



Enzymatic hydrolysis studies of arabinogalactan-protein structure from Acacia gum: The self-similarity hypothesis of assembly from a common building block



D. Renard^{a,*}, L. Lavenant-Gourgeon^a, A. Lapp^b, M. Nigen^c, C. Sanchez^c

^a INRA, UR1268 Biopolymères Interactions Assemblages, F-44300 Nantes, France

^b Laboratoire Léon Brillouin, CEA Saclay, F-91191 Gif-sur-Yvette cedex, France

^c UMR1208 Ingénierie des Agropolymères et Technologies Emergentes, INRA-Montpellier SupAgro-CIRAD-Université Montpellier, 2 Place Pierre Viala, F-34060 Montpellier, France

ARTICLE INFO

Article history:

Received 25 April 2014

Received in revised form 16 June 2014

Accepted 18 June 2014

Available online 24 June 2014

Keywords:

HPSEC-MALLS

Small-angle neutron scattering

Circular dichroism

Arabinogalactan-protein

Acacia gum

Self-ordering

ABSTRACT

Arabinogalactan (AG) and arabinogalactan-protein (AGP) fractions were treated enzymatically using several proteases in acidic (pH 4) and alkaline (pH 7) conditions in order to go deeper insight into the structure and conformations of the two main fractions of *Acacia senegal* gum. Endoproteinase Glu-C, pepsin and phosphatase acid were thus used in acidic conditions while subtilisin A, pronase, trypsin, papain and proteinase K were used in alkaline conditions to cleave protein moieties of the two fractions. Structures of AG and AGP were probed using HPSEC-MALLS, small angle neutron scattering and far-UV circular dichroism.

Enzymes did not affect AG fraction structure whatever the pH conditions used, highlighting the inaccessibility of the peptide backbone and the remarkable stability of this fraction in acidic and alkaline conditions. This result was in agreement with the thin oblate ellipsoid model we previously identified for the AG fraction where the 43 amino-acid residues peptide sequence was supposed, based on spectroscopic methods, to be totally buried.

Contrary to AG fraction, AGP protein component was therefore cleaved using enzymes in alkaline conditions, the absence of enzymatic efficiency in acidic conditions being probably ascribed to long range electrostatic repulsions occurring between negatively charged AGP and enzymes at pH 4. The decrease of AGP molecular weight after hydrolysis in alkaline conditions went from $1.79 \times 10^6 \text{ g mol}^{-1}$ for control AGP to as low as $1.68 \times 10^5 \text{ g mol}^{-1}$ for papain-treated AGP. The overall structure of the enzyme-treated AGPs was found to be surprisingly quite similar whatever the enzyme used and close, with however some subtle differences, to AG unit. A tri-axial ellipsoid conformation was found in enzyme-treated AGPs and the two main preferential distances identified in the pair distance distribution function would claim in favor of rod-like or elongated particles or alternatively would indicate the presence of two particles differing in dimensions. The secondary structures content of control and enzyme-treated AGPs were similar, highlighting both the high rigidity of the protein backbone and the overall symmetry of AGP. This conclusion was reinforced by the more compact structures found when AGP was intact compare to the more elongated structures found when AGP was enzymatically cleaved.

Finally, the structural similarities found in enzyme-treated AGP together with the theoretical calculations to analytically probe the type of branching would suggest that AGP would be made of a self-similar assembly of two types of building blocks, the second being a five-fold repetition of the first one, for which palindromic amino acid sequence would ensure a self-ordering of carbohydrate moieties along the polypeptide chains. The cleavage would therefore lead to hydrolysed building blocks with similar secondary structures and conformations whatever the enzyme used.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

AGPs are the most structurally complex hydroxyproline-rich protein and are implicated to function in plant growth, development, signaling, and plant–pathogen interactions (Ellis, Egelund,

* Corresponding author. Tel.: +33 2 40 67 50 52; fax: +33 2 40 67 50 25.

E-mail addresses: drenard@nantes.inra.fr, denis.renard@nantes.inra.fr (D. Renard).

Schultz, & Bacic, 2010; Liu & Mehdy, 2007). They are found in plasma membranes, cell walls, and plant exudates (Seifert & Roberts, 2007). AGPs are defined by three criteria: the presence of type II arabino-3,6-galactan chains, a hydroxyproline-rich protein backbone, and the ability of most AGPs to bind to a class of synthetic phenylazo dyes, the β -glycosyl Yariv reagents (Clarke, Gleeson, Harrison, & Knox, 1979). AGP protein backbones are typically rich in hydroxyproline (Hyp) alternating with Ala, Thr, and Ser, whereas their carbohydrate moieties are mostly composed of galactose and arabinose with varying amounts of rhamnose, fucose, and glucuronic acid (Qu et al., 2008). The polysaccharide chains of AGPs are composed of β -(1,3)-galactan chains interrupted with β -(1,6)-galactopyranose (Gal) side chains and terminated mostly with arabinofuranose residues (Bacic, Churms, Stephen, Cohen, & Fincher, 1987; Tan, Qiu, Lampert, & Kieliszewski, 2004; Tan et al., 2010). Recently, Tryfona, Liang, Kotake, Tsumuraya, Stephens, and Dupree (2012) reported that *Arabidopsis* AGPs are similarly decorated with a linear β -(1,3)-galactan backbone with β -(1,6)-D-galactan side chains.

AGP from Acacia gum, an exudate produced by many trees and shrubs as a natural defense mechanism, was also defined as an arabinogalactan-protein as macromolecules also fulfilled the three criteria classifying AGPs (Churms, Merrifield, & Stephen, 1983; Connolly, Fenyo, & Vandeveld, 1987a, 1987b). Acacia gum is a highly heterogeneous material defined, on a hydrophobic separation criterion, by three molecular fractions identified as an arabinogalactan-peptide (AG), an arabinogalactan-protein (AGP) and a glycoprotein (GP) fraction (Randall, Phillips, & Williams, 1989; Renard, Lavenant-Gourgeon, Ralet, & Sanchez, 2006). Analysis of the fractions showed that each contained similar proportions of the various sugars and differed essentially in their molecular masses and protein contents. The tertiary structure of AGP has been described in literature in terms of a wattle-blossom macromolecular assembly by virtue of which few (~ 5) discrete polysaccharide blocks of $M_w \sim 2 \times 10^5 \text{ g mol}^{-1}$ are held together by a short peptide backbone chain (Fincher, Stone, & Clarke, 1983). More recently, Mahendran, Williams, Phillips, Al-Assaf, and Baldwin (2008) identified smaller carbohydrate blocks of $\sim 4.5 \times 10^4 \text{ g mol}^{-1}$ linked by O-serine and O-hydroxyproline residues to the polypeptide chain of approximately 250 residues in length. This result, using a chemical treatment assumed to hydrolyze only the polypeptidic backbone of AGP, contradicted the previous enzyme degradation studies, using pronase to degrade polypeptidic backbone, which concluded that the carbohydrate blocks were of the order of $\sim (1.8\text{--}4.5) \times 10^5 \text{ g mol}^{-1}$ (Aoki, Al-Assaf, Katayama, & Phillips, 2007; Beltran et al., 2005; Chikamai, Banks, Anderson, & Weiping, 1996; Churms et al., 1983; Connolly, Fenyo, & Vandeveld, 1988; Duvallet, Fenyo, & Vandeveld, 1989; Fincher et al., 1983; Osman, Menzies, Williams, Phillips, & Baldwin, 1993; Randall et al., 1989). These apparent contradictory results coming from protease or chemical treatments on AGP could mean that the enzymatic kinetics conditions (e.g. pH, T, enzyme:substrate ratio, time of degradation) would be responsible for the difference of molecular weights found between chemical and enzymatic treatment. A closer examination of the enzymatic kinetics conditions in previous literature was in favor of a complete hydrolysis of the protein in the AGP component after 24 h of pronase incubation (Aoki et al., 2007; Flindt, Al-Assaf, Phillips, & Williams, 2005; Mahendran et al., 2008). Pronase is a mixture of several proteases with unspecific endogenous and exogenous cleavage of proteins, which generally proceed to the level of single amino acids. This is not the case when chemical treatment is used, such as sodium hydroxide/sodium borohydride used by Mahendran et al. (2008), where the chemical compounds specifically cleave O-glycans at polysaccharide-polypeptide linkages involving Ser or Thr residues

but not those involving Hyp residue. Why thus AGP chemical treatment which is more specific than pronase for peptidic cleavages can lead to lower molecular weight components? One strong explanation to this apparent contradiction would come from the works of Aoki et al. (2007) and Al-Assaf, Sakata, McKenna, Aoki, and Phillips (2009) who clearly demonstrated that AGP is not composed of a single polypeptide chain with various carbohydrate units attached to it. The AGP would be mainly composed of arabinogalactan (AG) units associated through hydrophobic association and this is why the protein sequence and sugar contents is similar in both fractions but different to that in the glycoprotein (GP) fraction as reported previously (Mahendran et al., 2008; Renard et al., 2006; Williams, Phillips, & Stephen, 1990). However, key to the stability and coherent of the AGP structure would be a small proportion of low molecular weight highly proteinaceous components often reported in the literature as the GP fraction (Randall et al., 1989). This interpretation would reconcile chemical and enzymatic degradation where the presence of polypeptide backbones with different lengths and as a consequence different carbohydrate blocks attached to it would lead to different molecular weight components after treatment whatever the conditions used. Al-Assaf et al. (2009) provided evidence that the high molecular weight fraction (AGP) in gum arabic was indeed highly associated molecular structure. In addition, when subjected to physical treatments, such as filtration or high pressure homogenization, AGP can be dissociated to yield the basic AG molecular units which make up the gum and a low molecular weight fraction highly proteinaceous. Further evidence to support this proposal came from enzyme digestion of gum arabic which was shown to be dependent on the starting molecular weight. The reduction of the molecular weight following enzyme digestion was as low as 1.8×10^5 or $5.0 \times 10^5 \text{ g mol}^{-1}$ when molecular weight of *A. senegal* gum probed by HPSEC-MALLS was processed as one peak (Al-Assaf et al., 2009). Finally, the recent results reported by Mahendran et al. (2008) clearly established by electrophoresis after deglycosylation of gum arabic the presence of two polypeptidic chains, one corresponding to a core protein of ~ 250 amino acid residues that was consistent with the literature for AGP core proteins (Williams & Langdon, 1995) and a second putative core protein of 45 amino acid residues that would be consistent with the proteinaceous component associated to the AG gum fraction (Mahendran et al., 2008). We identified recently two types of conformations for AGP in solution, a low molar mass population with long-chain branching and a compact conformation and a high molar mass population with short-chain branching and an elongated conformation (Renard, Garnier, Lapp, Schmitt, & Sanchez, 2012). The maximum dimension of AGP was found to be 64 nm and electron microscopy observations clearly revealed the presence of interacting branched chains, aggregate chains and ring-like structures that rendered the identification of single AGP macromolecules very difficult (Renard, Garnier, Lapp, Schmitt, & Sanchez, 2013). These structural studies would lead to the main conclusions that AGP could be an association of at least two different units and that these molecular associations would be prone to aggregation when adsorbed on solid surfaces.

The main objectives of the present study were to probe the structure of the AGP fraction from Acacia gum in solution, after purification by hydrophobic interaction chromatography, and enzymatic degradation using several proteases, in order to try to reveal new insights into the structure of this complex assembly. The combination of high-performance size exclusion chromatography coupled to on-line multi-angle laser light scattering (HPSEC-MALLS), small angle neutron scattering (SANS) and circular dichroism (CD) measurements was performed in order to reveal some fundamental knowledge of the carbohydrate-polypeptide linkages occurring in AGP.

