Carbohydrate Polymers 98 (2013) 522–531

Contents lists available at SciVerse ScienceDirect

Carbohydrate Polymers

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

Arabinogalactan protein cluster from *Jatropha curcas* seed embryo contains fasciclin, xylogen and LysM proteins

Maria Sehlbach[a], Simone König[b], Michael Mormann[c], Jandirk Sendker[a], Andreas Hensel[a,*]

[a] University of Münster, Institute for Pharmaceutical Biology and Phytochemistry (IPBP), Hittorfstraße 56, D-48149 Münster, Germany
[b] University of Münster, Interdisciplinary Centre for Clinical Research, Röntgenstr. 21, D-48149 Münster, Germany
[c] University of Münster, Institute for Hygiene, Robert-Koch-Strasse 41, D-48149 Münster, Germany

ARTICLE INFO

Article history:
Received 11 March 2013
Received in revised form 3 June 2013
Accepted 4 June 2013
Available online 24 June 2013

Keywords:
Jatropha curcas L.
Arabinogalactan proteins
Fasciclin
Xylogen
LysM
N-/*O*-glycosylation

ABSTRACT

An non-GPI-anchored AGP cluster (Y2) was isolated from the seeds of *Jatropha curcas* L. (Euphorbiaceae) composed of 4.8% polypeptides (mainly Ala, Ser, Gly, Hyp, Glu) and a carbohydrate moiety composed of Gal, Ara, GlcA, Rha, Man and GlcN. Besides the typical structural features of arabinogalactan proteins, typical *N*-glycan linker of the complex type ($GlcNAc_4Man_3Gal_2Fuc_1Xyl_1$) were identified. *O*-glycosylation occurred mainly *via* Hyp and to a lesser extent *via* Thr and Ser. N-glycans from the complex type, carrying at the innermost GlcNAc at position O-3 one α-Fuc-residue, were also present.

MS analysis of the tryptic digest assigned peptides of three major protein groups: fasciclin-like arabinogalactan proteins, xylogen-like proteins and LysM domain-containing proteins. They could not be separated further and it is indicated that various homologous protein forms co-exist. Histological investigation of *J. curcas* seeds revealed the presence of AGPs in the vessels of cotyledons and in the procambium ring of the embryo.

1. Introduction

Jatropha curcas L. (Euphorbiaceae) is a deciduous and drought resistant tree, widely distributed in tropical and subtropical areas of Central and South America, Africa, India and Southeast Asia. Plants have been used for ecosystem restoration of disturbed areas and biodiesel production (Fairless, 2007). The seeds contain high amounts of fatty oil, a lectin (Lin, Zhou, Wang, Jiang, & Tang, 2010) and diterpens from the tigliane-type in form of phorbol esters with tumour-promoting and inflammative activity (Goel, Makkar, Francis, & Becker, 2007). The seeds of *J. curcas* are used traditionally for wound-healing, fractures and burns. Recently a crude arabinogalactan protein (AGP) fraction from *J. curcas* seeds has been described to influence human skin cell physiology with the stimulation of mitochondrial activity of keratinocytes and dermal fibroblasts and significant increase of the ATP status of primary keratinocytes (Zippel, Wells, & Hensel, 2010). Additionally, the fraction induced keratinocyte differentiation by stimulation of growth factors GM-CSF, HGF, KGF and TGFβ. This *in vitro* activity profile pointed to a potent induction of cellular differentiation *via* stimulation of growth hormones and TGFβ-induced cell signalling, rationalizing the traditional medicinal use of aqueous plant extracts from *J. curcas* for improved wound-healing. Therefore, the present study aimed at the isolation, purification and advanced structural analysis of the AGP by detailed analysis.

2. Experimental

2.1. General

If not stated otherwise all chemicals were purchased from Sigma (Deisenhofen, Germany) and VWR (Darmstadt, Germany). *J. curcas* L. seeds were obtained from KPR – Gardeners Club, Slovakia. Identification was performed by senior author A. Hensel. A voucher species is deposited under code number 283 in the archives of the Institute of Pharmaceutical Biology and Phytochemistry (IPBP), University of Münster, Germany.

