 6** and indicate that both the AGP and AG components are degraded. The fact that the RI and UV profiles are more or less superimposable after treatment indicates that they are chemically very similar and hence that the proteinaceous moieties have been completely hydrolyzed and were removed during the dialysis step. The results confirm the previous work of Osman et al. (*15*) and cast doubt on the conclusions drawn by Qi et al. (*32*), who assumed that the polysaccharide moieties would not be degraded. Sugar analysis

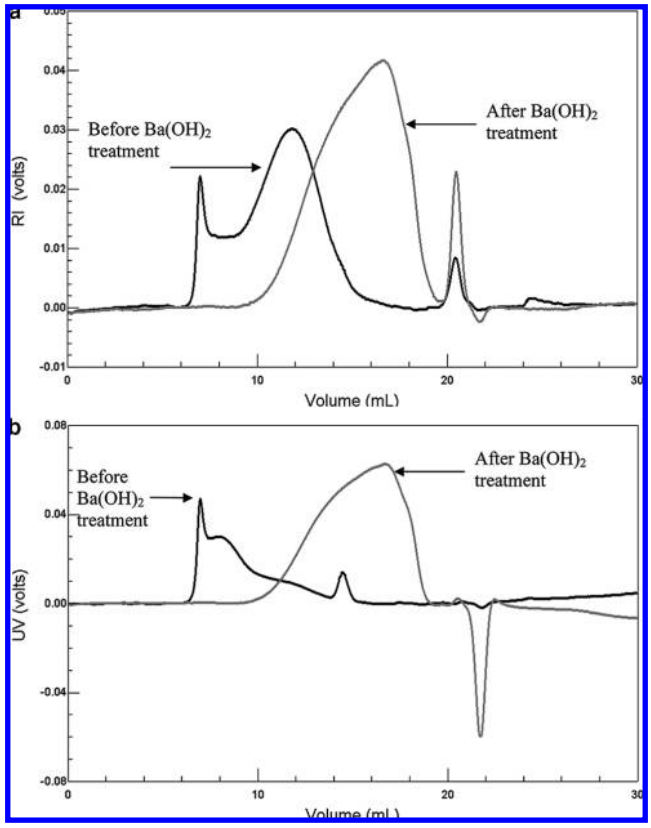

Figure 6. (**a**) GPC RI elution profiles for gum arabic before and after treatment with 0.22 M $Ba(OH)_2$. (**b**) GPC UV (214 nm) elution profiles for gum arabic before and after treatment with 0.22 M $Ba(OH)_2$.

Table 3. Sugar Composition of GA before and after Treatment with 0.2 M $NaBH_4$/ 0.5 M NaOH and 0.22 M $Ba(OH)_2$[a]

sample	% w/w arabinose	% w/w rhamnose	% w/w galactose
gum arabic	40	12	34
GA/0.2 M $NaBH_4$/NaOH	37	15	31
GA/0.22 M $Ba(OH)_2$	29	9	41

[a] The total is <100% as it does not include the glucuronic acid, protein, and metal ion contents.

on the gum arabic after hydrolysis (**Table 3**) indicates that the proportion of arabinose and rhamnose remaining has decreased, indicating that they are at the periphery of the molecules as suggested by Anderson et al. (*2*) and/or there are hydroxyproline−oligoarabinosides linkages as found by Qi et al. (*32*) and Akiyama et al. (*12*).

Alkaline Hydrolysis by NaOH/$NaBH_4$. Figure 7 shows the RI elution profiles for dextran before and after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h and 2 M NaOH/1 M $NaBH_4$ at 100 °C for 6 h. It is noted that the profiles are unaffected by either of the treatments, indicating that the dextran is not degraded to any extent. The peak at an elution volume of ~21 mL is the salt peak.

