Carbohydrate Polymers 89 (2012) 749–758

Contents lists available at SciVerse ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Structural investigation of a glycoprotein from gum ghatti

Ji Kang[a], Steve W. Cui[a,b,*], Qingbin Guo[b], Jie Chen[a], Qi Wang[b], Glyn O. Phillips[c], John Nikiforuk[d]

[a] State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu 214100, China
[b] Guelph Food Research Centre, Agriculture and Agri-Food Canada, Guelph, Ontario N1G 5C9, Canada
[c] Phillips Hydrocolloids Research Centre, Glyndwr University, Wrexham, Wales LL11 2AW, UK
[d] East Cereal and Oil Research Centre, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 12 December 2011
Received in revised form 17 March 2012
Accepted 2 April 2012
Available online 9 April 2012

Keywords:
Gum ghatti
Glycoprotein
NMR
Structure characterization

ABSTRACT

The glycoprotein structure of gum ghatti was investigated. The covalent bonding between polysaccharides and proteins was firstly confirmed by high performance size exclusion chromatography (HPSEC) using refractive index (RI) and UV detectors. Partial acid hydrolysis and enzymatic degradation were also utilized. Several structural fragments such as $AraHex_2$-HexA-HexNAc, $Hex_4HexNAc_2$, $AraHex_4HexNAc$, $Hex_{10}HexNAc$ and $Hex_4HexNAc$-Asn were identified from MALDI-TOF spectrum; using 1D and 2D NMR spectra, the linkage site of amino acids and polysaccharides was determined as N-linked (Hex)n-GlcNAc-Asn. Combined with the polysaccharide structure obtained before, a glycoprotein structure model was proposed. It was composed of 1,6-linked galactose backbone, which were attached by numerous of sugar side chains and peptide chains.

1. Introduction

The major application of gum ghatti in foods is mainly due to its emulsification ability, which was reported to be superior to that of gum Arabic when used at lower concentrations (Al-Assaf, Amar, & Phillips, 2008). Recently, some work has been done to explain the mechanism of superior emulsification property of gum ghatti. Firstly, more gum component was found to be adsorbed onto the oil droplets in the emulsions of gum ghatti (30%, w/w) than that of gum Arabic (7–10%) (Katayama et al., 2008). Secondly, the adsorbed components of gum ghatti in the emulsion distributed evenly in the whole molecular weight range: from high molecular to low molecular weight; on the contrary, only the high Mw fraction of gum Arabic was adsorbed on to the oil surface (Katayama et al., 2008). This could explain why gum ghatti is a better stabilizer even at relatively lower gum content. Thirdly, the protein content of gum ghatti is higher than that of gum Arabic (Al-Assaf et al., 2008). However, to our best knowledge, little work was done to explain the emulsification properties of gum ghatti from the structural perspective. We recently fractionated gum ghatti into four homogenous fractions (Kang, Cui, et al., 2011a) and characterized the two major structures (Kang, Cui, et al., 2011b; Kang, Cui, Chen, et al., 2011). The current study is a continuation of previous studies in order to understand why gum ghatti is a better emulsifier than gum Arabic.

For comparison purposes, the glycoprotein structure of gum Arabic is summarized here. It was a highly heterogeneous polysaccharide and consisted of three main fractions. These fractions were arabinogalactan protein complex (AGP), arabinogalactan (AG) and glycoprotein (GP). One of the proposed models for AGP was that of hydrophilic carbohydrate blocks linked to a protein chain with a wattle-blossom-type structure which was readily degradable by proteolytic enzyme (Osman, Menzies, Williams, Phillips, & Baldwin, 1993).

In order to clarify whether the protein and polysaccharide parts in gum ghatti were covalently linked or not, the composition and physicochemical properties before and after treatment (with mild acid, protease, α-L-arabinofuranosidase, and β-D-galactosidase) were determined. Molecular weight profiles for both polysaccharide and protein were monitored by high performance size exclusion chromatography (HPSEC) coupled with refractive index, UV absorbance, right and low angle light scattering, and viscometer detectors. The weight average molecular weight (Mw) values were calculated using a conventional calibration for relatively small molecular weight (<10,000 Da), whereas, a triple detection method is used for large molecular weight peaks (≥10,000 Da). The detailed protein and polysaccharide linkage information in the protein-rich fractions were obtained using 1D and 2D NMR (COSY, TOCSY, HMQC and HMBC) and MALDI-TOF-MS analysis. Understanding the linkage pattern between protein and polysaccharide of gum ghatti could lead to preparation of natural high-efficiency emulsifier and stabilizer through chemical modification.

* Corresponding author at: Guelph Food Research Centre, Agriculture and Agri-Food Canada, Guelph, Ontario N1G 5C9, Canada. Tel.: +1 519 780 8028; fax: +1 519 829 2600.
E-mail addresses: cuis@agr.gc.ca, Steve.Cui@AGR.GC.CA (S.W. Cui).

0144-8617/$ – see front matter.
http://dx.doi.org/10.1016/j.carbpol.2012.04.004

2. Materials and methods

2.1. Materials

2.1.1. Samples

Gatifolia SD, a specially selected, cleaned and spray dried product of gum ghatti, was from San-Ei Gen F.F.I., Inc., Osaka, Japan.

2.1.2. Enzyme and chemicals

The α-L-arabinofuranosidase from *Aspergillus niger* (Lot. 10401), β-D-galactosidase from *A. niger* (80601) and protease from *Bacillus licheniformis* (Lot 50901) were purchased from Megazyme.

2.2. Methods

2.2.1. Amino acid composition

The Gatifolia SD was hydrolyzed by 6 M HCl in screw-capped tubes. Tubes were flushed with N_2 and then heated at 110 °C for 24 h. The hydrolysates were rotary-evaporated to dryness under vacuum at 40 °C, then analyzed on an Agilent 1100 high performance liquid chromatography for the amino acid composition (Agilent Corporation, America) (4.0 mm × 125 mm, Amersham Parmacia Biotech, Switzerland). The elution programme consisted of a gradient system with buffer A (40 mM Na_2HPO_4, pH 7.8) and buffer B (ACN:MeOH:water, 45:45:10, v/v/v), respectively, at a flow-rate of 1 mL/min. The elution pattern was monitored by measuring the absorbance at 450 nm when excited at 340 nm.

2.2.2. Dearabinosylation of Gatifolia SD

Gatifolia SD (800 mg) was dissolved in 40 mM sodium acetate buffer (400 mL, pH 4.0) with 0.04% NaN_3, and incubated with 100 μl *A. niger* α-L-arabinofuranosidase at 40 °C for 3 days. The reaction was terminated by heating at 100 °C for 5 min. After dialysis, ethanol precipitation, centrifugation and lyophilization, the precipitate and supernatant could be obtained and designated AP and AS, respectively. Then the samples were dissolved in deionized water, filtered through a 0.45 μm filter membrane, and analyzed by high performance size exclusion chromatography (HPSEC) coupled with refractive index, UV absorbance, right and low angle light scattering, and online-viscometer detectors.