Abbreviations: AEC, anion exchange chromatography; AGP, arabinogalactan protein; DP, degree of polymerization; FLA, fasciclin-like arabinogalactans; GPC, gel permeation chromatography; HPAEC-PAD, high pressure anion exchange chromatography with pulsed-amperometric detection; LC, liquid chromatography; LYSM, LysM domain-containing proteins; MS, mass spectrometry; MALLS, multi-angel laser light scattering; MW, molecular weight; SPB, sodium phosphate buffer; TFMS, trifluoromethanesulphonic acid; XLP, xylogen-like proteins.

* Corresponding author. Tel.: +49 251 8333380; fax: +49 251 8338341.
E-mail address: ahensel@uni-muenster.de (A. Hensel).

0144-8617/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.carbpol.2013.06.003

2.2. Isolation of Y2

2.2.1. Isolation procedure

The testa of *J. curcas* seeds (3 kg) were removed. The resulting material (1.474 g) was ground in a mortar and defatted with petroleum benzine for 24 h in a Soxhlet apparatus. 630 g of the defatted endosperm were extracted with the 14-fold amount of water for 24 h at 4 °C. After centrifugation (14,000 × *g*, 13 min) high molecular weight material of the supernatant was precipitated by drop-wise addition of the 4-fold amount of ice-cold 96% ethanol under stirring. The precipitation was completed overnight at 4 °C. After centrifugation (2300 × *g*, 10 min) the remaining residue was dissolved in water, dialyzed (molecular weight cut-off 3.5 kDa) and centrifuged at 5300 × *g* for 13 min to remove any insoluble residue. This process yielded 1.1% (related to the starting seed material) of soluble fraction JC1. Non-glycoproteins were depleted by a 5 min treatment of JC1 solution at 100 °C, followed by immediate ice-cooling. The precipitated protein was removed by centrifugation (10 min at 3100 × *g*). The remaining supernatant was lyophilized to obtain JC2.

2.2.2. Chromatographic purification of JC1

Anion exchange chromatography (AEC) of JC2 (4.5 g in 60 mL water) was performed on DEAE-Sephacel® (GE Healthcare, Freiburg, Germany): 5.2 cm × 16 cm i.d., phosphate form, flow 3 mL/min, fraction size 10 mL, 85 fractions per gradient step, stepwise gradient with deionized water and $NaPO_4$ buffers (pH 5, ion strength 0.1, 0.25, 0.5, 1 mol/L). Carbohydrate-containing fractions were pooled, concentrated *in vacuo*, dialyzed and lyophilized. The main fraction named R2 was isolated from the 0.1 mol/L eluate (yield 0.08% related to the starting seed material).

2.2.3. Precipitation of AGPs from R2 by Yariv reagent

AGPs were isolated from R2 by precipitation with β-D-glucosyl-Yariv reagent, which was synthesized from *p*-nitrophenyl-β-D-glucopyranoside according to Yariv, Rapport, and Graf (1962). Identity and purity of the product was confirmed by UV–vis, HR-MS, ^1H and ^{13}C NMR (COSY, HSQC, HMBC) in DMSO-d_6.

100 mg of β-D-glucosyl Yariv reagent were dissolved in 100 mL of NaCl solution (0.15 M) and 400 mg R2 were added. The solution was mixed, AGPs were allowed to precipitate for 18 h at 4 °C, followed by centrifugation (1700 × *g*, 5 min). The resulting sediment was washed 4 times with 50 mL NaCl (0.15 M) with centrifugation after each washing step. The residue was suspended in 30 mL of water. Solid sodium dithionite was added to a final concentration of 10% (w/v) and the mixture was heated under stirring to 50 °C until the red colour disappeared. The solution was extensively dialyzed (MWCO 12–14 kDa) against water, centrifuged (3100 × *g*, 5 min) and lyophilized to yield Y2 (yield 0.01% related to the starting seed material).