Panels **a** and **b** of **Figure 8** show the RI and molecular mass elution profiles and **Figure 9** shows the UV elution profiles of gum arabic before and after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h and 2 M NaOH/1 M $NaBH_4$ at 100 °C for 6 h. The RI profiles in **Figure 8a** clearly demonstrate that the high molecular mass AGP peak (peak 1) disappears and that an additional peak appears at a higher elution volume (peak 4). **Figure 8b** confirms that this component has a molecular mass of ~4.5×10^4 Da. Peak 2, which corresponds to the AG component, remains more or less the same, which

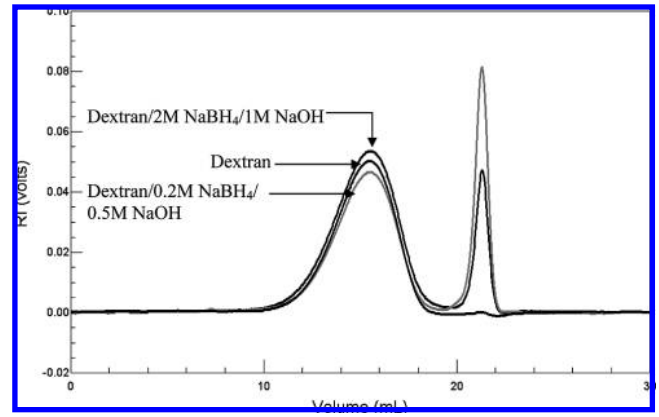

Figure 7. GPC RI elution profiles of dextran before and after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h and with 2 M NaOH/1 M $NaBH_4$ at 100 °C for 6 h.

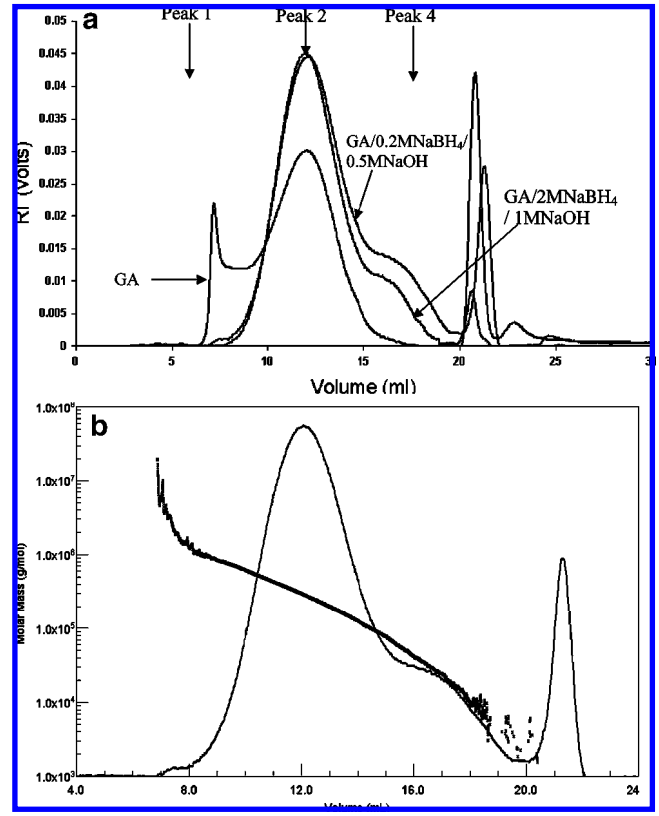

Figure 8. (**a**) GPC RI elution profiles of gum arabic before and after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h and with 2 M NaOH/1 M $NaBH_4$ at 100 °C for 6 h. (**b**) GPC RI and molecular mass elution profiles of gum arabic after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h.

may be indicative of different links between the polypeptide and the carbohydrate moieties. This is supported by the fact that the relative compositions of the main amino acids are slightly different in the AGP (Hyp_4 Ser_2 Thr_1 Pro_1 Leu_1 $Hist_1$) compared to the AG (Hyp_3 Ser_3 Thr_2 Glu_2 Pro_2 $Glyc_2$ Asp_1 Arg_1 Leu_1 Ala_1) (*14*). The sodium hydroxide/sodium borohydride treatments are expected to hydrolyze any proteinaceous components present but are not expected to degrade the carbohydrate components. This is consistent with the fact that the protein-rich AGP component degrades, whereas the protein-deficient AG component does not. As noted previously, the treatment with 0.2 M $NaBH_4$/0.5 M NaOH cleaves *O*-glycans at polysaccharide−polypeptide linkages involving Ser or Thr but