2. Materials and methods

2.1. Materials

Spray-dried Acacia gum (lot 97J716) from *A. senegal* trees was a gift from CNI company (Rouen, France). Before purification, Acacia gum was dialyzed against deionized water and then freeze-dried. All chemicals were of analytical grade and were obtained from Merck, Sigma and Aldrich.

2.2. Purification of AGP molecular fraction

The Acacia gum arabinogalactan-protein (AGP) molecular fraction was purified by hydrophobic interaction chromatography (HIC), as previously described (Renard et al., 2006). Purified AGP was dialysed and then freeze-dried before further characterization.

2.3. Acidic and alkaline enzyme hydrolyses

Both acidic and alkaline enzymatic treatments were applied on AGP (and AG) molecular fraction from Acacia gum. For acidic hydrolyses, AGP and enzymes were dissolved in 0.1 M acetate buffer pH 4. Enzymes used were endoproteinase Glu-C from *Staphylococcus aureus* strain V8 (lot 1278409, Fluka Biochemika), pepsin from porcine gastric mucosa (lot 035K76701, Sigma–Aldrich) and phosphatase acid type IV-S from potato (lot 026K7014, Sigma–Aldrich). For alkaline hydrolyses, AGP and enzymes were dissolved in 0.1 M phosphate buffer pH 8. Enzymes used were subtilisin A, type VIII from *Bacillus licheniformis* (lot 016K1295, Sigma–Aldrich), pronase from *Streptomyces griseus* (lot 084K1843, Sigma–Aldrich), trypsin, TPCK treated, from bovine pancreas (lot 026K7770, Sigma–Aldrich), papain from *Carica papaya* (lot 108014, Boehringer Mannheim) and proteinase K from *Tritirachium album* (lot P-6556, Sigma–Aldrich). For all assays, 3 mg of substrate (AG or AGP) was dissolved in 0.3 mL of the appropriate buffer and 0.3 mL of enzyme at a concentration of 0.2 mg/mL was added, giving a final substrate concentration of 5 mg/mL and a substrate:enzyme ratio of 25. Hydrolyses were performed one night at $T = 37^\circ\text{C}$. Controls were performed using untreated AG and AGP. Experiments were triplicated and samples were then stored at 4°C or extensively dialysed against distilled water and freeze-dried before further analyses.

2.4. HPSEC-MALLS

The determination of molecular weight and size distributions of AGP (and AG) after enzymatic hydrolyses were performed by coupling on line a high-performance size exclusion chromatography to a multi-angle laser light scattering detector, a differential refractometer and a differential viscosimeter. AGP (or AG) dispersions at 0.05 wt% stored at 4°C and containing enzymes or solubilized after freeze-drying at 0.04 wt% into 50 mM NaNO_3 buffer containing 0.02% NaN_3 as preservative were filtered through 0.2 μm Anotop membranes (Anotop, Alltech, France), and injected on a HPSEC system constituted of a Shodex OH SB-G pre-column followed by two Shodex OH-pack 804 HQ and 805 HQ columns used in series. The samples were eluted at 0.7 mL min^{-1} with a 50 mM NaNO_3 solution. On-line molecular weight and intrinsic viscosity determinations were performed at room temperature using a multi-angle laser light scattering (MALLS) detector (mini-Dawn, Wyatt, Santa Barbara, CA, operating at three angles: 41° , 90° and 138°), a differential refractometer (ERC 7547 A) ($\text{dn/dc} = 0.153\text{ mL g}^{-1}$) and a differential viscosimeter (T-50A, Viscotek). Weight (M_w) and number (M_n) average molar mass (g mol^{-1}), and radius of gyration (R_g , nm) were calculated using Astra 1.4 software (Wyatt). Intrinsic viscosity $[\eta]$ was calculated using TriSEC software, version 3.0 (Viscotek). No differences were observed in the results whatever the physical state

and the presence or not of enzyme in the samples (hydrated samples containing enzymes vs. solid samples without enzymes).

2.5. SANS experiments and data treatment

SANS experiments and data treatment were performed as described previously (Sanchez, Schmitt, Kolodziejczyk, Lapp, Gaillard, & Renard, 2008). Briefly, scattering experiments on filtrated AGP (intact or enzymatically hydrolysed) ($C = 1\text{ wt\%}$) dissolved in 50 mM NaCl solution in 99.8% D_2O in order to get the form factor $P(q)$ were performed on the PAXY instrument at the Laboratoire Léon Brillouin (LLB) (Saclay, France) using two spectrometer configurations: $\lambda = 8\text{ \AA}$ (incident wavelength), $d = 2\text{ m}$ (distance of the sample to the detector) and $\lambda = 8\text{ \AA}$ and $d = 6.74\text{ m}$. The wave-vector $q = 4\pi/\lambda \sin(\theta/2)$, θ being the scattering angle, ranged between 4×10^{-2} and 3 nm^{-1} . The data were normalized for transmission and sample path length and divided by the water spectrum. An absolute intensity scale was obtained using the absolute value of water intensity in units of cross-section (cm^{-1}). A subtraction of the appropriate incoherent background, taking H/D exchange on AGP into account, was realized. The form factor $P(q)$ of intact or cleaved AGP was obtained at 50 mM NaCl concentration where intermolecular interactions were negligible and therefore $S(q) = 1$.

The pair distance distribution function $P(r)$ together with the structural parameter derived from $P(r)$, i.e. the maximum dimension of the particle (D_{max}), was computed from the form factor by the indirect Fourier transform program GNOM (Svergun, 1992; Svergun, Semenyuk, & Feigin, 1988). Fit of experimental AGP form factor by different theoretical form factors was assayed using the SansView software v2.0.1 (<http://danse.chem.utk.edu/sansview.html>).

2.6. Far UV-CD

The circular dichroic spectra of intact or cleaved AGP in 0.1 M phosphate buffer pH 8 were recorded on a CD6 Dichrograph Instrument (Jobin-Yvon, Longjumeau, France) at 25°C , calibrated using ammonium d_{10} -camphorsulfonic acid. Spectra were averaged over two scans with a bandwidth of 1 nm and step resolution of 0.5 nm. All spectra were reported in terms of theta machine units (mdeg) within the 185–260 nm spectral area using a 1-mm path-length cylindrical cuvette. AGP concentrations used (after filtration on 0.2 μm pore sizes) were of 0.5 wt%. All spectra were corrected for the solvent (buffer + enzyme); noise reduction was applied according to the Jobin-Yvon software. CD spectra acquisition was repeated twice.

CD spectra were analyzed in terms of secondary structure content using Selcon3 program (Sreerama & Woody, 1993) in Dichroweb (Lobley, Whitmore, & Wallace, 2002) and using the protein data set 3 (185–240 nm, 37 proteins), 6 (185–240 nm, 17 proteins) and 7 (190–240 nm, 42 proteins) as the reference set (Whitmore & Wallace, 2004, 2008).

3. Results

The enzymatic hydrolyses of the two main molecular fractions, AG and AGP, from *A. senegal* gum, were performed in acidic (pH 4) and alkaline (pH 8) conditions using enzymes in the appropriate pH-optima conditions. Table 1 summarized the mode of action of each enzyme used in 0.1 M acetate buffer pH 4 and 0.1 M phosphate buffer pH 8 conditions. The enzymatic conditions used in this study thus covered a wide range of possibility for peptide and ester bonds cleavage in the polypeptidic backbone of AG and AGP unless the polypeptidic chains were accessible to enzymes in the

Table 1

Mode of action of the enzymes used to cleave polypeptidic backbones of arabinogalactan (AG) and arabinogalactan-protein (AGP) molecular fractions from Acacia gum.

Buffer conditions	Enzyme	Specificity ^a	Proportion in AGP (%) ^b
0.1 M acetate buffer, pH 4	Endoproteinase Glu-C	Glu, Asp-protease	6.5
0.1 M acetate buffer, pH 4	Pepsin	Phe, Tyr, Trp protease	4.4
0.1 M acetate buffer, pH 4	Phosphatase acid	Esterase (one ester bond and a phosphate group)	– ^c
0.1 M phosphate buffer, pH 8	Subtilisin A	Serine protease	12.2
0.1 M phosphate buffer, pH 8	Pronase	Mixture of several proteases	100
0.1 M phosphate buffer, pH 8	Trypsin	Arg, Lys protease	3.6
0.1 M phosphate buffer, pH 8	Papain	Arg, Lys, Glu, His, Gly, Tyr protease	20.7
0.1 M phosphate buffer, pH 8	Proteinase K	Aliphatic, aromatic, hydrophobic protease	24

^a Glu: glutamic acid; Asp: aspartic acid; Phe: phenylalanine; Met: methionine; Leu: leucine; Trp: tryptophan; Arg: arginine; Lys: lysine. Aromatic: Phe, Tyr, Trp. Hydrophobic: Ile, Leu, Val, Cys, Met. Aliphatic: Gly, Ala.

^b From Renard et al. (2006).

^c Unknown.

two fractions. Aliquots of treated samples were then analyzed by HPSEC-MALLS, SANS and far UV-CD as described above.

3.1. HPSEC-MALLS measurements

It was first observed that acidic and alkaline enzymatic treatments did not affect the structure of AG whatever the enzyme used (data not shown). HPSEC-MALLS results for AG before and after enzymatic hydrolyses gave a molecular weight M_w of $2.82 \times 10^5 \text{ g mol}^{-1}$, a polydispersity index M_w/M_n of 1.26 and a radius of gyration R_g of $8 \pm 5 \text{ nm}$, values in close agreement with those reported previously (Sanchez et al., 2008). This result would mean that the peptide chain of AG would not be accessible to enzyme whatever the pH conditions and enzymes used. From our previous result and those obtained by Mahendran et al. (2008), the protein core in the AG component would consist of 43 (or 45) amino acids (Mahendran et al., 2008; Renard et al., 2006). The protein core would be however completely buried by the oligo- and polysaccharide blocks attached to hydroxyproline (Hyp), serine (Ser) or threonine (Thr) residues, and as a consequence, not accessible to enzymes. Indeed, arabinose and rhamnose were found to be at the periphery of the macromolecules (Anderson, Hirst, & Stoddart, 1966; Mahendran et al., 2008). Moreover, it was clearly established by Mahendran et al. (2008) that the chemical treatment used to degrade proteinaceous components in Acacia gum did not affect the protein-deficient AG reinforcing the fact that oligosaccharides and polysaccharide blocks were preferentially located at the periphery of the macromolecule. In addition, it was also interesting to note that Williams and Phillips (2000) demonstrated that the third minor glycoprotein (GP) fraction was unaffected by a protease treatment. The question for which we should answer through this study is the following: if AG and GP fractions are unaffected by protease treatments, at least in alkaline conditions, and if AGP is a molecular association of AG and GP units (Al-Assaf et al., 2009), which will be the reactivity of AGP toward protease treatment?