2.3. Carbohydrate analysis

Standard methods of carbohydrate analysis (carbohydrate quantification, uronic acids, polysaccharide and protein hydrolysis, HPAEC-PAD for carbohydrates and amino acids, GC-FID analysis of alditol acetates, linkage analysis by GC–MS identification of partially methylated alditol acetates, D/L-configuration, low pressure GPC on Superose®) were performed according to methods previously described by Hermann et al. (2012) and Zippel, Deters, Pappai, and Hensel (2009). Quantitative analysis of monosaccharide composition of Y2 after TFA hydrolysis was performed by HPAEC-PAD using external standard calibration with the respective reference carbohydrates and, for verification, by GC-FID of the respective alditol acetates. The total amount of uronic acids was determined by colourimetric assay according to Blumenkrantz and Asboe-Hansen (1973). Individual uronic acids were determined after hydrolysis of the polymers by HPAEC-PAD. D-/L-configuration was determined *via* capillary electrophoresis after hydrolysis and derivatisation with S-(−)-1-phenylethylamine (Noe & Freissmuth, 1995). Linkage analysis was performed *via* methylation analysis after reductive deuteration of acidic polymers and evaluation of partially methylated alditol acetates by GC/MS.

2.3.1. Determination of molecular weight

HP-SEC was performed on a SECurity GPC system (PSS Polymer Standard Service, Mainz, Germany) consisting of three Suprema® columns from PSS (i.d. 8 mm, particle size; 10 μm 100 Å, 300 mm; 3000 Å, 300 mm; guard column, 50 mm) coupled online to a refractive index detector (Agilent series 1200 RID, Agilent Technologies, Santa Clara, USA), a UV detector (Agilent 1200 Series Variable Wavelength Detector, Agilent Technologies), a viscosimetric detector (PSS SECurity ETA2010) and MALLS detection (PSS SECurity SLD7000 MALLS) equipped with a 5 mW HeNe laser, operating at λ = 632.8 nm. Phosphate buffer (50 mM, pH 7) served as eluent at 0.7 mL/min; calibration was performed with pullulans (PSS Calibration Kit (1.32, 5.9, 10, 22.8, 47.3, 112, 212, 404, 710 kDa) dn/dc = 0.149 (pullulan)) and data analysis by PSS WinGPC Unity V.7.3.0.

2.3.2. NMR spectroscopy

For NMR, 16 mg of Y2 were dissolved in 0.8 mL D_2O (Uvasol®, Merck, Darmstadt, Germany) and filtered through prewashed cotton. ^1H and ^{13}C NMR measurements were obtained at 500 MHz and 125 MHz, respectively (Agilent VNMRS 500). The NMR spectra were referenced to the C6 signal of rhamnose (17.20 ppm in ^{13}C NMR, 1.25 ppm in ^1H NMR, 26 °C). The signals of the anomeric centre were compared to reference standards. To that end, 40 mg of the soluble gum arabic fraction (Caelo, Hilden, Germany) (Defaye & Wong, 1986) and 40 mg of sugar beet arabinan (Westphal, Kuehnel, Schols, Voragen, & Gruppen, 2010) (Südzucker, Obrigheim, Germany) were measured in D_2O containing 3-(trimethylsilyl)-propionic-2,2,3,3-d_4 acid sodium (TMSP). Signals were referenced to the C-6 of rhamnose at 1.25 ppm in ^1H and 17.20 ppm in ^{13}C NMR spectrum.

Data were processed by MestReNova® 8.0.1, Mestrelab Research S.L., Santiago de Compostela, Spain.

2.3.3. Mass spectrometry

Mass spectrometric analysis of oligosaccharides was performed as described recently for the characterization of *N*-glycopeptides derived from plant lectins by use of a quadrupole time-of-flight (Q-TOF) mass spectrometer (Micromass, Manchester, UK) equipped with a nanoelectrospray ionisation (nanoESI) ion source (Kumar, Shivappa, Pohlentz, Mormann, & Kumar, 2013). Neutral sugars were dissolved in methanol/water/formic acid (49/49/2; v/v/v) and analysed in the positive ion mode using a capillary voltage of 1100 V and a cone voltage of 30 V to minimize in-source fragmentation. For the characterization of acidic oligosaccharides analytes were dissolved in methanol/water (50/50; v/v) and submitted to ionisation in the negative ion mode employing a capillary voltage of −1100 V and a cone voltage of 30 V. The source temperature was set to 80 °C and a desolvation gas (N_2) flow rate of 75 l/h was used. Homemade nanospray capillaries were used. For low-energy CID experiments, the glycan-derived precursor ions were selected in the quadrupole analyzer and fragmented in the collision cell using a collision gas (Ar) pressure of 3.0×10^{-5} mbar and collision energies of 15–40 eV (E_{lab}).