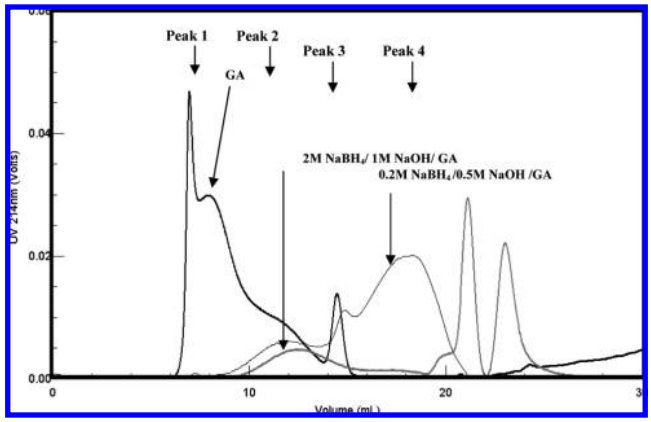

Figure 9. GPC UV (214 nm) profiles of gum arabic before and after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h and 2 M NaOH/1 M $NaBH_4$ at 100 °C for 6 h.

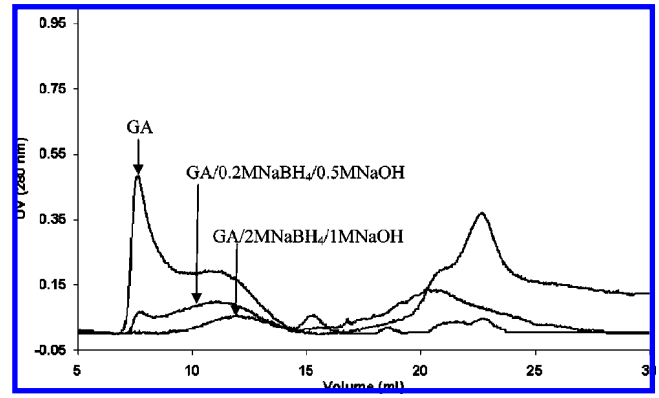

Figure 10. GPC UV (280 nm) profiles of gum arabic before and after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h and with 2 M NaOH/1 M $NaBH_4$ at 100 °C for 6 h.

not those involving Hyp. The predicted consensus sequence for N-linked addition of glycans is (Asn-Xaa-Ser/Thr, where Xaa is any amino acid except Pro), but such linkages are rare in AGPs. However, the treatment with 2 M $NaBH_4$/1 M NaOH is expected to cleave all such links. The fact that the AGP is degraded by the milder treatment confirms the presence of *O*-serine−polysaccharide linkages, consistent with the findings of Akiyama et al. (*12*).

The AGP fraction degrades from a molecular mass of ~2 × 10^6 to 4.5 × 10^4 Da. This, therefore, suggests that the AGP consists of carbohydrate blocks of 4.5 × 10^4 Da linked to the polypeptide chain and contradicts the enzyme degradation studies described above and reported by others, which concluded that the carbohydrate blocks are of the order of ~(2.5−4.5) × 10^5 Da. Clearly the enzyme is unable to fully degrade the proteinaceous components present.

The UV elution profiles at 214 nm (**Figure 9**) also demonstrate that the AGP component (peak 1) and that some of the proteinaceous elements corresponding to the AG component (peak 2) are degraded by treatment with NaOH/$NaBH_4$. It is noted that the new peak formed (peak 4) is significantly more intense for the sample subjected to the milder alkaline treatment, suggesting that it still contains proteinaceous elements. As noted above, the milder treatment is expected to hydrolyze O-linkages to serine and hence indicates that the material at peak 4 contains other types of protein−carbohydrate links, presumably *O*-hydroxyproline as reported by Qi et al. (*32*). Peak 3, which corresponds to the GP with a high protein content, is not degraded by the milder 0.2 M NaOH/0.5 M $NaBH_4$ treatment but is degraded by the more severe 2 M $NaBH_4$/1 M NaOH treatment, also demonstrating the presence of other amino acid−polysaccharide linkages.