2.2.3. Degalactosylation of Gatifolia SD

Gatifolia SD (10 mg) was dissolved in 20 mM sodium acetate buffer (5 mL, pH 4.5) with 0.04% NaN_3 and incubated with 25 μl *A. niger* β-D-galactosidase at 60 °C for 3 days. The procedure was repeated and the solution dialyzed against deionized water. After enzyme inactivation (15 min, 100 °C), 3 volumes of ethanol was added to the solution, then centrifuged and lyophilized, the supernatant was designated GS, and the precipitate was named as GP. Both samples were analyzed for protein content and molecular weight distribution.

2.2.4. Partial acid hydrolysis of Gatifolia SD

Gatifolia SD (2 g) was hydrolyzed with 0.1 M trifluoroacetic acid (10 mL) at 100 °C for 20 min, 60 min, 120 min, and 210 min. After hydrolysis samples were cooled down to room temperature and precipitated with three volumes of ethanol (v/v) followed by freeze drying. The precipitate and supernatant for each sample were designated 20P/20S, 60P/60S, 120P/120S and 210P/210S correspondingly (Fig. 1).

2.2.5. Fractionation of 210P by Bio-gel

The 210P, the precipitate fraction after 210 min acid hydrolysis, was applied on a Bio-Gel P-4 (Hercules, CA) and eluted with distilled water at a flow rate of 0.5 mL/min at 25 °C. Fractions of every 2.6 mL were collected and analyzed by HPSEC.

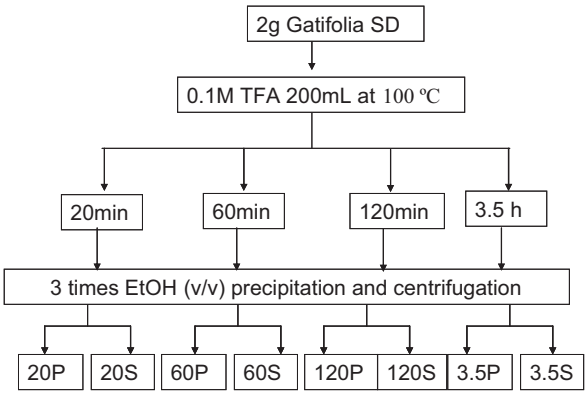

Fig. 1. Procedure for partial acid hydrolysis of Gatifolia SD.

2.2.6. Protease hydrolysis of Gatifolia SD

The Gatifolia SD, 210P and *Artemisia sphaerocephala* Krasch seed gum (used as control for verifying the enzyme activity), were dissolved seperatedly in 80 mM phosphate buffer (pH 7.5), then the protease was added and incubated at 60 °C for 3 days with constant stirring, followed by dialysis (1000 Da Mw cut-off) and then freeze drying.

2.2.7. Chemical and monosaccharide analyses

Protein content was analyzed by a NA2100 Nitrogen and Protein Analyzer (Strada Rivoltana, Milan, Italy) using the factor of 6.25 to convert measured nitrogen into protein content. Ash and moisture contents were examined according to the AOAC methods (Anonymous, 2000).

Monosaccharide composition analysis was conducted using a modified method of Mopper et al. (1992). Detailed procedure was described earlier (Kang, Cui, et al., 2011a).

Uronic acid content was carried out using a high performance anion exchange chromatographic method described earlier (Kang, Cui, et al., 2011a). The sample was prepared by hydrolysis in 1 M H_2SO_4 at 100 °C for 2 h with galacturonic acid and glucuronic acid as standards.

2.2.8. Molecular weight characterization

The molecular weight distribution profile was determined using a high performance size-exclusion chromatograph (HPSEC) equipped with multiple detectors: a differential pressure viscometer (DP) for viscosity determination; a refractive index detector (RI) and a UV detector for concentration determination; a right angle laser light scattering detector (RALLS) and a low angle laser light scattering detector (LALLS) (Viscotek, triple detector array from Malvern company). Two columns in series: a Shodex Ohpak KB-806M (Showa Denko K.K., Tokyo, Japan), and an Ultrahydrogel linear (Waters, Milford, CT, USA) were equipped. The columns, viscometer and RI detector were maintained at 40 °C. The eluent was 0.1 M $NaNO_3$ containing 0.03% (w/w) NaN_3 at a flow rate of 0.6 mL/min. Data was obtained and analyzed using the OmniSEC 4.6.1 software.

2.2.9. Matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry

The matrix chosen for MALDI-TOF analysis was 2,5-dihydroxybenzoic (DHB) acid, 10 mg of recrystallized DHB was dissolved in 1 mL of 20% ethanol solution in HPLC grade water. Samples were re-suspended in an equal volume of matrix solution

Table 1
Amino acid composition of Gatifolia SD.

Amino acids	Content (%)	Amino acids	Content (%)	Amino acids	Content (%)
Asx	18.62	Ala	5.64	Arg	3.52
Lys	9.59	Leu	5.27	Cys-s	3.49
Glu	9.41	Thr	4.55	His	3.48
Gly	9.14	Ser	4.38	Phe	2.81
Val	6.73	Ile	3.81	Met	0.11
Pro	5.81	Tyr	3.6		

Table 2
Relative monosaccharide composition of original Gatifolia SD, AP and AS.

	Yield (%)[a]	Rha	Ara	Gal	Glc	Xyl	Man
Gatifolia SD		2.56	57.27	33.83	0.98	2.62	2.75
AP	73.72	1.28	46.70	44.35	0.00	4.49	3.18
AS	26.18	5.25	87.08	6.67	1.00	0.00	0.00

[a] Calculated from the monosaccharide content.

and dried on a stainless steel plate prior to MALDI analysis. The sample/matrix mixtures were allowed to air-dry, and were analyzed in a Bruker Reflex III (Bruker Daltonik, Germany) equipped with a 337 nm nitrogen laser (Mass Spectrometry Facility, University of Guelph), operating in reflectron and linear detection mode with an external calibration. Samples were analyzed in positive ion modes scanning from 0 to 10,000 *m*/*z* using ion suppression up to 500 *m*/*z*. For all experiments, the ion sources 1 and 2 were held at 20 kV and 16.35 kV, respectively, and the guiding lens voltage at 9.75 kV. The reflector detection gain was set up at 5.3 with pulsed ion extraction at 200 ns. 15% of laser energy was used, which is equal to ~7 mJ per laser shot. Spectra were acquired from an average of at least 100 shots. The spectra were processed using the mMass software (Strohalm, Kavan, Novak, Volny, & Havlicek, 2010).

2.2.10. NMR methods

Samples were dried in a vacuum oven (~80 °C) for 6 h, then dissolved in 5 mL deuterium oxide (D_2O) (~4%, w/v) and freeze dried. This procedure was repeated three times to completely replace H with D, and samples were finally dissolved in D_2O at room temperature for 3 h before NMR analysis.

Both ^1H and ^{13}C NMR spectra were recorded on a Bruker AMX 500 FT spectrometer at 25 °C, the detailed procedure was described earlier (Cui, Wood, Blackwell, & Nikiforuk, 2000).