The second results we obtained were that AGP was also unaffected by acidic enzymatic treatment whatever the enzyme used. The molecular weight M_w and the radius of gyration R_g were $1.83 \times 10^6 \text{ g mol}^{-1}$ and 33 nm, respectively, values very close to those reported previously for native AGP (Renard et al., 2006, 2012). The absence of reactivity of enzymes used in acidic conditions, endoproteinase Glu-C, pepsin and phosphatase acid, would mean that the enzymes were unspecific toward amino acid residues in the polypeptide chains of AGP or were unable to proceed due to steric hindrance. The first hypothesis was not true since all amino acid residues cleaved by the three enzymes were present in AGP (Renard et al., 2006). The second hypothesis would therefore be more conceivable, as previously discussed for the AG fraction from Acacia gum. However, all previous enzymatic studies on AGP and whole gum performed in alkaline conditions demonstrated partial

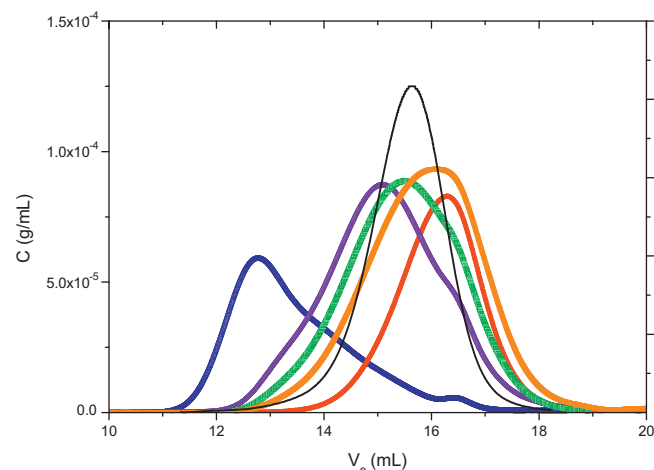


Fig. 1. HPSEC chromatogram showing the elution profile monitored by refractive index (C , g/mL) and light scattering (M_w , g mol⁻¹) for Acacia gum arabinogalactan-protein (AGP) before and after enzymatic hydrolyses as a function of elution volume (V_e , mL) obtained from HPSEC coupled to on-line viscosimeter and multi-angle light scattering (MALLS) detector. HPSEC chromatogram for arabinogalactan-peptide (AG) molecular fraction is displayed for comparison (continuous line). AGP (○), AGP + proteinase K (□), AGP + subtilisin (△), AGP + pronase (◇), AGP + papain (*).

degradation of AGP. The steric hindrance would thus not be the main reason to the absence of enzymes reactivity in acidic conditions. One possible explanation to understand these results would be that the low content of amino acid residues susceptible to be cleaved by enzymes (Asp, Glu, Phe, Tyr, Trp residues), residues that in addition could also be buried and inaccessible to enzyme, would not affect the whole structure of AGP. The second probable explanation would be that non covalent interactions between sugar moieties and/or sugar moieties/amino acid residues would take place maintaining the overall structure of AGP even in the presence of enzyme cleavages.

AGP was therefore affected by enzymatic treatments in alkaline conditions as illustrated in Fig. 1 which displayed the elution profiles of AGP obtained by HPSEC-MALLS after one night of enzymatic treatment at $T = 37^\circ \text{C}$. HPSEC chromatograms for control arabinogalactan-peptide (AG) and arabinogalactan-protein (AGP) were displayed for comparison. It was first evidenced from the maxima in the concentration peak as a function of elution volume that AGP was degraded to some extent depending on the enzyme used. In addition, one single peak was observed whatever the enzyme used with however a higher polydispersity index M_w/M_n value compare to control AGP, indicating that AGP macromolecules would be uniformly cleaved in solution giving rise to homogeneous fragments. While control AGP had a M_w of $1.79 \times 10^6 \text{ g mol}^{-1}$, a decrease of M_w was observed after enzymatic hydrolyses except for

Table 2
Summary of arabinogalactan-protein (AGP) from Acacia gum, before and after enzymatic hydrolyses using proteinase K, subtilisin, pronase and papain, structural parameters determined by HPSEC-MALLS and SANS, or calculated.

Structural parameters	AGP	AGP + pronase	AGP + proteinase K	AGP + subtilisin	AGP + papain
<i>HPSEC-MALLS</i>					
Non-hydrolysed AGP ^a	100	23.5	11.4	5.8	0
M_w (g mol ⁻¹)	1.79×10^5	6.0×10^5	4.9×10^5	3.67×10^5	1.68×10^5
M_w/M_n	1.32	1.43	1.47	1.75	1.46
R_g (nm) ^b	33	21.2	18	17.7	11
$[\eta]$ (mL g ⁻¹)	73.5	25.4	26.3	18.2	15.4
R_h (nm) ^c	27.5	13.4	12.7	10.2	7.4
$\rho(R_g/R_h)$	1.2	1.6	1.4	1.7	1.5
<i>SANS</i>					
$P(0)$ (cm ⁻¹)	3.3	1.77	1.52	1.25	1.22
R_g (nm) ^d	18.4	14.8	14.4	14.5	16.2
R_g (nm) ^e	18.5	14.7	14.6	15.6	16.8
D_{max} (nm) ^e	64	49.4	49.5	51.5	54.2
R_a (nm) ^f	0.3	0.4	0.5	0.3	2
R_b (nm) ^f	9.3	7.1	6.5	5.8	4.2
R_c (nm) ^f	32	27	26.5	23	27
χ^2 ^g	2.67	0.84	0.85	0.72	2.26

^a Estimation of non-hydrolysed AGP from HPSEC-MALLS (see text for details).

^b From SEC-MALLS by light scattering.

^c Calculated from the equation $[\eta] = 6.308 \times 10^{24} R_h^3 / M_w$.

^d From the Guinier region of the form factor $P(q)$.

^e From the pair distance distribution function $P(r)$.

^f R_a , R_b , R_c , semi-axis of the triaxial ellipsoid model that best fits the form factors $P(q)$.

^g Parameter of best fit quality.

trypsin treatment (data not shown) for which M_w was unchanged (Table 2). A molecular weight M_w decrease from 1.79×10^6 to as low as 1.68×10^5 g mol⁻¹ was observed for papain-treated AGP. Moreover, the intrinsic viscosity $[\eta]$ value decreased from 73.5 to 15.4 mL g⁻¹. A rough estimation of the enzyme efficiency was done by calculating the percentage of non-hydrolysed AGP after one night at $T = 37^\circ\text{C}$. For that, it was considered that intact AGP was still present after enzymatic hydrolysis if macromolecules with M_w higher than $\sim 1 \times 10^6$ g mol⁻¹ (i.e. $V_e \sim 13.5$ mL) were present. The ratio of the sum of concentrations for macromolecules eluted for $M_w > 1 \times 10^6$ g mol⁻¹ over total concentrations given by the refractometer gave the results in Table 2 where papain appeared to be the most efficient enzyme with 100% hydrolysed AGP while pronase was found to be the less efficient with 23.5% residual intact AGP after one night of enzymatic reaction. These rough estimations of enzymatic efficiencies were however in accordance with the lowest M_w found for papain-treated AGP, $M_w = 1.68 \times 10^5$ g mol⁻¹, while the highest one was found for pronase-treated AGP, $M_w = 6.0 \times 10^5$ g mol⁻¹.

These results were in partial agreement with previous literature data stating that pronase treatment, for time of reaction equal or higher than one night and enzyme:substrate ratio of 56 or 60, lead to $M_w \sim 2\text{--}3 \times 10^5$ g mol⁻¹ and $[\eta] \sim 11\text{--}12$ mL g⁻¹, values close to those of AG unit (Al-Assaf et al., 2009; Connolly et al., 1988; Elmanan, Al-Assaf, Phillips, & Williams, 2008). The higher M_w found in our study was clearly explained by the lower enzyme:substrate ratio used. The absence of cleavage of peptidic bonds using trypsin could be related, as for enzymes in acidic conditions, to the very low content of arginine and lysine residues in AGP (Table 1), residues that in addition could also be buried and inaccessible to enzyme (Renard et al., 2006). An analysis of the conformation adopted by AGP before and after enzymatic treatment revealed that enzyme-treated AGP was very similar to AG conformation contrary to control AGP that adopted two different conformations (Fig. 2). AGP was therefore found to adopt a compact conformation for the low molar mass population and an elongated conformation for the high molar mass population (Renard et al., 2012). The HPSEC chromatograms showing the M_w dependence with the elution volume V_e clearly revealed the great similarity found between control AG and hydrolyzed AGP, highlighting the hypothesis of

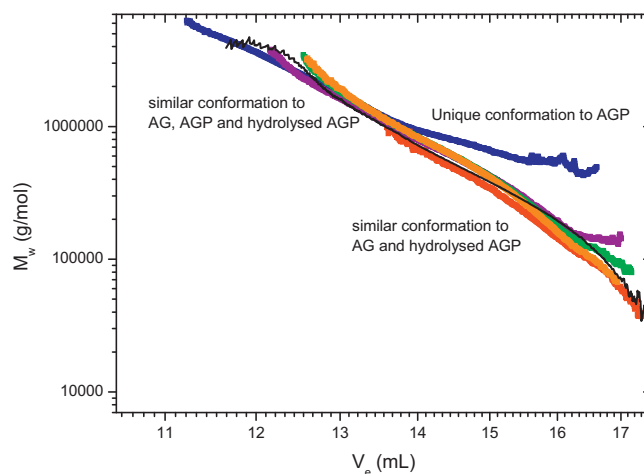


Fig. 2. HPSEC chromatogram showing the molecular weight (M_w , g mol⁻¹) dependence with the elution volume (V_e , mL) for Acacia gum arabinogalactan-protein (AGP) before and after enzymatic hydrolyses. Molecular weight dependence for arabinogalactan-peptide (AG) molecular fraction is displayed for comparison (continuous line). AGP (—○—), AGP + proteinase K (—□—), AGP + subtilisin (—△—), AGP + pronase (—◇—), AGP + papain (—*—).

the dissociation of AGP into building blocks similar to AG units after enzymatic treatment. Differences in M_w of hydrolyzed AGP evidenced several degree of hydrolysis according to enzyme, as demonstrated by the content of the non-hydrolysed AGP (Table 2), and highlighted the existence of a building block with low M_w , i.e. 1.68×10^5 g mol⁻¹ for papain-treated AGP. The association of these building blocks leading to AGP macromolecular assembly could originate from covalent linkages and/or from hydrophobic associations and hydrogen bonding, covalent linkages being a reasonable hypothesis considering the mode of action of proteases used in this study and the mixed composition of AGP containing glycomodules of both extensin and arabinogalactan-glycoproteins (Goodrum, Patel, Leykam, & Kieliszewski, 2000). In particular, disulfide intermolecular bridges, neither however identified in AGP, and/or isodityrosine intramolecular cross-links, found in extensins (Kieliszewski & Lampert, 1994), could be present in AGP

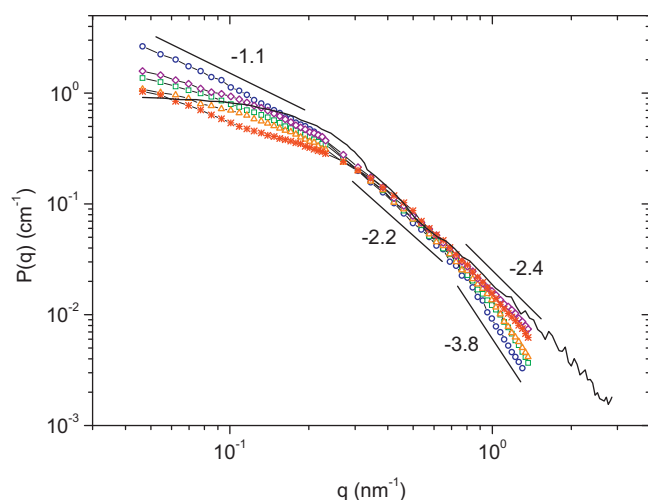


Fig. 3. Form factors obtained by SANS at 25 °C and 50 mM NaCl on arabinogalactan-protein (AGP) dispersions ($C = 1$ wt%) from Acacia gum before and after enzymatic hydrolyses. (AG) molecular fraction is displayed for comparison (continuous line). AGP (○), AGP + proteinase K (□), AGP + subtilisin (△), AGP + pronase (◇), AGP + papain (*). Form factor on arabinogalactan-peptide (AG) (—) is displayed for comparison. High q range of form factors can be described by power law exponents which range is indicated in the figure (−3.85 for AGP and 2.4 for AGP + pronase). In the intermediate q range, power law exponents were comprised between 2.13 (AGP) and 2.2 (AGP + papain).

macromolecular assembly giving a better efficiency toward hydrolysis to papain. The unique globular conformation however adopted by control AGP for V_e higher than 13 mL compared to control AG and enzyme-treated AGP was in accordance with our previous results stating that AGP would behave in solution as a branched polymer with conformations ranging from globular to elongated shape depending on the M_w range considered (i.e. size of the carbohydrate branches) (Renard et al., 2012). From these preliminary results, it was concluded that enzyme-treated AGPs adopted one single conformation in solution close to the AG unit whatever the molecular weight, contrary to control AGP that adopted two distinct conformations depending on the molecular weight. The apparent similar conformation adopted by enzyme treated AGPs with AG unit should therefore be shaded by the fact that lower M_w compare to AG were obtained for hydrolysed AGPs. This important result would therefore mean that, structural blocks, different from AG units, could be at the origin of the macromolecular assembly leading to the complex AGP macromolecule.