Figure 10 shows the elution profiles monitored by UV absorbance at 280 nm of gum arabic before and after treatment with 0.2 M NaOH/0.5 M $NaBH_4$ at 50 °C for 18 h and with 2 M NaOH/1 M $NaBH_4$ at 100 °C. Measurements at this wavelength are sensitive only to the protein present. For the milder treatment with 0.2 M $NaBH_4$/0.5 M NaOH, it is expected that β-elimination occurs through *O*-serine linkages as discussed above; however, significant proportions of AGP (peak 1) and AG (peak 2) remain, indicating that O-links through lysine, hydroxyproline, or hydroxylysine and/or N-links must be present. Amino acid analysis after hydrolysis (see **Table 1**, column 2) shows a decrease in serine and an increase in alanine (compared to hydroxyproline), confirming serine−carbohydrate linkages as proposed by Akiyama et al. (*12*). Qi et al. (*32*) reported only the presence of hydroxyproline linkages. A

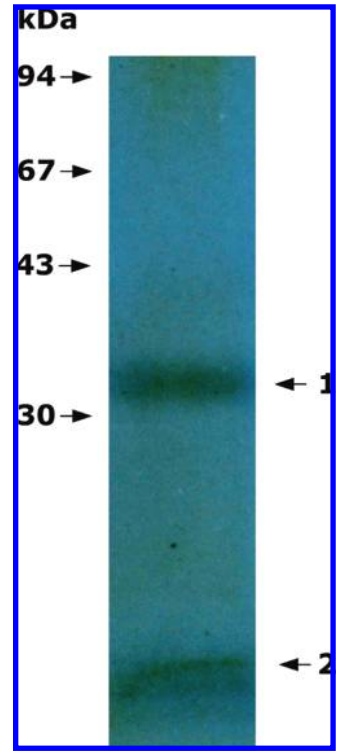

Figure 11. SDS-PAGE analysis of HF-deglycosylated gum arabic. The numbers on the left are the molecular masses of protein markers.

significant amount of proteinaceous material elutes between 16 and 20 mL, which corresponds to peak 4, and also at elution volumes corresponding to the salt peak. On more severe treatment with 2 M $NaBH_4$/1 M NaOH, which cleaves both N-linked and O-linked proteins, most of the protein moieties elute close to the salt peak, indicating they are present in the form of small peptide chains and/or single amino acids.

Deglycosylation by HF Reveals Two Putative Core Polypeptides. Figure 11 shows the SDS-PAGE results obtained for the gum arabic sample deglycosylated using HF. The results indicate that there are possibly two core protein bands present. The highest molecular mass band (band 1) is a fuzzy band (possibly still partially glycosylated) corresponding to a molecular mass of approximately 3 × 10^4 Da. A second, much more discrete, band (band 2) has a molecular mass of approximately 5 × 10^3 Da. This suggests the presence of two core polypeptides of ~250 and ~45 amino acids, respectively. Qi et al. (*32*) determined that the gum arabic high molecular mass component contained a polypeptide chain of ~400 amino acids (*30*). Renard et al.