3. Results and discussion

3.1. Amino acid composition of Gatifolia SD

Amino acid analysis of the original gum ghatti (Gatifolia SD) (Table 1) showed that Asx, Lys, Glx and Gly were the major amino acids and accounted for 18.62%, 9.59%, 9.41% and 9.14% of the total amino acids. In the hydrolysis process, amino acids Asn and Gln were converted to Asp and Glu, respectively. Therefore, the Asn or Asp content was shown as Asx, and Glx represents Glu or Gln.

This result was significantly different from the amino acids composition of gum Arabic, which is mainly composed of Hyp (32.7%), Ser (16.3%), Leu (6.8%), Pro (7.6%) and Thr (7.0%) with only 3.9% of Asp (Qi, Fong, & Lamport, 1991). This difference in amino acid composition indicated a different protein structure of gum ghatti compared with gum Arabic.

3.2. Degradation of Gatifolia SD by enzyme and chemical method

3.2.1. Degradation of Gatifolia SD by α-L-arabinofuranosidase

The total arabinose content was 57.27% (w/w) based on total sugar of Gatifolia SD. Degradation by α-L-arabinofuranosidase removed 22.80% of arabinose relative to the total sugar of Gatifolia SD. After ethanol precipitation, the arabinose content was determined as 46.70% in the AP fraction (Table 2). Meanwhile, the methylation results of gum ghatti after uronic acid reduction showed that the amount of t-ara*f* in gum ghatti was 27.32% (Kang, Cui, Chen, et al., 2011) of the total sugar. Not all of the terminal arabinose was removed, which might be caused by steric hindrance of the molecule due to the large molecular weight and highly branched structure; namely, the enzyme could not approach the site to hydrolyze the glycosidic bond.

The HPSEC coupled with multiple detectors was used to monitor the polysaccharide and protein component distribution in the original gum ghatti (Gatifolia SD) and its hydrolysates. RI response was used to monitor the concentration (both polysaccharide and protein), while the UV detector was used to trace the proteinaceous component. The RI and UV elution profiles of original gum and its hydrolysates (before ethanol precipitation) are shown in Fig. 2.

Three peaks were observed from the HPSEC profile in the oligosaccharide region for the hydrolyzed sample: the peak molecular weights were 1027, 247 and 145 Da, respectively, based on dextran standards. These results indicated that mono-, disaccharide and 4–8 units of oligosaccharide were removed from the Gatifolia SD. It is worth noting that the UV profile of the Gatifolia SD and α-L-arabinofuranosidase degraded sample were very similar, as shown in Fig. 2. This result suggested that the protein groups were not linked to the arabinose located at the outer layer (branches) of the polysaccharide molecule.

3.2.2. Degradation of Gatifolia SD by β-D-galactosidase

Three small fractions were generated by the treatment of the Gatifolia SD using β-D-galactosidase, as shown in Fig. 3. The peak molecular weight of these three fractions was 143, 617 and 1578 Da, respectively. These small fractions were composed of monosaccharide, 2–4 and 6–10 units of oligosaccharides. By comparing the RI profile of Gatifolia SD with degraded product by β-D-galactosidase (Fig. 3), the Mw (calculated by triple detection method using RI, LS and Viscometer detector) after degradation was slightly reduced (from 958 to 869 kDa), suggesting that large amount of oligosaccharide or monosaccharide were removed; the backbone of the polysaccharides remained intact. In addition, the UV profile of the hydrolyzed polysaccharide did not change (Fig. 3). Further suggesting that the protein moiety was not covalently linked to the

Table 3
Protein content of partial acid hydrolysates.

Hydrolysis time	Precipitate	Supernatant
20 min	13.70 ± 1.36	3.26 ± 2.13
60 min	10.73 ± 2.67	1.40 ± 0.26
120 min	16.23 ± 1.12	1.76 ± 0.11
3.5 h	13.79 ± 0.02	1.16 ± 0.07

galactose residues on the periphery of the molecules; and suggests that they are linked to the backbone of the polysaccharides.

3.2.3. Partial acid hydrolysis of Gatifolia SD

As described in Section 2.2.4, Gatifolia SD (2 g) was hydrolyzed with 0.1 M TFA (10 mL) at 100 °C for 20 min, 60 min, 120 min, and 210 min, respectively. The hydrolysates were neutralized and precipitated with three volumes of ethanol (v/v) to give precipitate and supernatant fractions which are designated as 20P/20S, 60P/60S, 120P/120S and 210P/210S, respectively.

As shown in Table 3, the protein contents in the precipitated fractions (20P, 60P, 120P and 210P) were all higher than 10%; on the contrary, the protein content in the soluble fractions were very low: 3.26%, 1.40%, 1.76%, and 1.16% (w/w) for the corresponding 20S, 60S, 120S and 210S fractions. These results showed that the protein moieties were mainly associated with the precipitate fractions. The results of monosaccharide composition analysis of the hydrolysates are shown in Table 4. After 20 min of acid hydrolysis, most of the arabinosyl units were removed; as a result, the level of galactosyl units in the precipitate fraction increased. With increasing hydrolysis time, rhamnose content was also decreased in the precipitate fraction; on the contrary, the relative amounts of xylose and mannose were increased. All these results indicated that galactose, mannose and xylose were likely located at the backbone or core part of the molecule, while the arabinosyl and rhamnosyl units were located in the periphery of the molecule.

The HPSEC coupled with multiple detectors was used to monitor the molecular weight distribution of the polysaccharide and proteinaceous components in the gum ghatti (Gatifolia SD) and the precipitate fractions after partial acid hydrolysis; the results are shown in Fig. 4.

The original gum ghatti gave a symmetrical peak at the retention volume around 15 mL in both RI and UV (280 nm) profiles. After hydrolysis (0.1 M TFA at 100 °C) for 20 min, the retention volume of the peak shifted to 17 mL in both RI and UV profiles,

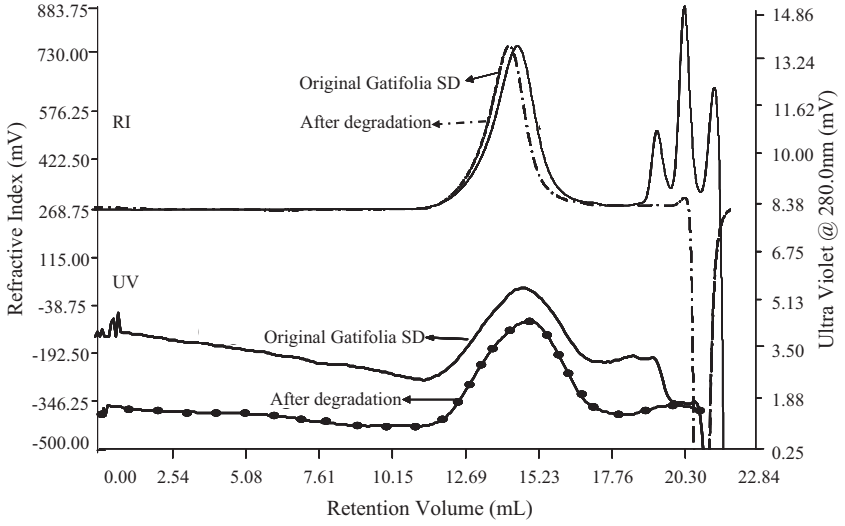

Fig. 2. HPSEC RI (a) and UV (b) profiles of Gatifolia SD before and after treatment 4 with α-L-arabinofuranosidase.