3.2. SANS measurements

Additional informations on conformation and structure were probed using small angle neutron scattering experiments on a distance $d = 2\pi/q$ ranging from 2 to ~150 nm. Form factors $P(q)$ of enzyme-treated AGP and control AGP in 50 mM NaCl as a function of the wave vector q (nm^{-1}) were displayed in Fig. 3. At low scattering wave vectors q , the scattering function $P(q)$ vs. q of control AGP followed a power law with an exponent of 1.1, indicative of a linear conformation typical of a rod or an elongated cylinder, as described previously (Renard et al., 2012). Enzyme-treated AGPs displayed lower scattering intensity at low q values indicative of lower M_w compare to control AGP, in agreement with M_w determined by HPSEC-MALLS. This decreasing molecular mass with enzyme treatment was further confirmed through the radii of gyration R_g values that decreased from 18.4 nm for control AGP to 14.4 nm for proteinase K-treated AGP. In addition, it was noticed that R_g values were quite constant whatever the enzyme used with however an intermediate value found for papain-treated AGP.

These preliminary results would favor again a consensus in the whole conformation of enzyme-treated AGP whatever the type of enzyme used. At intermediate scattering wave vectors q , the scattering function of control AGP and enzyme-treated AGP were superimposed and followed a power law with an exponent value of 2.2. A value of 2 is indicative of the presence of an isolated population of disk-like particles, and more generally of a 2D particles morphology (Svergun & Koch, 2003), or of a fractal structure. The affinity of control AGP and enzyme-treated AGP for the solvent was thus estimated to be of 0.45 (1/exponent value), value corresponding to a theta solvent, intermediate between a poor affinity (0.33) and a very good affinity (0.6) for the solvent. In the high q range, where the local structure of scattering objects is probed, scattering intensity followed a power law with an exponent ranging from 3.8 for control AGP to 2.4 for control AG and papain and pronase-treated AGPs. These exponent values clearly indicated that control AGP had a fractal surface structure while enzyme-treated AGPs and control AG had an internal structure rather dense and ramified (Renard et al., 2012; Sanchez et al., 2008). It was interesting to notice that the scattering intensity $I(q)$ vs. q of control AG was superimposed to enzyme-treated AGPs on a q range ranging from ~0.3 to ~0.8 nm^{-1} , covering a distance between 2.5 and 20 nm (Fig. 3). This was particularly true for papain and pronase-treated AGPs, slight deviations being observed at high q for proteinase K and subtilisin-treated AGPs. The slight differences observed at low q values highlighted the structural changes occurring after enzymatic treatment on AGP macromolecules compare to control AG. The structural differences identified by SANS and not by HPSEC-MALLS when control AG was compared to enzyme-treated AGPs were related to the distance probed using neutron scattering compare to light scattering. The higher distances therefore probed by light scattering masked some subtle structural differences observed by neutron scattering at lower distances. These results would reinforce the hypothesis of the association of structural building blocks different from AG units leading to the formation of AGP macromolecular assembly.

Since the arabinogalactan-peptide (AG) and arabinogalactan-protein (AGP) fractions were described as a thin oblate ellipsoid (Sanchez et al., 2008) and a tri-axial ellipsoid or an elliptical cylinder (Renard et al., 2012), respectively, a fit of the enzyme-treated AGPs form factors was attempted taking into account the fact that HPSEC-MALLS revealed elongated conformations and SANS identified a rather elongated structure and particles with 2D morphologies. Finally, it appeared that a tri-axial ellipsoid form factor reasonably fitted experimental data (Fig. 4). Details of the semi-axes values of the tri-axial ellipsoids were given in Table 2. The fits were quite satisfying and reasonably approached the overall characteristic of the form factor. It was therefore surprising to notice that, even if a 10 fold reduction of molecular mass occurred after papain treatment on AGP for instance, the overall structure of the macromolecule was maintained. The structural similarities observed before and after enzymatic treatment, leading to a tri-axial ellipsoid morphology, even if subtle differences were probed at small distances (i.e. high q values by SANS), could be related to the lack of enzyme specificity and/or the steric hindrance, due to the presence of large carbohydrate blocks attached to the polypeptide backbone, avoiding the enzymes to freely diffuse in the core of the AGP. The consensus in the mode of enzyme action, due to the repetition of amino acid sequences in AGP and also steric hindrance, would therefore lead to the same structural blocks after degradation whatever the type of enzyme used.

Additional information about the structural morphology of the control and enzyme-treated AGPs came from the $P(r)$ functions, shown in Fig. 5, obtained by the GNOM treatment of the SANS data. From this function, the maximum distances ($D_{\max}$) found for control AGP and enzyme-treated AGPs varied from 64 to 49.4 nm

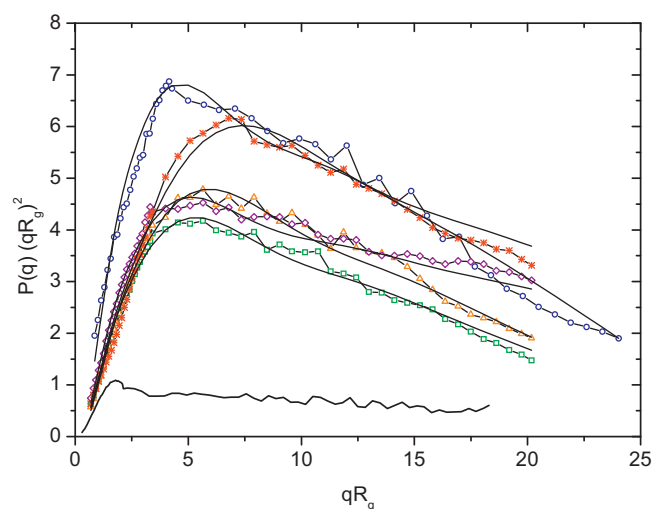


Fig. 4. Kratky-type plots of form factors and curve-fitting by a tri-axial ellipsoid model of form factors arabinogalactan-protein (AGP) dispersions (from Acacia gum) before and after enzymatic hydrolyses. AGP (○), AGP + proteinase K (□), AGP + subtilisine (△), AGP + pronase (◇), AGP + papain (*). Form factor on arabinogalactan-peptide (AG) (—) is displayed for comparison.

(Table 2). In addition, R_g values calculated from the distance distribution functions were in very good agreement with the R_g values obtained in the Guinier analysis (Table 2). Note that these values were used to fit the experimental form factors by the tri-axial ellipsoid theoretical form factor. The extended, tail-like decay of the $P(r)$ functions indicated elongated AGPs that vary from globular/ellipsoid to cylindrical (Glatter, 1979, 2002), conformations in agreement with those reported from HPSEC-MALLS data. Two preferential distances, three distances in the case of enzyme-treated AGPs, were therefore identified as two, or three, maxima were observed in the $P(r)$ functions. The second maxima were common to control and enzyme-treated AGPs and corresponded to a half-distance of about 32 nm. The first maxima was found to be at a distance of 10 nm for control AGP while a distance of about 7.5 nm was obtained for enzyme-treated AGPs except

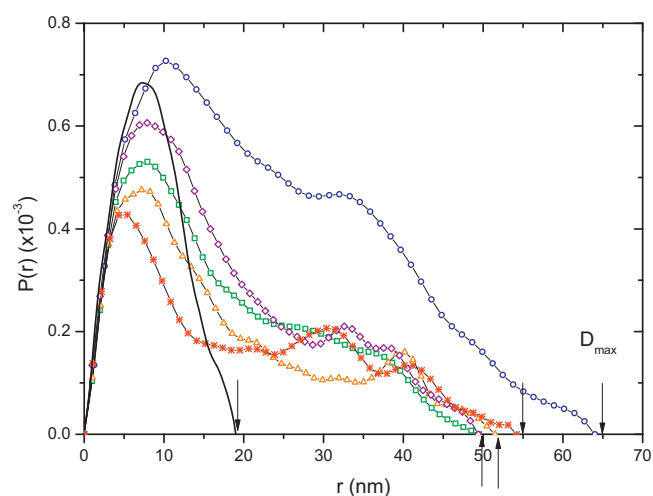


Fig. 5. Pair distance distribution function $P(r)$ calculated by indirect Fourier transform for the arabinogalactan-protein (AGP) from Acacia gum form factor using GNOM software before and after enzymatic hydrolyses. AGP (○), AGP + proteinase K (□), AGP + subtilisine (△), AGP + pronase (◇), AGP + papain (*). Form factor on arabinogalactan-peptide (AG) (—) is displayed for comparison. Maximum dimensions (D_{max}) indicated by arrows were comprised between 64 nm (AGP) and 49.5 nm (AGP + subtilisine). The R_g obtained from the integrals of the profiles were comprised between 18.5 nm (AGP) and 14.6 nm (AGP + proteinase K) (see Table 2).

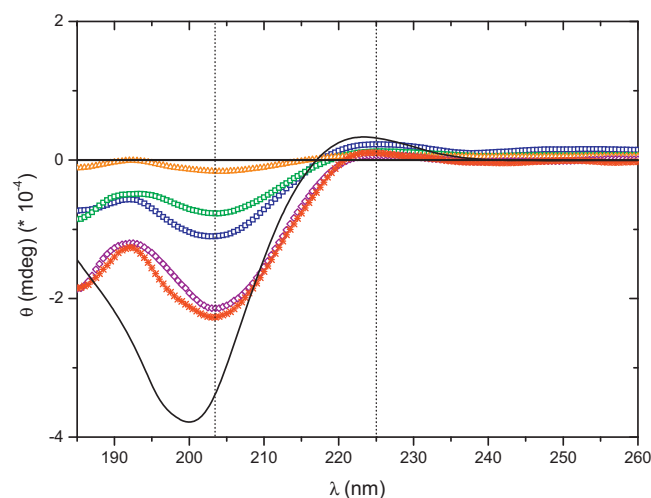


Fig. 6. Far-UV circular dichroism spectra obtained at 25 °C and phosphate buffer pH 8 for arabinogalactan-protein (AGP) dispersions from Acacia gum before and after enzymatic hydrolyses. AGP (○), AGP + proteinase K (□), AGP + subtilisine (△), AGP + pronase (◇), AGP + papain (*), (—) collagen.

papain-treated AGP for which a lower distance of 5 nm was found. Note that the first maxima were close in shape and dimension to the control AG fraction, with however a larger asymmetric distribution in the case of enzyme-treated AGPs. The third maxima, only identified in enzyme-treated AGPs, corresponded to a half-distance of about 40 nm. The distances identified in the $P(r)$ functions would correspond, in the case of rod-like or elongated particles, to a cross-sectional dimension of 10 nm (or 7.5 or 5 nm for enzyme-treated AGPs) and a major dimension length of 64 nm for control AGP and about 50 nm for enzyme-treated AGPs. The cross-sectional dimensions were therefore in very good agreement with the R_b semi-axis values calculated in the case of a tri-axial ellipsoid model (Table 2). Alternatively, the presence of two or three maxima in the $P(r)$ functions can also indicate the presence of two or three particles differing in dimensions or, in the particular case of proteins, distances separating covalently linked domains in a single protein or monomers in a multimeric protein (Wang et al., 2009). In the particular case of AGP, the macromolecular assembly could come from the association of covalently-linked structural building blocks.