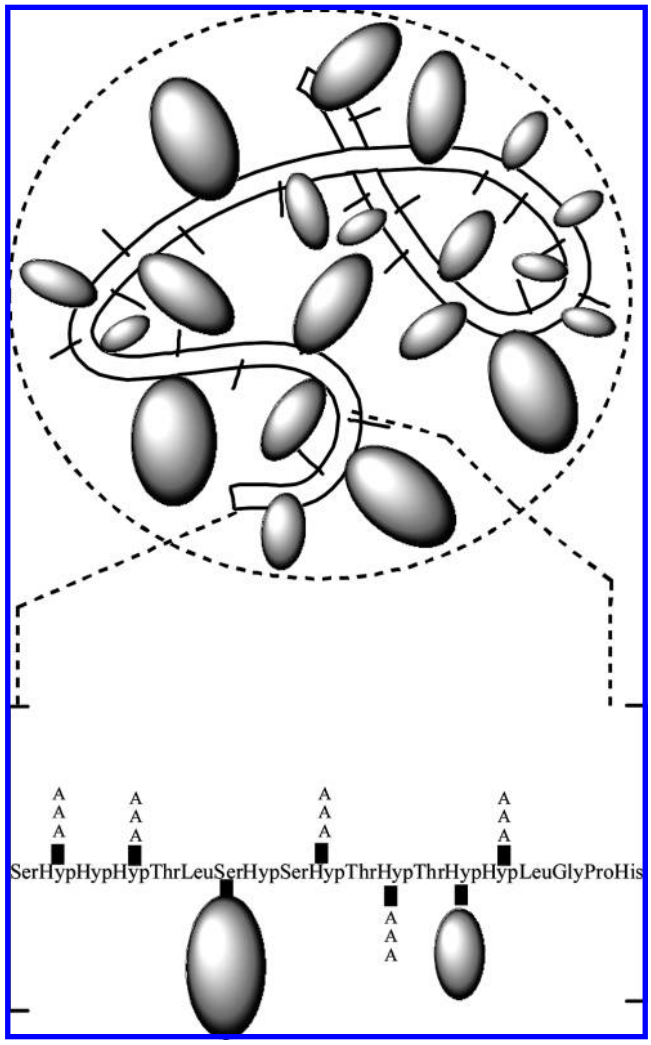

Figure 12. Schematic illustration of the structure of the gum arabic arabinogalactan protein complex. The molecule has a molecular mass of $(1-2) \times 10^6$ Da and consists of a polypeptide chain possibly containing ~250 amino acids with short arabinose side chains and much larger blocks of carbohydrate of molecular mass $\sim 4.0 \times 10^4$ Da and attached. The carbohydrate is highly branched (*31*) and may adopt the thin oblate ellipsoid structure proposed for the AG component. The amino acid sequence for the polypeptide chain has been determined by Goodrum et al. (*44*). The molecule adopts a very compact conformation with R_g of ~36 nm.

(*10*) estimated that the backbone for this fraction consisted of 2253 amino acids on the basis of its overall M_w and amino acid content and composition, assuming a single polypeptide chain. It is possible in the HF deglycosylation studies that the polypeptide chain length is hydrolyzed, or it could be that more than one polypeptide chain is present. Interestingly, Renard et al. (*10*) also estimated that the protein in the AG component consisted of 43 amino acids, which corresponds to our band 2. The presence of this short peptide indicates that the AG fraction of the gum may contain an AG−peptide(s). Such AG−peptides have recently been identified as members of the arabinogalactan−protein gene family in *Arabidopsis thaliana*. *Acacia* species may also possess and express genes that encode AG−peptides, which would explain the presence of such peptide sequences in the gum (*45*).

Conclusions. The AGP component of gum arabic has a molecular mass of $\sim(1-2) \times 10^6$ Da and an R_g of ~45 nm. It has a highly branched compact structure (*31*). Mild alkaline hydrolysis of the gum followed by GPC analysis has indicated that the AGP fraction consists of carbohydrate blocks of molecular mass $\sim 4.5 \times 10^4$ Da linked to a polypeptide chain. A significant number of the linkages are through serine residues with some linkages through hydroxyproline. SDS-PAGE studies on the HF-deglycoslylated whole gum sample have indicated that there are two putative core proteins present. The larger band occurs at around 3×10^4 Da, which corresponds to a core protein of approximately 250 amino acids and is consistent with the literature for AGP core proteins (*29*). This band is slightly fuzzy, which may indicate that it is still partially glycosylated. The second putative core protein is a more discrete band of around 5×10^3 Da, which would correspond to around 45 amino acids. This proteinaceous component may be associated with the AG gum fraction. The data presented for the high molecular mass AGP fraction are consistent with the wattle blossom model proposed for AGPs in which carbohydrate blocks are linked to a common polypeptide chain (*27*). Our results indicate that the carbohydrate blocks are of the order of 4×10^4 Da, much less than we assumed previously from enzyme degradation studies (*43*), but significantly greater than reported by Qi et al. (*32*). These carbohydrate blocks may have a thin oblate ellipsoid carbohydrate structure, which has been reported recently for the AG component (*11*). We propose that the structure of the AGP fraction can be schematically represented as shown in **Figure 12**.