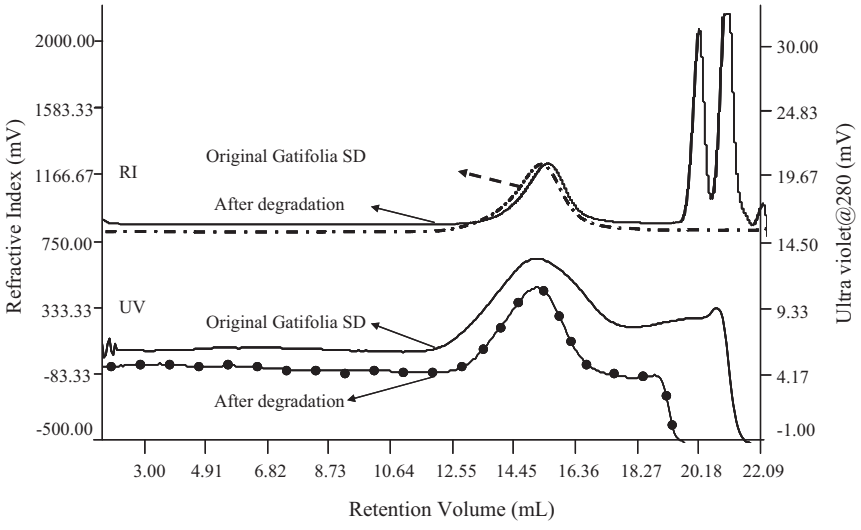

Fig. 3. HPSEC RI and UV profile of Gatifolia SD after treatment with β-D-6 galactosidase.

Table 4
Relative monosaccharide composition of partial acid hydrolysates.

	Precipitate				Supernatant			
	20P	60P	120P	210P	20S	60S	120S	210S
Rha	1.48	0.49	0.36	0.4	2.18	2.24	2.26	2.45
Ara	13.37	8.77	10.71	6.84	80.58	73.21	64.6	72.61
Gal	72.62	77.1	72.93	80.4	14.43	20.06	27.75	21.05
Glc	0	0.78	0.78	1.19	0	0.25	0.24	0.14
Xyl	6.96	5.24	4.45	6.06	2.78	4.05	4.73	3.51
Man	5.57	7.61	10.77	5.11	0.02	0.17	0.42	0.25

Table 5
Molecular weight and proportion of each fraction from acid hydrolysis.

	Higher Mw fraction		Lower Mw fraction	
	Mw (Da)	Weight proportion (w/w)	Mw (Da)	Weight proportion (w/w)
Original	958,357	100%	–	–
20P	77,748	78.10%	3167	21.90%
60P	27,349	70.00%	2217	30.00%
120P	10,791	71.60%	2393	28.40%
210P	8,959	76.90%	2153	23.10%

accompanied with a smaller peak at 19.7 mL, observed only in the RI profile. The Mw of the two peaks in the RI chromatograph of the 20P was 78 and 3.1 kDa, respectively. This observation indicated that after 20 min hydrolysis, some of the branches were removed from the gum ghatti molecule; however, the protein moieties were only linked with the high molecular weight fraction. In the same way, the 60P, 120P and 210P fractions also yielded two peaks. Their corresponding molecular weights were 27.3 and 2.2 kDa for 60P, 10.8 and 2.4 kDa for 120P, 9.0 and 2.2 kDa for 210P, respectively, which are summarized in Table 5. Further observations could be drawn from Fig. 4 and Table 5: with an increase of hydrolysis time, more branches were removed from the backbone or core part of the polysaccharide molecule; the Mw of the remaining portion decreased gradually. It is important to notice that the retention volumes of the UV peaks always coincided with that of the RI peaks, suggesting that protein moieties were still attached to the backbone or core part of the polysaccharide after extensive acid hydrolysis.

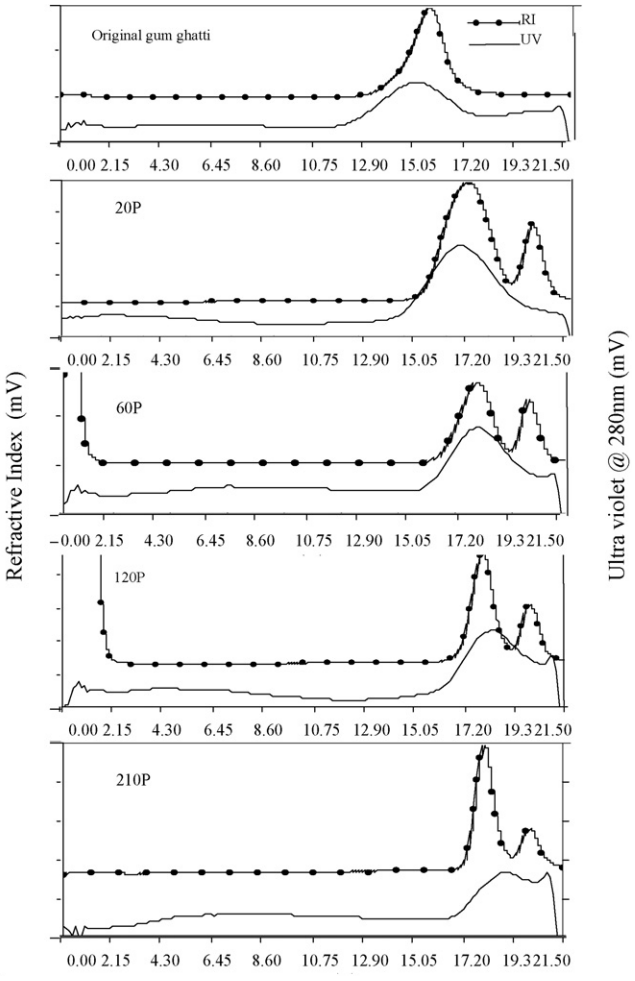

Fig. 4. HPSEC profiles from RI and UV signals of mild acid hydrolysates of Gatifolia 8 SD with different hydrolysis time (20 min, 60 min, 120 min and 210 min); RI: (-●-); UV (—).

3.3. Elucidation of protein–polysaccharide linkage of the high protein content fraction 210P

3.3.1. Fractionation of the 210P

210P was chosen as a representative of Gatifolia SD for the investigation of the linkage between protein and polysaccharide, because it has the lowest molecular weight among all the precipitates (Fig. 4) and relatively higher protein content (Table 3). The two molecular weight fractions of 210P were separated by eluting through a Bio-gel P400 column: the higher Mw fraction was designated 210PH, and the lower Mw was named as 210PL. The molecular weight distributions of 210P, 210PH and 210PL are

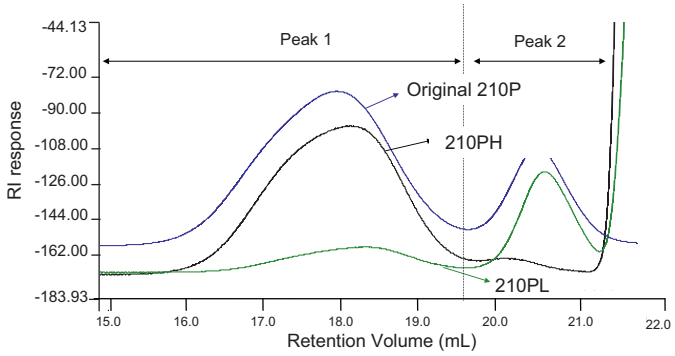

Fig. 5. SEC RI profile of original 210P and its two components (high Mw fraction and 11 low Mw fraction).