3.3. Far-UV circular dichroism

The molecular scale of AGP and enzyme-treated AGPs was studied using far-UV circular dichroism in order to probe the secondary structures of the polypeptide backbone after protease cleavage. A collagen standard was also displayed for comparison (Fig. 6). A minimum at 203 nm and a maximum at 225 nm were identified in the CD spectra in the 185–260 nm spectral region of AGP and enzyme-treated AGPs. Collagen, an example of protein with a PPII helix conformation, displayed a strong negative band at 200 nm and a weak positive band at 220 nm (Fig. 6). A minimum at 205 nm and a maximum at 226 nm were previously found in a polyhydroxyproline standard, corresponding to a polyproline II (PPII) conformation (Renard et al., 2006). The PPII-helix structure is generally found on the surfaces of proteins where it is able to hydrogen bond with the surrounding water molecules (Sreerama & Woody, 1993). This PPII type secondary structure would be partially adopted by AGP and enzyme-treated AGPs. It is well known that though the conformational freedom of PPII can be partially limited when it includes proline due to its pyrrolidine ring, the extended PPII helix is markedly more flexible in comparison with the α -helix and β -sheet (Adzhubei, Sternberg,

& Makarov, 2013). CD secondary structure analyses of AGP and enzyme-treated AGPs, using the self-consistent method (Sreerama & Woody, 1993) in Dichroweb (Lobley et al., 2002), predicted similar secondary structures content: 11% α -helix, 25 β -sheet, 22% β -turn, 26% PPII helix and 16% unordered structure. The presence of unordered and PPII secondary structures is the basic characteristic of natively unfolded proteins (Uversky, 2013a, 2013b), highlighting the extended and flexible conformation of the polypeptide backbone in AGPs. Theoretical secondary structure predictions made on the 19-residues sequence identified by Goodrum et al. (2000) in the AGP consensus glycopeptide, using algorithms such as i-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>), FoldIndex, DIS-EMBL, GLOBPLOT, SPRITZ, PONDR-FIT, DISPRO-Scratch protein predictor or PSIPRED (<http://expasy.org/>), gave however 100% unordered structure. The result was unchanged whatever the length of the consensus glycopeptide (secondary structures were also predicted for the consensus sequence repeated several times in tandem). In addition, PSIPRED algorithm found a high flexibility in the consensus sequence whatever its length. From these theoretical predictions and those found experimentally, it clearly appeared that the distribution of arabinosides and polysaccharide substituents along the polypeptide chain lead, contrary to what was theoretically predicted, to a certain stiffness of the building blocks. Indeed, as suggested by Goodrum et al. (2000), the amino acid sequence of the 19-residues peptide lead to a symmetrical distribution of arabinosides and polysaccharide substituents directed by the palindrome-like arrangement of the Hyp residues in the peptide backbone. The conformational consequence is that the AGP consensus glycopeptide is slightly rigidified by contiguous Hyp blocks. This stiffness found on the consensus glycopeptide would be much more pronounced on the whole AGP macromolecule and would easily justify the lower level of unordered structures we experimentally found on intact and enzyme-cleaved AGPs.

It was however interesting to notice that the cleavage by proteases of the polypeptide backbone in AGP did not affect the secondary structure content. This result would be in accordance with the palindromic sequence identified in AGP (Goodrum et al., 2000), highlighting the fact that the protease cleavage would induce the generation of very similar (poly)peptide fragments. In addition, the similarities found in the secondary structures before and after hydrolyses would mean that glycomodules attached to the native polypeptide and polypeptide fragments would be very similar and would favor again the repetitive and palindromic nature of the polypeptide sequences found in intact and cleaved AGP. Indeed, Goodrum et al. (2000) gave a final length for the polypeptide of 380 amino acids with about 20 peptide repeats and occasional partial repeats to take into account the higher serine content in native AGP. This repetition of the consensus peptide together with the Hyp contiguity hypothesis (Shpak, Leykam, & Kieliszewski, 1999) would lead to similar structures for the glycomodules after enzymatic hydrolyses. The new original results obtained in the present study would mean that these similar (poly)peptide fragments would adopt the same conformation as the uncleaved polypeptide backbone, illustrating a sort of self-ordering in the building of the glycomodules leading to the AGP macromolecular assembly. This interesting result would reinforce the absolute necessity to perform in a near future a thorough analysis by modern tandem mass spectrometry of the exact polypeptide sequence of AGP after deglycosylation in order to confirm or not the existence of repetitive and palindromic sequences suggested in the literature (Goodrum et al., 2000; Kieliszewski, 2001; Kieliszewski & Shpak, 2001). Tandem mass spectrometry would also be very useful for the identification of the presence of one or several polypeptide sequences in the AGP in order to confirm whether AGP is a molecular association of different structural units as suggested by Al-Assaf et al. (2009).

4. Discussion

Arabinogalactan (AG) and arabinogalactan-protein (AGP) fractions from *A. senegal* gum were hydrolysed using several proteases differing in their mode of action to cleave polypeptide backbones. We aimed to better characterize the structure of the complex AG and AGP fractions by deeply studying the secondary and tertiary structures of the native and enzymatically cleaved AG and AGP. Endoproteinase Glu-C, pepsin and phosphatase acid were thus used in acidic conditions while subtilisin A, pronase, trypsin, papain and proteinase K were used in alkaline conditions to cleave the two fractions. Enzymes did not affect AG fraction structure whatever the pH conditions used, highlighting the inaccessibility of the peptide backbone and the remarkable stability of this fraction in acidic and alkaline conditions. This result was in agreement with previous published data (Randall et al., 1989) and with the thin oblate ellipsoid model we previously identified for the AG fraction where the 43 amino-acid residues peptide sequence was supposed, based on spectroscopic methods, to be totally buried (Renard et al., 2006; Sanchez et al., 2008). The enzyme-treated AG reported here in this study was finally not conclusive in terms of new structural information and in particular in the understanding of the glycosylation arrangements along the peptide backbone due to the absence of cleavage of the short peptide.

Contrary to the remarkable stability of AG-peptide against hydrolysis, even in alkaline conditions, numerous authors clearly demonstrated the efficiency of pronase in the complete hydrolysis of the protein in the AGP component with a ten-fold decrease in the weight average molecular mass of the macromolecule (Al-Assaf et al., 2009; Aoki et al., 2007; Chikamai et al., 1996; Connolly et al., 1988; Duvallet et al., 1989; Elmanan et al., 2008; Flindt et al., 2005; Mahendran et al., 2008; Osman et al., 1993; Randall et al., 1989). Recently, it was also demonstrated that papain-treated Acacia gum also showed complete disappearance of the AGP fraction with a 40% decrease of its original weight average molecular mass (Abdalla et al., 2012). However, a reduction in M_w was also noticed by the authors in absence of enzyme due to the temperature applied during hydrolysis casting some doubts about the real efficiency of papain to really degrade the protein component of Acacia gum. In addition, a too much high substrate:enzyme ratio could be at the origin of the lack of enzyme efficiency.

Our present results, using a substrate:enzyme ratio of 25 in alkaline conditions, indicated that the protein moiety could be efficiently degraded whatever the enzyme used including papain. The lowest M_w was therefore reached in treated AGP-papain with a M_w value of $1.68 \times 10^5 \text{ g mol}^{-1}$, a M_w/M_n polydispersity value of 1.46 and a radius of gyration value of 11 nm. The M_w/M_n polydispersity values were always slightly higher for enzyme-treated AGPs compare to control AGP highlighting the broader distribution of AGPs fragments after hydrolysis due presumably to the differences in efficiency toward hydrolysis of the proteases used in these conditions (see Table 1) (Renard et al., 2012). The differences in enzymatic efficiency could also be related to local steric constraints giving some enhanced accessibility of cleavage sites to enzymes and consequently increasing their efficiency. However, a single conformation in solution close to the AG unit was identified in enzyme-treated AGPs by HPSEC-MALLS (see Fig. 2). The overall structure of the enzyme-treated AGPs was found to be surprisingly quite similar whatever the enzyme used and close, with however some subtle differences, to AG unit (see Fig. 4). A tri-axial ellipsoid conformation with semi-axes values slightly differing from one enzyme to another was found in enzyme-treated AGPs (see Table 2). More interestingly, the two main preferential distances identified in the pair distance distribution function (see Fig. 5) of control and enzyme-treated AGPs would claim in favor of rod-like or elongated particles or alternatively would indicate the

presence of two particles differing in dimensions. These preliminary structural data would therefore reinforce the hypothesis of the molecular associations of structural units in Acacia gum through covalent linkages and/or hydrophobic associations and hydrogen bonding leading to the formation of the AGP macromolecular assembly. Another hypothesis would be that AGP macromolecular assembly, naturally present in exudate, would be intrinsically hydrolysed during the maturation of nodules by exo/endogeneous enzymes, giving rise to molecular fractions with lower M_w , as AG fraction for instance. These structural units would thus be revealed due to their dissociation after polypeptide cleavage. As AG was previously shown to be unaffected by enzyme treatment in alkaline conditions, we may easily conclude that the hydrolysis only concerned the polypeptide sequence of other structural units. A rough calculation based on previous M_w determinations (Renard et al., 2006), and assuming that AGP would be a molecular association of AG and GP units (Al-Assaf et al., 2009), would give one AG ($M_w = 2.86 \times 10^5 \text{ g mol}^{-1}$) and one polydisperse GP units (average $M_w = 1.59 \times 10^6 \text{ g mol}^{-1}$) or one AG and three GP₃ units (third population identified by HPSEC-MALLS, $M_w = 2.95 \times 10^5 \text{ g mol}^{-1}$) to form AGP macromolecular assembly. However, the present study revealed that structural units close to AG units but however with some structural differences would be involved in the formation of AGP macromolecular assembly. In addition, the very large polydispersity found in GP fraction (Renard et al., 2006) would cast some doubt about its possible involvement in the formation of AGP macromolecular assembly in particular because several populations with M_w higher than intact AGP were identified. In addition, the different results found in literature regarding the structure of AGP from Acacia gum would suggest in a near future to give a better term to define this molecular fraction, AGP being often ascribed to a unique macromolecule (Al-Assaf et al., 2009; Goodrum et al., 2000; Qi, Fong, & Lampert, 1991; Renard et al., 2006, 2012, 2013).