LITERATURE CITED

(1) Williams, P. A.; Phillips, G. O. *Handbook of Dietary Fibre in Food Development*; Choo, S. S., Ed.; Dekker: New York, 2001; pp 675−694.
(2) Williams, P. A.; Phillips, G. O.; Stephen, A. M.; Churms, S. C. In *Food Polysaccharides and Their Applications*; Stephen, A. M., Phillips, G. O., Williams, P. A., Eds.; CRC Press Taylor and Francis Group: Boca Raton, FL, 2006; pp 455.
(3) Williams, P. A.; Phillips, G. O. Gum arabic In *Handbook of Hydrocolloids*; Phillips, G. O.; Williams, P. A., Eds.; Woodhead Publishing: Cambridge, U.K., 2000; pp 155.
(4) Islam, A. M.; Phillips, G. O.; Sljiyo, A.; Snowden, M. J.; Williams, P. A. *Food Hydrocolloids* **1997**, *11*, 493–506.
(5) Anderson, D. M. W.; Hirst, E. L.; Stoddart, J. F. *J. Chem. Soc. C* **1966**, 308.
(6) Assaf, S. A.; Phillips, G. O.; Williams, P. A. *Food Hydrocolloids* **2005**, *19*, 647–660.
(7) Flindt, C.; Osman, M. E.; Assaf, S. A.; Phillips, G. O.; Williams, P. A. *Food Hydrocolloids* **2005**, *19*, 687–702.
(8) Hassan, E. A.; Assaf, S. A.; Phillips, G. O.; Williams, P. A. *Food Hydrocolloids* **2005**, *19*, 669–678.
(9) Kieliszewski, M. J.; Goodrum, L. J.; Patel, A.; Leykam, J. F. *Phytochemistry* **2000**, *54*, 99–106.
(10) Renard, D.; Lavenant-Gourgeon, L.; Ralet, M.; Sanchez, C. *Biomacromolecules* **2006**, *7*, 2637–2649.
(11) Sanchez, C.; Schmitt, C.; Kolodziejczyk, E′.; Lapp, A.; Gaillard, C.; Renard, D. *Biophys. J.* **2008**, *94*, 629–639.
(12) Akiyama, Y.; Eda, S.; Kato, K. *Agric. Biol. Chem.* **1984**, *48*, 235–237.
(13) Randall, R. C.; Phillips, G. O.; Williams, P. A. *Food Hydrocolloids* **1988**, *2*, 131–140.
(14) Randall, R. C.; Phillips, G. O.; Williams, P. A. *Food hydrocolloids* **1989**, *3*, 65–75.
(15) Osman, M. E.; Menzies, A. R.; Williams, P. A.; Phillips, G. O.; Baldwin, T. C. *Carbohydr. Res.* **1993**, *246*, 303–318.
(16) Idris, O. H. M.; Williams, P. A.; Phillips, G. O. *Food Hydrocolloids* **1998**, *12*, 379.
(17) Siddig, N. E.; Osman, M. E.; Assaf, S. A.; Phillips, G. O.; Williams, P. A. *Food Hydrocolloids* **2005**, *19*, 687–702.
(18) Churms, S. C.; Merrifield, E. H.; Stephen, A. M. *Carbohydr. Res.* **1977**, *55*, 3.