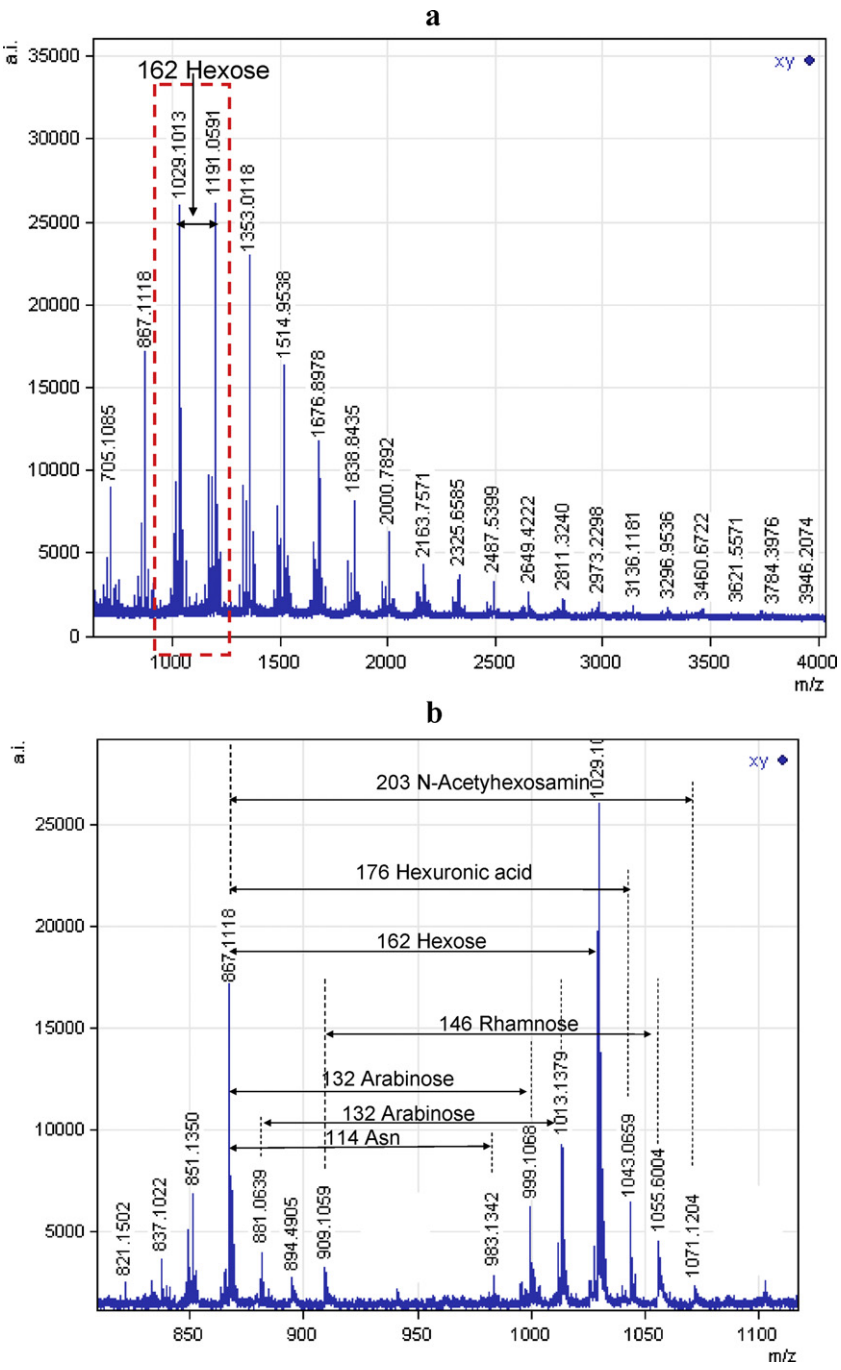

Fig. 6. MALDI-TOF mass spectra of 210P high Mw fraction (a) whole spectrum and (b) 13 expansion of the region indicated by the box in (a).

shown in Fig. 5, which indicated a good separation of the two fractions.

3.3.2. Protease hydrolysis of 210PH

Gatifolia SD and 210PH were subjected to protease hydrolysis. Protease did not decrease the Mw of the 210PH and Gatifolia SD significantly, and the protein content of the ethanol precipitate after protease hydrolysis was 12.28%, which was slightly lower than that of the original 210PH (13.79%). The effectiveness of the protease was verified in a previous study (Guo et al., 2011). Our results showed that the protease could only hydrolyze a small part of the protein in gum ghatti; in other words, the majority of the protein components in 210PH were not affected by protease probably because of the steric hindrance which prevented the protease from accessing the bonds.

3.3.3. MALDI-TOF-MS study of 210PH

The higher Mw fraction 210PH was subjected to MALDI-TOF MS. A series of ions were detected with relatively strong intensity in the mass spectrum (Fig. 6a). The *m*/*z* values were 705, 867, 1029, 1191, 1353, 1514, 1676, 1838, etc. as labeled in Fig. 6; the mass differences between the neighboring peaks were 162 Da, suggesting that they were due to the elimination of 1 mol hexose, which was mainly galactose according to the monosaccharide composition results of 210P (Table 4). In addition, the ions at *m*/*z* 867.1 corresponded to the theoretical molecular weight of four hexoses (specifically galactose or mannose in the current