The molecular structure of control and enzyme-treated AGPs surprisingly predicted similar secondary structures content with 11% α -helix, 25% β -sheet, 22% β -turn, 26% PPII helix and 16% unordered structure. These findings would indicate that the enzyme cleavage would induce the formation of similar (poly)peptide fragments identified previously in AGP after hydrogen fluoride (HF)-deglycosylation as being formed of 19-residues repetitive and palindromic sequences (Goodrum et al., 2000). In addition, the similar conformations adopted by control and enzyme-cleaved AGPs would be in favor of a high regularity in the amino acid residues sequence of the polypeptide backbone in accordance with the palindromic nature of the peptide sequence and the overall symmetry of the consensus glycopeptide found by Goodrum et al. (2000). Indeed, authors found that palindromic symmetry rigidified by contiguous hydroxyproline building blocks may impart self-ordering properties in AGP. These results would question however about the type of oligo- and polysaccharide branching occurring on the AGP polypeptide backbone. This question motivated some simple calculations to determine the theoretical molecular weight of AGP taking into account the type of branching on hydroxyproline (Hyp) residues identified by Qi et al. (1991) and those identified by Goodrum et al. (2000) extended to serine (Ser) and threonine (Thr) residues as demonstrated by Mahendran et al. (2008). The theoretical molecular weights, taking into account our previous result of the composition and the total number of amino acid residues present in AGP (Renard et al., 2006), were of $1.87 \times 10^6 \text{ g mol}^{-1}$ on a Qi et al. (1991) basis for the type of branching and of $1.61 \times 10^6 \text{ g mol}^{-1}$ on a Goodrum et al. (2000) basis for the type of branching (Table 3). These calculations simply highlighted the fact that Hyp, Thr and Ser residues would be the only three amino acid residues involved in the glycosylation motifs that direct arabinoside oligomers and carbohydrate blocks addition to the protein backbone of AGP. The existence of

repetitive and palindromic sequences in the protein moiety of AGP would therefore be confirmed through these simple calculations as only 4.7% difference in the theoretical and experimental M_w was found taking into consideration the type of branching identified by Qi et al. (1991). In addition, the calculations performed to determine the M_w of intact AGP confirmed the Hyp contiguity hypothesis which predicts arabinosides on contiguous Hyp residues and arabinogalactan polysaccharides on clustered non-contiguous Hyp residues. These calculations combined with the structural similarities found after enzyme-treated AGP would reinforce the self-ordering arrangements of the building blocks along the polypeptide sequence due to a certain self-similar arrangement during assembly of the elementary building blocks (i.e. the consensus glycopeptide found by Goodrum et al. (2000), that we can defined as the elementary building block, can be repeated n times, on a self-similar basis, to form whole AGP, giving rise to a structural self-similarity for the AGP macromolecular assembly whatever the length or observation scale). Self-similar behavior was therefore confirmed by SANS at a length scale ranging from ~ 1 to 20 nm, length scale in accordance with some self-similarity arrangement of elementary building blocks. Beyond the self-similarity or self-ordering arising from the association of elementary building blocks, which we will refer later to build a 3D model of AGP macromolecular assembly, we may wonder how many elementary building blocks constitute AGP macromolecular assembly. To answer this question, we performed calculations to determine the number of short and long polypeptide chains involved in AGP macromolecular assembly. We made the assumption that two types of building blocks would be present in AGP to perform these calculations. This hypothesis came from the identification by Mahendran et al. (2008) of two polypeptide chains by SDS-PAGE electrophoresis after HF-deglycosylation of Acacia gum, one short peptide chain and one long polypeptide chain with 45 and 250 amino acid residues, respectively. The long polypeptide chain, with M_w of $2.9\text{--}3 \times 10^4 \text{ g mol}^{-1}$, corresponding to $\sim 260\text{--}270$ residues, was also identified in AGP by Osman et al. (1993). The calculations were thus performed taking into account these two polypeptide chains and the type of branching on Hyp residues determined by Qi et al. (1991) extended to branching on Ser and Thr residues as demonstrated by Mahendran et al. (2008). Finally, two equations with two unknown parameters, x the number of short peptide chains and y the number of long polypeptide chains, were resolved and gave $x=6$ and $y=8$ (Table 4). AGP macromolecular assembly would thus be made of the association of 6 glycomodules with $M_w = 3.88 \times 10^4 \text{ g mol}^{-1}$ and 8 glycomodules with $M_w = 2.07 \times 10^5 \text{ g mol}^{-1}$, each glycomodule (i.e. building block) being made of arabinoside oligomers and polysaccharide blocks attached to Hyp, Ser and Thr residues of the short and long (poly)peptide chain. It was therefore interesting to notice that the molecular weight calculated for the building block 1 was close to those found experimentally after chemical hydrolysis using sodium borohydride/sodium hydroxide treatments by Mahendran et al. (2008). A value of $4.5 \times 10^4 \text{ g mol}^{-1}$ was found for the molecular weight of the building block after complete AGP degradation. To go further insight into the structural elucidation of AGP macromolecular assembly, we wanted to know how much arabinoside oligomers and polysaccharide blocks were present on each (poly)peptide chain.

The number of branches (arabinoside oligomers and carbohydrate blocks) on AGP macromolecular assembly was directly given by the number of amino acid residues involved in (poly)saccharide substitutions (Table 4). 23 and 119 Hyp, Thr and Ser amino acid residues would be involved in the short and long (poly)peptide chain, respectively. Considering the percentages of substitutions (arabinoside oligomers and carbohydrate blocks) given by Qi et al. (1991) on Hyp residues, and extended to Ser and Thr residues, the number of branches for each building block and the total

Table 3
Calculations of arabinogalactan-protein (AGP) molecular weight M_w and comparison with experimental M_w determined by HPSEC-MALLS. Calculations were based on type of branching on hydroxyproline (Hyp) residues from Qi et al. (1991) and Goodrum et al. (2000) extended to branching on serine (Ser) and threonine (Thr) residues as demonstrated by Mahendran et al. (2008). The number of Hyp, Ser and Thr amino acid residues were taken from Renard et al. (2006).

Number f residues involved in oligo- and polysaccharides branching	Type of branching on amino acid residue (from Qi et al., 1991)	%Glycosylation	%Glycosylation × number of residues	Aminoacid M_w involved in branching	Carbohydrate M_w involved in glycosylation	Total M_w	Type of branching on aminoacid residue (from Goodrum et al., 2000)	%Glycosylation	%Glycosylation × number of residues	Aminoacid M_w involved in branching	Carbohydrate M_w involved in glycosylation	Total M_w
Number of Hyp												
	Non-glycosylated	0.12	92	110	0	1.01E+04	Non-glycosylated	0.11	84	110	0	9.24E+03
766	Hyp-Ara ₃	0.64	490	110	440	2.70E+05	Hyp-Ara ₄	0.05	38	110	590	2.66E+04
	Hyp-polysaccharide	0.24	184	110	4400	8.30E+05	Hyp-Ara ₃	0.27	207	110	440	1.14E+05
							Hyp-Ara ₂	0.27	207	110	290	8.28E+04
							Hyp-Ara	0.1	77	110	146	1.97E+04
							Hyp-polysaccharide	0.2	153	110	4400	6.90E+05
Number of Ser												
	Non-glycosylated	0.12	33	110	0	3.63E+03	Non-glycosylated	0.11	30	110	0	3.30E+03
275	Ser-Ara ₃	0.64	176	110	440	9.68E+04	Ser-Ara ₄	0.05	14	110	590	9.80E+03
	Ser-polysaccharide	0.24	66	110	4400	2.98E+05	Ser-Ara ₃	0.27	74	110	440	4.07E+04
							Ser-Ara ₂	0.27	74	110	290	2.96E+04
							Ser-Ara	0.1	28	110	146	7.17E+03
							Ser-polysaccharide	0.2	55	110	4400	2.48E+05
Number of Thr												
	Non-glycosylated	0.12	21	110	0	2.31E+03	Non-glycosylated	0.11	19	110	0	2.09E+03
176	Thr-Ara ₃	0.64	113	110	440	6.22E+04	Thr-Ara ₄	0.05	9	110	590	6.30E+03
	Thr-polysaccharide	0.24	42	110	4400	1.89E+05	Thr-Ara ₃	0.27	47	110	440	2.59E+04
							Thr-Ara ₂	0.27	48	110	290	1.92E+04
							Thr-Ara	0.1	18	110	146	4.61E+03
							Thr-polysaccharide	0.2	35	110	4400	1.58E+05
Sum of aminoacid-oligo(poly)saccharides branching						1.75E+06						1.48E+06
Sum of Hyp, Ser, Thr residues not involved in saccharide branching						1.61E+04						1.46E+04
Sum of additional non-glycosylated aminoacid residues						1.14E+05						1.14E+05
M_w (calculated)						1.88E+06						1.61E+06
M_w (from HPSEC-MALLS)						1.79E+06						1.79E+06
Error%						4.77						11.13

Table 4
Calculations of the number of building blocks 1 and 2 involved in the formation of arabinogalactan-protein (AGP) macromolecular assembly taking into account the type of branching on hydroxyproline (Hyp) residues determined by Qi et al. (1991) extended to branching on serine (Ser) and threonine (Thr) residues as demonstrated by Mahendran et al. (2008). Building blocks 1 and 2 were composed of 45 and 250 amino acid residues, respectively. The number of Hyp, Ser and Thr amino acid residues were taken from Renard et al. (2006).

Building block 1 (with the short chain made of 45 amino acid residues)						
Number of residues involved in oligo-and polysaccharides branching	Type of branching on amino acid residue (from Qi et al., 1991)	%Glycosylation	%Glycosylation × number of residues	Amino acid M_w involved in branching	Carbohydrate M_w involved in glycosylation	Total M_w
Number of Hyp						
15	Non-glycosylated	0.12	1	110	0	1.10E+02
	Hyp-Ara3	0.64	10	110	440	5.50E+03
	Hyp-polysaccharide	0.24	4	110	4400	1.80E+04
Number of Ser						
6	Non-glycosylated	0.12	1	110	0	1.10E+02
	Ser-Ara ₃	0.64	4	110	440	2.20E+03
	Ser-polysaccharide	0.24	1	110	4400	4.51E+03
Number of Thr						
4	Non-glycosylated	0.12	0	110	0	0.00E+00
	Thr-Ara ₃	0.64	3	110	440	1.65E+03
	Thr-polysaccharide	0.24	1	110	4400	4.51E+03
Sum of amino acid-oligo(poly)saccharides branching						3.64E+04
Sum of Hyp, Ser, Thr residues not involved in saccharide branching						2.20E+02
Sum of non-glycosylated additional amino acid residues						2.20E+03
Building block 1 M_w (calculated)						3.88E+04
Total AGP M_w						1.88E+06
Total number of residues						2.25E+03
Number of building blocks 1 ^a						6
Building block 2 (with the long chain made of 250 amino acid residues)						
Number of residues involved in oligo-and polysaccharides branching	Type of branching on amino acid residue (from Qi et al., 1991)	%Glycosylation	%Glycosylation × number of residues	Amino acid M_w involved in branching	Carbohydrate M_w involved in glycosylation	Total M_w
Number of Hyp						
85	Non-glycosylated	0.12	11	110	0	1.21E+03
	Hyp-Ara3	0.64	54	110	440	2.97E+04
	Hyp-polysaccharide	0.24	20	110	4400	9.02E+04
Number of Ser						
31	Non-glycosylated	0.12	4	110	0	4.40E+02
	Ser-Ara ₃	0.64	20	110	440	1.10E+04
	Ser-polysaccharide	0.24	7	110	4400	3.16E+04
Number of Thr						
20	Non-glycosylated	0.12	2	110	0	2.20E+02
	Thr-Ara ₃	0.64	13	110	440	7.15E+03
	Thr-polysaccharide	0.24	5	110	4400	2.26E+04
Sum of amino acid-oligo(poly)saccharides branching						1.92E+05
Sum of Hyp, Ser, Thr residues not involved in saccharide branching						1.87E+03
Sum of non-glycosylated additional amino acid residues						1.25E+04
Building block 2 M_w (calculated)						2.07E+05
Total AGP M_w						1.88E+06
Total number of residues						2.25E+03
Number of building blocks 2 ^a						8

^a Total AGP M_w (calculated) = $x(M_w \text{ (building block 1)}) + y(M_w \text{ (building block 2)})$ and total number of residues = $x(45) + y(250)$ with x = number of building blocks 1 and y = number of building blocks 2.

number of building blocks 1 and 2 constituting AGP macromolecular assembly, were finally determined (see details in Table 5). Six building blocks 1, with a $M_w = 38\,880 \text{ g mol}^{-1}$, and 8 building blocks 2, with a $M_w = 207\,000 \text{ g mol}^{-1}$, would contribute to the assembly of

the arabinogalactan-protein (AGP) from Acacia gum. The molecular associations of these building blocks would be driven by covalent and/or non-covalent interactions through the interactions occurring between amino acid residues to link neighboring glycomodules