(19) Churms, S. C.; Merrifield, E. H.; Stephen, A. M. *Carbohydr. Res.* **1978**, *64*, 61.
(20) Churms, S. C.; Stephen, A. M. *Carbohydr. Res.* **1975**, *45*, 291.
(21) Anderson, D. M. W.; Stoddart, J. F. *Carbohydr. Res.* **1966**, *2*, 104.
(22) Yariv, J.; Lis, H.; Katchalski, E. *Biochem. J.* **1967**, *105*, 1c–2c.
(23) Yariv, J.; Rapport, M. M.; Graf, L. *Biochem. J.* **1962**, *85*, 383–388.
(24) Schultz, C. J.; Johnson, K. L.; Currie, G.; Bacic, A. *Plant Cell* **2000**, *12*, 1751–1767.
(25) Schultz, C. J.; Rumsewicz, M. P.; Johnson, K. L.; Jones, B. J.; Gaspar, Y. M.; Bacic, A. *Plant Physiol.* **2002**, *129*, 1448–1463.
(26) Seifert, G. J.; Roberts, K. *Annu. Rev. Plant Biol.* **2007**, *58*, 137–161.
(27) Nothnagel, E. A. *Int. Rev. Cytol.* **1997**, *174*, 195–291.
(28) Showalter, A. M. *Cell. Mol. Life Sci.* **2001**, *58*, 1399–1417.
(29) Williams, P. A.; Langdon, M. J. Characterisation of gum arabic by gel permeation chromatography. *Chromatogr. Anal.* **1995**, *5* (Oct/Nov), 5–7.
(30) Connolly, S.; Fenyo, J. C.; Vandevelde, M. C. *Carbohydr. Res.* **1988**, *8*, 23.
(31) Williams, P. A.; Phillips, G. O.; Stephen, A. M. *Food Hydrocolloids* **1990**, *4*, 305.
(32) Qi, W.; Fong, C.; Lamport, D. T. A. *Plant Physiol.* **1991**, *96*, 848–855.
(33) Goodrum, L. J.; Patel, A.; Leykam, J. F.; Kieliszeswki, M. J. *Phytochemistry* **2000**, *54*, 99.
(34) Baldwin, T. C.; McCann, M. C.; Roberts, K. *Plant Physiol.* **1993**, *103*, 1953–1956.
(35) Cheung, A. Y.; Wang, H.; Wu, H. M. *Cell. Mol. Life Sci.* **1995**, *82*, 383–393.
(36) Yadav, M. P.; Johnston, D. B.; Hicks, K. B.; Nothnagel, E. A. *Foods Food Ingredients J. Jpn.* **2006**, *211* (3), 243–254.
(37) Lewis, B. A.; Smith, F. *J. Am. Chem. Soc.* **1957**, *79*, 3929.
(38) Carlson, D. M. *Biol. Chem.* **1968**, *243* (3), 616–620.
(39) Lloyd, K. O.; Burchell, J.; Kudryashov, V.; Yin, B. W. T.; Papadimitriou, J. T. Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines — Demonstration of simpler and fewer glycan chains. *J. Biol. Chem.* **1996**, *271*, 33325–33334.
(40) Morelle, W.; Guyetant, R.; Strecker, G. Structural analysis of oligosaccharide alditols released by reductive b elimination from oviducal mucins of *Rana dalmatina*. *Carbohydr. Res.* **1998**, *306*, 435–443.
(41) Mort, A. J. *Carbohydr. Res.* **1983**, *122*, 315–321.
(42) Ansorge, W. *J. Biochem. Biophys. Methods* **1986**, *11*, 21–30.
(43) Morelle, W.; Guyetant, R.; Strecker, G. *Carbohydr. Res.* **1998**, *306*, 435–443.
(44) Goodrum, L. J.; Patel, A.; Leykam, J. F.; Kieliszeswki, M. J. *Phytochemistry* **2000**, *54* (1), 99–106.
(45) Shultz, C. J.; Rumsewicz, M. P.; Johnson, K. L.; Jones, B. J.; Gaspar, Y. M.; Bacic, A. *Plant Physiol.* **2002**, *129*, 1448–1463.

Received for review March 18, 2008. Revised manuscript received June 11, 2008. Accepted July 28, 2008.

JF800849A