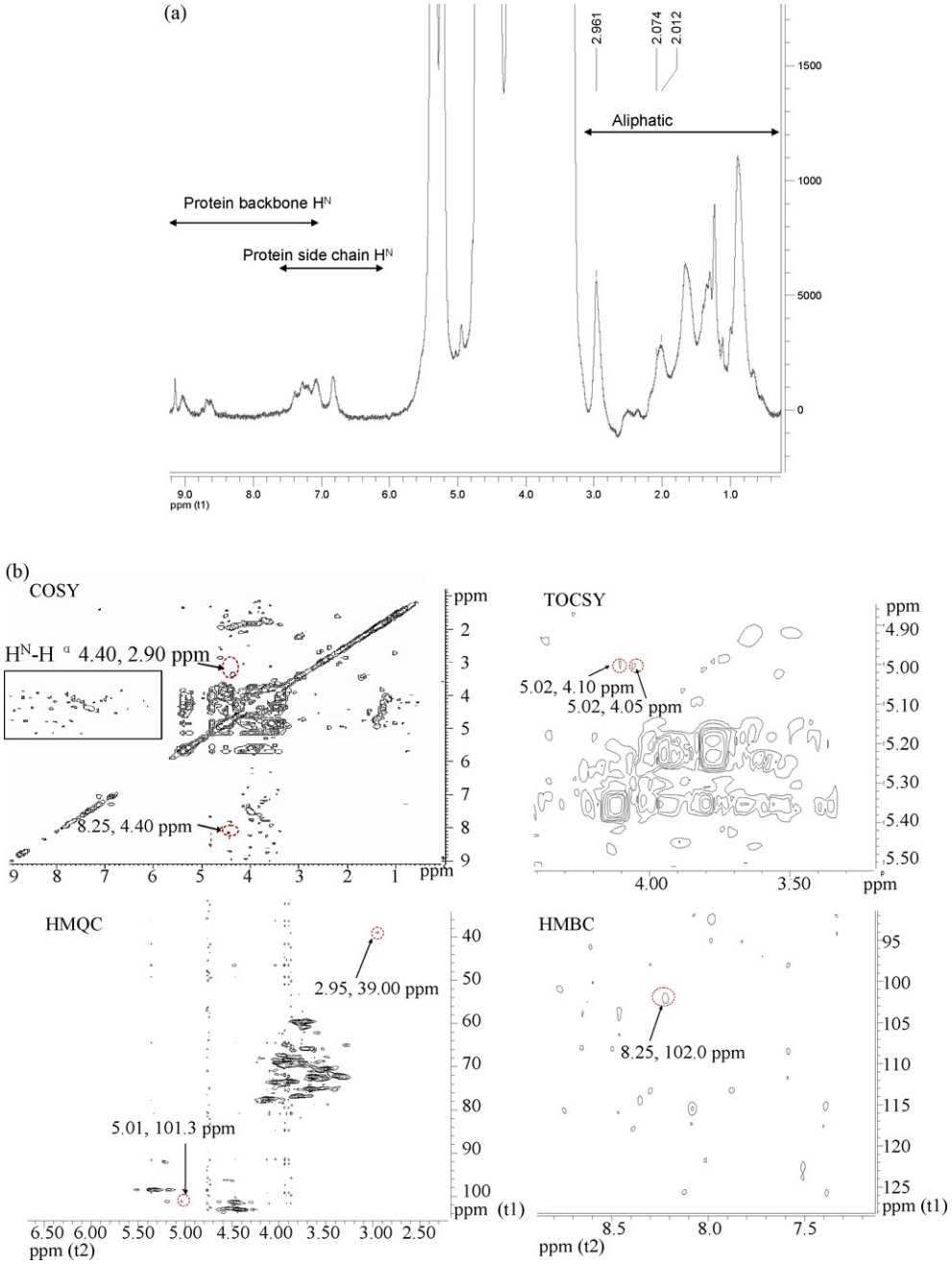

Fig. 7. 1D NMR (^{1}H) spectrum (a) and 2D NMR (COSY, TOCYS, HMQC and 15 HMBC) spectra (b) of 210PH.

study) and one N-acetylhexosamine ($Hex_4HexNAc$). Consequently, *m*/*z* 705, 1029, 1191, 1353, 1514, 1676 and 1838 were derived from $Hex_3HexNAc$, $Hex_5HexNAc$, $Hex_6HexNAc$, $Hex_7HexNAc$, $Hex_8HexNAc$, $Hex_9HexNAc$ and $Hex_{10}HexNAc$, respectively. In addition, several ion peaks were observed with less intensity, as shown in Fig. 6b, which was the expansion of the marked region in Fig. 6a. The mass difference between 867.1 and 999.1 was 132 Da, indicating that a pentose (specifically arabinose in the current study) was cleaved from the molecule. Since *m*/*z* 867.1 was assigned to $Hex_4HexNAc$, it can be concluded that 999.1 *m*/*z* was $AraHex_4HexNAc$. The ions at *m*/*z* 1071.1 can be recognized as 867.1 plus 203, indicating the presence of N-acetyhexosamin. Based on the above analysis, ions at *m*/*z* 1071.1 would be of $Hex_4HexNAc_2$. Accordingly, ions at *m*/*z* 909.1 corresponded to $Hex_3HexNAc_2$. The mass difference between *m*/*z* 1055.1 and 909.1 was 146, which was derived from a loss of deoxyhexose (specifically rhamnose in the current study). Combined with previous results, ions at *m*/*z* 1055.1 could be assigned to $RhaHex_3HexNAc_2$. Ions at *m*/*z* 1043.1 were attributed to Hex_3-HexA-HexNAc due to the mass difference of 176 relative to ions at *m*/*z* 867.1, as shown in Fig. 6a, because fragment 176 was a loss of hexuronic acid. A less intensive ion peak at *m*/*z* 881.1 was observed to have a difference of 162 with *m*/*z* 1043.1, indicating the ions at *m*/*z* 881.1 were Hex_2-HexA-HexNAc. While a 132 *m*/*z* peak difference was detected between *m*/*z* 881.1 and 1013.1, leading to a conclusion that the ions at *m*/*z* 1013.1 were $AraHex_2$-HexA-HexNAc. The mass difference between 983.1 and 867.1 was 114 Da, which was responsible for the elimination of asparagine, and the fragment of 983.1 could be attributed to $Hex_4HexNAc$-Asn. This result could also be verified by the highest content of Asx in 210P among all the amino acids (Table 1).