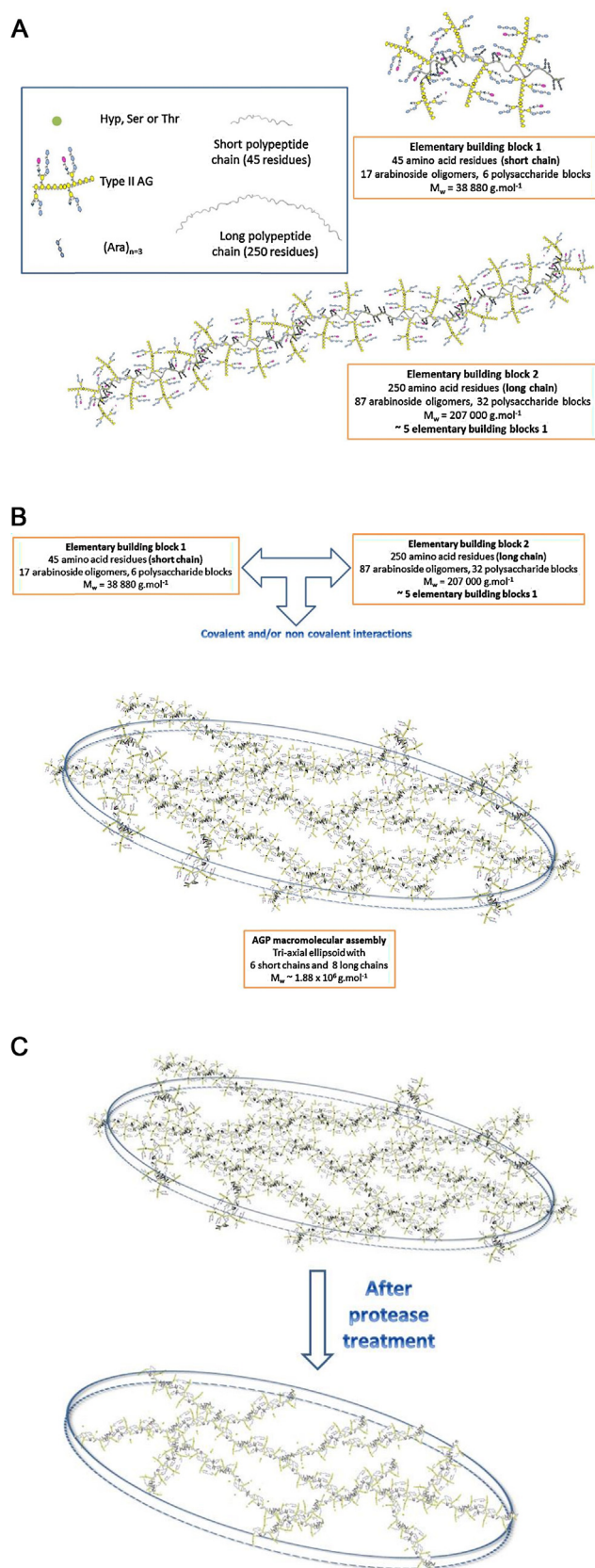


Fig. 7. (A) Elementary building blocks 1 and 2 used to build arabinogalactan-protein (AGP) macromolecular assembly from *Acacia Senegal* gum. Building block 2 was built by a linear arrangement of five building blocks 1. (B) Three-dimensional model of AGP made of the assembly of 6 building blocks 1 and 8 building blocks 2 by covalent and/or non-covalent interactions. The elementary building blocks were constructed taking into consideration the quasi-palindromic nature of polypeptide backbones

Table 5

Calculations of the number of (poly)saccharide substituents (arabinoside oligomers and carbohydrate blocks) in building blocks 1 and 2, attached to Hyp, Ser and Thr residues of the short and long polypeptide chains, involved in the formation of arabinogalactan-protein (AGP) macromolecular assembly.

Elementary building blocks involved in AGP macromolecular assembly	Number	M_w
Building block 1 (23 amino acid residues for (poly)saccharide substitutions) ^a		
-arabinoside oligomers	17	
-carbohydrate blocks	6	
Total M_w (short peptide chain + (poly)saccharide substituents) (g mol^{-1})		38 880
Building block 2 (119 amino acid residues for (poly)saccharide substitutions) ^a		
-arabinoside oligomers	87	
-carbohydrate blocks	32	
Total M_w (long polypeptide chain + (poly)saccharide substituents) (g mol^{-1})		207 000
Number of building blocks 1 in AGP ^b	6	
Number of building blocks 2 in AGP ^b	8	
Calculated AGP M_w (g mol^{-1})		1.88×10^6
Experimental AGP M_w from HPSEC-MALLS (g mol^{-1})		1.79×10^6

^a See Table 4 for the details of (poly)saccharide substitutions on Hyp, Thr and Ser amino acid residues.

^b Number of building blocks 1 (or building blocks 2) = $(M_w \text{ arabinoside oligomers}/440) + (M_w \text{ carbohydrate blocks}/4400)$ with M_w arabinoside oligomers in building block 1 (or 2) = $0.73 \times M_w^{\text{sugar}}$ and M_w carbohydrate blocks in building block 1 (or 2) = $0.27 \times M_w^{\text{sugar}}$. M_w^{sugar} (building blocks 1 and 2) were calculated from data available in Table 4 with M_w^{sugar} (building block 1) = $33\,880 \text{ g mol}^{-1}$ and M_w^{sugar} (building block 2) = $179\,080 \text{ g mol}^{-1}$. The values 0.73 and 0.27 corresponded to the percentages of glycosylation determined by Qi et al. (1991) but extended to 100% as sugars moieties attached to Hyp, Ser and Thr residues were only considered.

or through the numerous hydrogen bonds occurring between polypeptide and/or β -1,3 galactan backbones. A three-dimensional model of AGP macromolecular assembly was built based on the repetition of the peptide consensus sequence found by Goodrum et al. (2000) to get the short and long polypeptide chains and the self-ordering process occurring from these quasi-palindromic polypeptide sequences to attach carbohydrate moieties (Fig. 7). The elementary building block 2 was therefore built from the repetition of ~5 elementary building blocks 1 linearly linked to maintain the palindromic nature of the polypeptide backbone. The self-similar approach was then used to build the 3D model of AGP macromolecular assembly taking into consideration the self-similarity of the spatial organization of building blocks 2 but also the spatial confinement of building blocks 1 and 2 to fill a volume close to the tri-axial ellipsoid morphology we identified by SANS. However, the statement that a given biological structure is self-similar should imply that the iterative process of its construction has a real biological meaning, i.e. that its construction in nature is achieved by means of a single genetic, enzymatic, or biophysical mechanism successively repeated. Such an iterative process to reproduce a self-similar object was successfully applied to glycogen, with a theoretical fractal dimension value of 2 (Melendez, Melendez-Hevia, & Canela, 1999). The experimental d_f value of 2.2 probed by SANS for AGP macromolecular assembly was in agreement with the d_f values of 2–2.5 found for homogeneously branched

and as a consequence the overall symmetry, or self-ordering, of carbohydrate moieties and arabinoside oligomers attached to it. The self-similarity approach was then used to build the AGP macromolecular assembly from the elementary building blocks 1 and 2. (C) Three-dimensional model of papain-treated AGP.

polymers, where the low values refer to indifferent (Θ) solvents, the higher ones to thermodynamically good solvents (Burchard et al., 2012). Experimental d_f values of 2.23–2.47 were therefore found for hyperbranched polymers by small angle scattering (De Luca, Richards, Grillo, & King, 2003; Garamus et al., 2004; Gelade et al., 2001). The self-similar structure identified by SANS in AGP macromolecular assembly, property used to build the 3D model, would also explain why great similarities were found in AGP conformation and secondary structures after enzymatic degradation.

One could also argue that a star-like arrangement of the five elementary building blocks 2 could also be possible to build AGP macromolecular assembly. However, star-like polymer conformations, often at the origin of dendrimer conformation, lead to fractal dimension close to 1.7, value we did not find experimentally whatever the observation scale. Our model was therefore built taking into account the self-ordering process of carbohydrate moieties due to the palindromic arrangement of peptide sequence for each building block together with the self-similarity of these building blocks during their association leading to a macromolecular assembly at larger scale called AGP. A better definition of this class of macromolecules should be brought when identification of the polypeptide sequence(s) will be performed.

Unfortunately, we were unable to perform the same calculations and build 3D models on enzyme-treated AGPs to know their fine structure, in particular because the number of residual amino acid residues present in the different fragments after hydrolysis was unknown. The knowledge of only the molecular weight of enzyme-treated AGPs did not allow the calculation of the number of (cleaved) building blocks, several combinations of building blocks 1 and 2 being possible without the knowledge of the number of amino acid residues present in enzyme-treated AGPs. A hypothetical and very simple three-dimensional model of papain-treated AGP was built according to molecular dimensions and molecular weight probed by SANS and HPSEC-MALLS (Fig. 7C).

However, from the simple calculations we performed on intact AGP, and assuming that AGP macromolecular assembly would be a combination of two similar structural units, it would be very interesting to perform proteomic analyses on enzyme-treated AGPs after chemical deglycosylation in order to probe the structure and the sequence of the protein moieties and to confirm or not the presence of two distinct protein chains in AGP. The complementary analyses would be to degrade enzymatically the oligo- and carbohydrate blocks in order to study both glycosylation motifs and protein sequences after separation of each components using liquid chromatography–tandem mass spectrometry (LC–MS–MS).

5. Conclusion

AGP and AG molecular fractions were degraded enzymatically using acidic and alkaline proteases in order to probe the conformation and structure of the two main fractions of *A. senegal* gum. While AG unit kept intact whatever the enzymes and conditions used, AGP was found to be hydrolysed only in alkaline conditions. The decrease in molecular weight of AGP after enzymatic treatment confirmed the accessibility of enzymes toward polypeptide cleavages and papain was found to be the most efficient protease with a decrease of M_w from 1.79×10^6 to 1.68×10^5 g mol⁻¹. Results from HPSEC-MALLS and small angle neutron scattering measurements would claim in favor of rod-like or elongated particles or alternatively would indicate the presence of two particles differing in dimensions before and after enzymatic treatment. These preliminary structural data would therefore reinforce the hypothesis of the molecular associations of structural units in Acacia gum through covalent linkages and/or hydrophobic associations and hydrogen bonding leading to the formation of the AGP

macromolecular assembly. Moreover, the molecular structure of control and enzyme-treated AGPs surprisingly predicted similar secondary structures content. The similar conformations adopted by control and enzyme-cleaved AGPs probed at the molecular and mesoscopic scale would be in favor of a high flexibility of the polypeptidic backbone before and after enzymatic treatment in accordance with the repetitive and palindromic nature of peptide sequence and the overall symmetry of the carbohydrate moieties along the protein backbone. It was finally suggested that a self-similarity driven-process would be at the origin of the assembly of AGP from a consensus glycopeptide building block with a symmetrical distribution of arabinosides and polysaccharide substituents.

Acknowledgments

We would like to acknowledge M.-J. Crepeau and M.-C. Ralet for HPSEC-MALLS measurements and data treatments. Special thanks to M.-C. Ralet for the fruitful advices concerning the experimental part of enzymatic hydrolyses on AGP and A. Giuliani (Synchrotron radiation source Soleil, France) for the collagen CD spectrum.