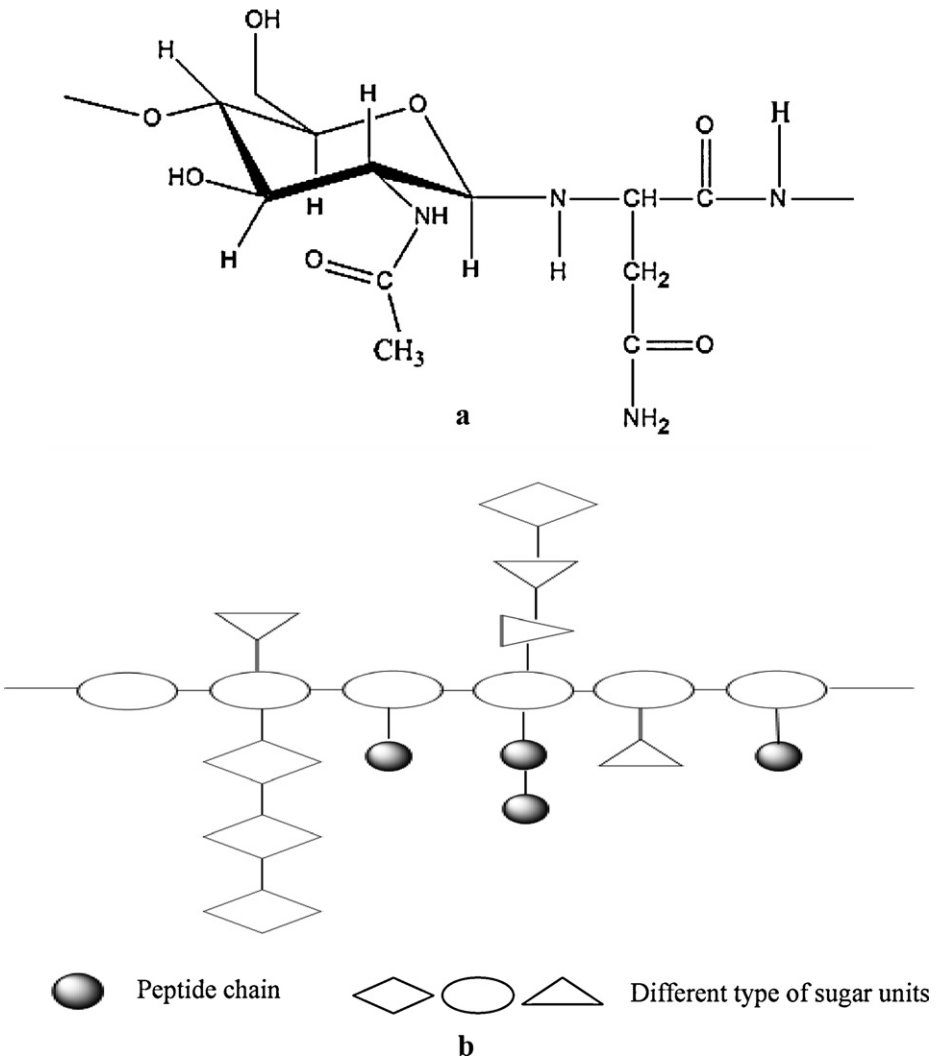

Fig. 8. The linkage pattern of amino acid and sugar unit (a). The schematic model of 17 gum ghatti glycoprotein (b).

3.3.4. 1D and 2D NMR analysis 210PH

Most of the NMR signals of protein and polysaccharide were located in different regions of 1D spectra, except the H^{α} signals of amino acids overlapped with the signals from polysaccharides. However, these overlapped peaks can be resolved in 2D NMR spectra. Therefore, NMR is still a powerful tool for analyzing the structure of glycoproteins or glycopeptides with relatively low protein or peptides content. In this study, 210PH was subjected to 1D and 2D NMR investigation including homonuclear $^{1}H/^{1}H$ correlations spectroscopy (COSY, TOCSY), heteronuclear $^{13}C/^{1}H$ multiple-quantum coherence spectroscopy (HMQC) and heteronuclear multiple bond correlation (HMBC) and the spectra are shown in Fig. 7b.

The fragment Hex_4HexNAc-Asn obtained from the MALDI-TOF results led to a hypothesis that Asn might link to the HexNAc. In order to verify this hypothesis, the linkage patterns and chemical shifts of Asn and GlcNAc and GalNAc were investigated in 1D and 2D NMR spectra.

The ^{1}H spectrum of 210PH is shown in Fig. 7a, the signals from 7.1 to 9.0 ppm were derived from the backbone signals of protein, while 7.5–6.0 ppm originated from the side chains of protein; the resonances from 0 to 3 ppm belonged to the aliphatic group of protein and polysaccharide. The broad peak around δ 2.96 (as marked in the ^{1}H spectrum) belonged to the signals H^{β} of Asn, which were the overlap of two resonances. The COSY spectrum gave the correlation between neighboring protons within one amino acid or sugar ring, in the region of 3–5 ppm and 7–9 ppm, several cross peaks were observed, most of them belonged to the correlation between H^{N} and H^{α} of amino acids. Most of the cross peaks were between 3–5 ppm and 3–5 ppm corresponded to the signals of polysaccharides as marked in Fig. 7b. The correlation between H^{N} and H^{α}, H^{α} and H^{β} of Asn was detected; the corresponding cross peaks were located at δ 8.25 and 4.4 ppm, 4.4 and 2.9 ppm indicating the chemical shifts of H^{N} and H^{α} were 8.25 and 4.40 ppm, the results also confirmed the above assignment of H^{β}. To obtain the ^{13}C chemical shifts of Asn, HMQC spectrum was analyzed and the cross peaks at δ 2.947 and 39.002 ppm were observed, as shown in Fig. 7b, indicating that peak at 39.002 ppm was the C^{β} chemical shift of Asn. However, the correlation between H^{α} and C^{α} could not be found due to the low intensity. The assignment of C^{β} is consistent with the results of a previous study (38.9 ppm) (Wishart, Bigam, Holm, Hodges, & Sykes, 1995).

The peaks at the chemical shift of 2.012 and 2.073 ppm in ^{1}H spectrum were assigned to the CH_3 of NAc group, which is in agreement with literature values (Halbeek et al., 1980). The peak at δ 5.024 ppm was attributed to the H-1 of GlcNAc by comparing with the literature value of δ 5.03 ppm (Davis, Hirani, Bartlett, & Reid,

1994) and 5.075 ppm (Sturm, Bergwerff, & Vliegenthart, 1992), and this assignment was verified by the J1,2 value of 7.698 Hz, which was in agreement with Halbeek's results (Halbeek et al., 1980). In the TOCSY spectrum, two peaks were observed to have the correlation with the anomeric proton of GlcNAc, which were 5.02 and 4.05, 5.02 and 4.10, respectively. This result indicated that the other non-anomeric proton chemical shifts were 4.05, and 4.10 ppm (Fig. 7b). However, three other proton chemical shifts could not be assigned due to weak signals. The corresponding cross peak of the anomeric proton and carbon was found in the HMQC spectrum at δ 5.01 and 101.3 ppm, indicating the chemical shift of C-1 of GlcNAc was 101.3 ppm.

The sequence of glycosyl residues and amino acids of 210PH was determined by a long-range HMBC experiment. A cross peak between δ 8.25 and 102.0 ppm in the HMBC was observed, where the signal at 8.25 ppm belonged to the resonance of H^N of Asn, while the peak at 102 ppm was attributed to the ^{13}C signal for the anomeric carbon of GlcNAc. This result confirmed that Asn was connected with GlcNAc by N-link, as shown in Fig. 8a.

In summary, the results of partial acid hydrolysis indicated that most of the protein was covalently linked to the backbone of the polysaccharide instead of the side chains. Both of the α-L-arabinofuranosidase and β-D-galactosidase could release small molecular weight fragments from the molecules of gum ghatti. No protein was detected in those fractions, which supported the conclusion that protein was linked to the main chain of the polysaccharide. Combination of MALDI-TOF analysis and HMBC NMR spectroscopy further confirmed that the link of the protein moiety with the polysaccharides mainchain was through GlcNAc N-linkage. Based on the structure work of gum ghatti elucidated in our previous reports (Kang, Cui, et al., 2011a, 2011b), the glycoprotein structure of gum ghatti was proposed as shown in Fig. 8b. It was composed of a galactose backbone with numerous side chains, and proteins are attached directly to the core part of the polysaccharide. This model could explain why gum ghatti exhibited a stronger emulsification capacity than gum Arabic at the same concentration; it was probably caused by more hydrophobic group present in the protein parts.

4. Discussion and conclusions

Using partial acid hydrolysis, enzymatic degradation, HPSEC, MALDI-TOF-MS and 1D and 2D NMR, the glycoprotein structure of Gatifolia SD was investigated in the present study. Results indicated that the protein was covalently linked to the polysaccharide at the core part of the molecule. The linkage pattern between protein and polysaccharide was also studied. Summarized all the results, three possible explanations for superior emulsification properties of gum ghatti are listed as follows: (1) the highly branched structure of the molecules could cause the steric effects between different oil droplets by which the molecules were attached. This steric effect could decrease the rate of coalescence, hence maintaining the stability of emulsion systems. (2) The amphiphilic character of the glycoprotein molecules could also enhance the stability of the emulsion system, as the proteinaceous moieties could be functioning as the anchoring point to adsorb to oil droplets; whereas the hydrophilic polysaccharide moieties could be surrounding the outer layer of the oil droplets and also create the steric effects to prevent coalescence. (3) The relatively low hydrodynamic radius (23 nm) of the molecules, compared to other nature hydrocolloids, make it easy to be attached to oil droplets before flocculation and coalescence phenomena happens in the system. Based on the proposed model, the peptide chains were attached to the backbone of the gum ghatti molecules, which could be used to explain why the whole molecular weight range of gum ghatti molecules adsorbed onto the oil droplets in the emulsions (Katayama et al., 2008). The small Mw molecules of gum ghatti might be the degraded product of large molecules. Since the proteins were linked to the backbone of the gum ghatti molecules, and difficult to be removed during the degradation. Therefore, most of the gum ghatti molecules have protein and can be absorbed to the oil droplets in the emulsion system. This is different from the wattle-blossom model of gum Arabic, in which the small molecular weight fraction only contains polysaccharide, hence only large molecular weight fractions could be detected on the surface of oil droplets in emulsion system (Katayama et al., 2008). The results from the current study could be used to explain why the adsorbed components of gum ghatti emulsion distributed evenly across the whole Mw range, from high molecular to low molecular weight while only the high Mw fraction of gum Arabic was adsorbed (Katayama et al., 2008).

Acknowledgements

The authors wish to thank Ms. Cathy Wang and Dr. Ying Wu from Guelph Food Research Centre, Agriculture and Agri-Food Canada for technical assistance and helpful discussion. Thanks also go to Yin Hing Ma from Guleph Food Research Centre for help on proof reading of the manuscript. Also we thank Phillips Hydrocolloids Research Ltd for their support.

References

Al-Assaf, S., Amar, V., & Phillips, G. O. (2008). Characterization of gum ghatti and comparison with gum Arabic. In P. A. Williams, & G. O. Phillips (Eds.), *Gums and stabilisers for the food industry 14* (pp. 281–290). Cambridge: RSC. Special publication.

Anonymous. (2000). *Official methods of analysis*. Washington, DC: Association of Official Analytical Chemists.

Cui, W., Wood, P. J., Blackwell, B., & Nikiforuk, J. (2000). Physicochemical properties and structural characterization by two-dimensional NMR spectroscopy of wheat [beta]-D-glucan—Comparison with other cereal [beta]-D-glucans. *Carbohydrate Polymers*, *41*(3), 249–258. http://dx.doi.org/10.1016/s0144-8617(99)00143-5

Davis, J. T., Hirani, S., Bartlett, C., & Reid, B. R. (1994). ^{1}H NMR studies of an Asn-linked glycopeptide. *Journal of Biological Chemistry*, *269*(5), 3331–3338.

Guo, Q., Cui, S. W., Wang, Q., Hu, X., Guo, Q., Kang, J., et al. (2011). Extraction, fractionation and physicochemical characterization of water-soluble polysaccharides from *Artemisia sphaerocephala* Krasch seed. *Carbohydrate Polymers*, *86*(2), 831–836. http://dx.doi.org/10.1016/j.carbpol.2011.05.034

Halbeek, H. V., Dorland, L., Veldink, G. A., Vliegenthart, J. F. G., Michalski, J. C., Montreuil, J., et al. (1980). Structure elucidation of oligomannoside-type asparagine-bound carbohydrate chains of glycoproteins by 500 MHz 1H NMR spectroscopy. *FEBS Letters*, *121*(1), 65–70. http://dx.doi.org/10.1016/0014-5793(80)81268-3

Kang, J., Cui, S. W., Chen, J., Phillips, G. O., Wu, Y., & Wang, Q. (2011). New studies on gum ghatti (*Anogeissus latifolia*). Part I. Fractionation, chemical and physical characterization of the gum. *Food Hydrocolloids*, *25*(8), 1984–1990. http://dx.doi.org/10.1016/j.foodhyd.2010.12.011

Kang, J., Cui, S. W., Phillips, G. O., Chen, J., Guo, Q., & Wang, Q. (2011a). New studies on gum ghatti (*Anogeissus latifolia*). Part II. Structure characterization of an arabinogalactan from the gum by 1D, 2D NMR spectroscopy and methylation analysis. *Food Hydrocolloids*, *25*(8), 1991–1998. http://dx.doi.org/10.1016/j.foodhyd.2010.11.021

Kang, J., Cui, S. W., Phillips, G. O., Chen, J., Guo, Q., & Wang, Q. (2011b). New studies on gum ghatti (*Anogeissus latifolia*). Part III. Structure characterization of a globular polysaccharide fraction by 1D, 2D NMR spectroscopy and methylation analysis. *Food Hydrocolloids*, *25*(8), 1999–2007. http://dx.doi.org/10.1016/j.foodhyd.2010.11.020

Katayama, T., Ido, T., Sasaki, Y., Ogasawara, T., Al-Assaf, S., & Phillips, G. O. (2008). Characteristics of the adsorbed component of gum ghatti responsible for its oil–water interface advantages. *Foods and Food Ingredients Journal of Japan*, *213*(4), 372–376.

Mopper, K., Schultz, C. A., Chevolot, L., Germain, C., Revuelta, R., & Dawson, R. (1992). Determination of sugars in unconcentrated seawater and other natural waters by liquid chromatography and pulsed amperometric detection. *Environmental Science & Technology*, *26*(1), 133–138. http://dx.doi.org/10.1021/es00025a014

Osman, M. E., Menzies, A. R., Williams, P. A., Phillips, G. O., & Baldwin, T. C. (1993). The molecular characterisation of the polysaccharide gum from *Acacia senegal*. *Carbohydrate Research*, *246*(1), 303–318.

Qi, W., Fong, C., & Lamport, D. T. A. (1991). Gum arabic glycoprotein is a twisted hairy rope: A new model based on O-galactosylhydroxyproline

as the polysaccharide attachment site. *Plant Physiology*, *96*(3), 848–855. http://dx.doi.org/10.1104/pp.96.3.848

Strohalm, M., Kavan, D., Novak, P., Volny, M., & Havlicek, V. (2010). mMass 3: A cross-platform software environment for precise analysis of mass spectrometric data. *Analytical Chemistry*, *82*(11), 4648–4651, 10.1021/ac100818g.

Sturm, A., Bergwerff, A. A., & Vliegenthart, J. F. G. (1992). ^{1}H-NMR structural determination of the N-linked carbohydrate chains on glycopeptides obtained from the bean lectin phytohemagglutinin. *European Journal of Biochemistry*, *204*(1), 313–316. http://dx.doi.org/10.1111/j.1432-1033.1992.tb16639.x

Wishart, D., Bigam, C., Holm, A., Hodges, R., & Sykes, B. (1995). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR*, *5*(1), 67–81. http://dx.doi.org/10.1007/bf00227471