References

- Abdalla, I. G., Ahmed, B.-E. I., Ishag, C. Y., Al-Assaf, S., Andres-Brull, M., Phillips, G. O., et al. (2012). Processing of gum arabic enzymatically. In J. F. Kennedy, P. A. Williams, & G. O. Phillips (Eds.), *Gum arabic* (pp. 145–152). London: RSC Publishing.
- Adzhubei, A. A., Sternberg, M. J. E., & Makarov, A. A. (2013). Polyproline-II helix in proteins: Structure and function. *Journal of Molecular Biology*, 425(12), 2100–2132.
- Al-Assaf, S., Sakata, M., McKenna, C., Aoki, H., & Phillips, G. O. (2009). Molecular associations in Acacia gums. *Structural Chemistry*, 20(2), 325–336.
- Anderson, D. M. W., Hirst, E. L., & Stoddart, J. F. (1966). Studies on uronic acid materials – Part XVII: Some structural features of *Acacia senegal* gum (gum arabic). *Journal of Chemical Society C*, 308, 1959–1966.
- Aoki, H., Al-Assaf, S., Katayama, T., & Phillips, G. O. (2007). Characterization and properties of *Acacia senegal* (L.) Willd. var. *senegal* with enhanced properties (*Acacia* (sen) SUPER GUM (TM)): Part 2 – Mechanism of the maturation process. *Food Hydrocolloids*, 21(3), 329–337.
- Bacic, A., Churms, S. C., Stephen, A. M., Cohen, P. B., & Fincher, G. B. (1987). Fine-structure of the arabinogalactan protein from *Lolium-multiflorum*. *Carbohydrate Research*, 162(1), 85–93.
- Beltran, O., de Pinto, G. L., Martinez, M., Picton, L., Cozic, C., Le Cerf, D., et al. (2005). Fractionation and characterization of gum from *Acacia tortuosa*. Effect of enzymatic and alkaline treatments. *Carbohydrate Polymers*, 62(3), 239–244.
- Burchard, W., Khalyavina, A., Lindner, P., Schweins, R., Friedel, P., Wiemann, M., et al. (2012). SANS investigation of global and segmental structures of hyperbranched aliphatic-aromatic polyesters. *Macromolecules*, 45(7), 3177–3187.
- Chikamai, B. N., Banks, W. B., Anderson, D. M. W., & Weiping, W. (1996). Processing of gum arabic and some new opportunities. *Food Hydrocolloids*, 10(3), 309–316.
- Churms, S. C., Merrifield, E. H., & Stephen, A. M. (1983). Some new aspects of the molecular-structure of *Acacia senegal* gum (gum arabic). *Carbohydrate Research*, 123(2), 267–279.
- Clarke, A., Gleeson, P., Harrison, S., & Knox, R. B. (1979). Pollen–stigma interactions – Identification and characterization of surface components with recognition potential. *Proceedings of the National Academy of Sciences of the United States of America*, 76(7), 3358–3362.
- Connolly, S., Fenyo, J. C., & Vandeveld, M. C. (1987a). The effect of pronase on the amino-acid composition of gum arabic. *Comptes Rendus des Seances de la Societe de Biologie et de Ses Filiales*, 181(6), 683–687.
- Connolly, S., Fenyo, J. C., & Vandeveld, M. C. (1987b). Heterogeneity and homogeneity of an arabinogalactan-protein: *Acacia senegal* gum. *Food Hydrocolloids*, 1(5–6), 477–480.
- Connolly, S., Fenyo, J. C., & Vandeveld, M. C. (1988). Effect of a proteinase on the macromolecular distribution of *Acacia senegal* gum. *Carbohydrate Polymers*, 8(1), 23–32.
- De Luca, E., Richards, R. W., Grillo, I., & King, S. M. (2003). Molecular characterization of a hyperbranched polyester. II. Small-angle neutron scattering. *Journal of Polymer Science Part B: Polymer Physics*, 41(12), 1352–1361.
- Duvallet, S., Fenyo, J. C., & Vandeveld, M. C. (1989). Meaning of molecular-weight measurements of gum arabic. *Polymer Bulletin*, 21(5), 517–521.
- Ellis, M., Egelund, J., Schultz, C. J., & Bacic, A. (2010). Arabinogalactan-proteins: Key regulators at the cell surface? *Plant Physiology*, 153(2), 403–419.
- Elmanan, M., Al-Assaf, S., Phillips, G. O., & Williams, P. A. (2008). Studies on *Acacia* exudate gums: Part VI. Interfacial rheology of *Acacia senegal* and *Acacia seyal*. *Food Hydrocolloids*, 22(4), 682–689.
- Fincher, G. B., Stone, B. A., & Clarke, A. E. (1983). Arabinogalactan-proteins – Structure, biosynthesis, and function. *Annual Review of Plant Physiology and Plant Molecular Biology*, 34, 47–70.

- Flindt, C., Al-Assaf, S., Phillips, G. O., & Williams, P. A. (2005). *Studies on Acacia exudate gums: Part V. Structural features of Acacia seyal*. *Food Hydrocolloids*, 19(4), 687–701.
- Garamus, V. M., Maksimova, T. V., Kautz, H., Barriau, E., Frey, H., Schlotterbeck, U., et al. (2004). Hyperbranched polymers: Structure of hyperbranched polyglycerol and amphiphilic poly(glycerol ester)s in dilute aqueous and nonaqueous solution. *Macromolecules*, 37(22), 8394–8399.
- Gelade, E. T. F., Goderis, B., de Koster, C. G., Meijerink, N., van Benthem, R., Fokkens, R., et al. (2001). Molecular structure characterization of hyperbranched polyesteramides. *Macromolecules*, 34(11), 3552–3558.
- Glatter, O. (1979). Interpretation of real-space information from small-angle scattering experiments. *Journal of Applied Crystallography*, 12(April), 166–175.
- Glatter, O. (2002). In P. Lindner, & T. Zemb (Eds.), *Neutrons, X-rays and light* (pp. 73–102). Amsterdam, The Netherlands: Elsevier Science B.V.
- Goodrum, L. J., Patel, A., Leykam, J. F., & Kieliszewski, M. J. (2000). Gum arabic glycoprotein contains glycomodules of both extensin and arabinogalactan-glycoproteins. *Phytochemistry*, 54(1), 99–106.
- Kieliszewski, M. J. (2001). The latest hype on hyp-O-glycosylation codes. *Phytochemistry*, 57(3), 319–323.
- Kieliszewski, M. J., & Lampion, D. T. A. (1994). Extensin – Repetitive motifs, functional sites, posttranslational codes, and phylogeny. *Plant Journal*, 5(2), 157–172.
- Kieliszewski, M. J., & Shpak, E. (2001). Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. *Cellular and Molecular Life Sciences*, 58(10), 1386–1398.
- Liu, C. G., & Mehdy, M. C. (2007). A nonclassical arabinogalactan protein gene highly expressed in vascular tissues AGP31, is transcriptionally repressed by methyl jasmonic acid in Arabidopsis. *Plant Physiology*, 145(3), 863–874.
- Lobley, A., Whitmore, L., & Wallace, B. A. (2002). Dichroweb: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, 18(1), 211–212.
- Mahendran, T., Williams, P. A., Phillips, G. O., Al-Assaf, S., & Baldwin, T. C. (2008). New insights into the structural characteristics of the arabinogalactan – Protein (AGP) fraction of gum arabic. *Journal of Agricultural and Food Chemistry*, 56(19), 9269–9276.
- Melendez, R., Melendez-Havia, E., & Canela, E. I. (1999). The fractal structure of glycogen: A clever solution to optimize cell metabolism. *Biophysical Journal*, 77(3), 1327–1332.
- Osman, M. E., Menzies, A. R., Williams, P. A., Phillips, G. O., & Baldwin, T. C. (1993). The molecular characterization of the polysaccharide gum from *Acacia senegal*. *Carbohydrate Research*, 246, 303–318.
- Qi, W., Fong, C., & Lampion, 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.
- Qu, Y. M., Egelund, J., Gilson, P. R., Houghton, F., Gleeson, P. A., Schultz, C. J., et al. (2008). Identification of a novel group of putative *Arabidopsis thaliana* beta-(1,3)-galactosyltransferases. *Plant Molecular Biology*, 68(1–2), 43–59.
- Randall, R. C., Phillips, G. O., & Williams, P. A. (1989). Fractionation and characterization of gum from *Acacia senegal*. *Food Hydrocolloids*, 3(1), 65–75.
- Renard, D., Garnier, C., Lapp, A., Schmitt, C., & Sanchez, C. (2012). Structure of arabinogalactan-protein from Acacia gum: From porous ellipsoids to supramolecular architectures. *Carbohydrate Polymers*, 90(1), 322–332.
- Renard, D., Garnier, C., Lapp, A., Schmitt, C., & Sanchez, C. (2013). Structure of arabinogalactan-protein from Acacia gum: From porous ellipsoids to supramolecular architectures. *Carbohydrate Polymers*, 97(2), 864–867.
- Renard, D., Lavenant-Gourgeon, L., Ralet, M. C., & Sanchez, C. (2006). *Acacia senegal* gum: Continuum of molecular species differing by their protein to sugar ratio, molecular weight, and charges. *Biomacromolecules*, 7(9), 2637–2649.
- Sanchez, C., Schmitt, C., Kolodziejczyk, E., Lapp, A., Gaillard, C., & Renard, D. (2008). The Acacia gum arabinogalactan fraction is a thin oblate ellipsoid: A new model based on small-angle neutron scattering and ab initio calculation. *Biophysical Journal*, 94(2), 629–639.
- Seifert, G. J., & Roberts, K. (2007). The biology of arabinogalactan proteins. *Annual Review of Plant Biology*, 58, 137–161.
- Shpak, E., Leykam, J. F., & Kieliszewski, M. J. (1999). Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. *Proceedings of the National Academy of Sciences of the United States of America*, 96(26), 14736–14741.
- Sreerama, N., & Woody, R. W. (1993). A self-consistent method for the analysis of protein secondary structure from circular-dichroism. *Analytical Biochemistry*, 209(1), 32–44.
- Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25, 495–503.
- Svergun, D. I., & Koch, M. H. J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, 66(10), 1735–1782.
- Svergun, D. I., Semenyuk, A. V., & Feigin, L. A. (1988). Small-angle-scattering-data treatment by the regularization method. *Acta Crystallographica Section A*, 44, 244–250.
- Tan, L., Qiu, F., Lampion, D. T. A., & Kieliszewski, M. J. (2004). Structure of a hydroxyproline (Hyp)-arabinogalactan polysaccharide from repetitive Ala-Hyp expressed in transgenic *Nicotiana tabacum*. *Journal of Biological Chemistry*, 279(13), 13156–13165.
- Tan, L., Varnai, P., Lampion, D. T. A., Yuan, C. H., Xu, J. F., Qiu, F., et al. (2010). Plant O-hydroxyproline arabinogalactans are composed of repeating trigalactosyl subunits with short bifurcated side chains. *Journal of Biological Chemistry*, 285(32), 24575–24583.
- Tryfona, T., Liang, H. C., Kotake, T., Tsumuraya, Y., Stephens, E., & Dupree, P. (2012). Structural characterization of Arabidopsis leaf arabinogalactan polysaccharides. *Plant Physiology*, 160(2), 653–666.
- Uversky, V. N. (2013a). A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science*, 22(6), 693–724.
- Uversky, V. N. (2013b). Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta: Proteins and Proteomics*, 1834(5), 932–951.
- Wang, J. B., Zuo, X. B., Yu, P., Byeon, I. J. L., Jung, J. W., Wang, X. X., et al. (2009). Determination of multicomponent protein structures in solution using global orientation and shape restraints. *Journal of the American Chemical Society*, 131(30), 10507–10515.
- Whitmore, L., & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32, W668–W673.
- Whitmore, L., & Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers*, 89(5), 392–400.
- Williams, P. A., & Langdon, M. J. (1995). Characterisation of gum arabic by gel permeation chromatography. *Chromatographic Analysis*, 5(October/November), 5–7.
- Williams, P. A., & Phillips, G. O. (2000). Gum arabic. In G. O. Phillips, & P. A. Williams (Eds.), *Handbook of hydrocolloids* (pp. 155–168). Cambridge: Woodhead Publishing Limited.
- Williams, P. A., Phillips, G. O., & Stephen, A. M. (1990). Spectroscopic and molecular comparisons of three fractions from *Acacia senegal* gum. *Food Hydrocolloids*, 4(4), 305–311.