2.3.4. X-ray fluorescence analysis

Total reflection X-ray fluorescence (TXRF) was carried out at the Institute of Inorganic and Analytical Chemistry, University of Münster, Germany. 1.6 mg of Y2 in nitric acid 0.2% were analysed using

Bruker S2 Picofox TXRF with a molybdenum anode and gallium as internal standard.

2.3.5. Enzymatic partial degradation

Enzymatic degradation was performed by using *endo*-arabinase (3.2.1.99, 9.0 U/mg, buffer 0.1 M Na acetate, pH 4) from *A. niger*, α-L-arabinofuranosidase (3.2.1.55, 19 U/mg, 0.1 M Na acetate, pH 4) from *A. niger*, α-rhamnosidase (3.2.1.40, 190 U/mg, 0.1 M Na phosphate, pH 6.5), all from Megazyme, Bray, Ireland. Partial purification of a β-(1,3)-galactanase from Driselase® (Sigma, St. Luis, U.S.A.) was carried out according to Tsumuraya, Mochizuki, Hashimoto, and Kovac (1990).

2.4. Peptide analysis

MS analysis of tryptic peptides was performed as described earlier (Hermann et al., 2012). Assignments of unknown proteins based on mass spectrometric peptide fragmentation (König, Herrmann, & Hensel, 2013) using Q-TOF Premier coupled to nanoAcquity LC (Waters Corp., Manchester, UK) was performed as described with the exception that an LC–MS/MS coupling was employed and spectra were acquired in a data-dependent fashion. Targeted experiments focussed on specific ions of interest to improve fragment ion quality.

2.5. Deglycosylation protocols

Trifluoromethanesulphonic acid (TFMS) treatment (Bosch, Knudsen, Derksen, & Mariani, 2001): Y2 (1 mg) dried *in vacuo* over P_2O_5, was dissolved in 1.8 mL of ice-cold anisole/TFMS (1:2) in a glass vial, flushed with nitrogen and stirred at 0 °C for 24 h. Ice-cold diethyl ether/hexane (8 mL, 9:1) were added and the deglycosylated protein was precipitated at −20 °C (12 h). After centrifugation (1 min, 17,000 × *g*) the pellet was washed twice with ice-cold 95% ethanol and centrifuged. Remaining ethanol was removed in a stream of nitrogen.

PNGase F: According to the instruction of the manufacturer (New England Biolabs®, Ipswich, USA) with 0.17 mg of AGP dissolved in 9 μL water. The functionality of the enzyme was proven by the digest of α-antitrypsin as control, which was cleaved properly.

N-glycosidase A: According to the instruction of the manufacturer (Roche Diagnostics, Mannheim, Germany).

β-elimination of O-glycosides: Standard procedure (Montreuil, Bouquelet, Debray, & Lemoine, 1994): 0.6 mg of the sample was dissolved in 50 μL $NaBH_4$ (1 M in 0.1 M NaOH) or in $NaBH_4$ (0.5 M in 0.05 M NaOH) and incubated under an atmosphere of nitrogen at 50 °C in a thermo mixer for 20 h. The reaction was stopped with 10 μL of acetic acid and the mixture was prepared for subsequent SDS-PAGE.

Deglycosylation by GlycoPofile™ β-elimination kit (Sigma, St. Louis, USA): 0.17 μg sample were dissolved in 170 μL water. 34 μL of β-elimination reagent mixture (37.6 μL β-elimination reagent + 2.4 μL of NaOH 5.0 M solution) were added and the reaction mixture was incubated at 5 °C for 20 h. The reaction was stopped by adding 10 μL of HCl 1 M. The sample was directly prepared for subsequent SDS-PAGE.

Purification of digested AGP mix after enzymatic treatment was performed by solid phase extraction on graphitised carbon (Carbograph, Alltech, Deerfield, IL, U.S.A.), conditioned with 240 μL of acetonitrile/trifluoroacetic acid (80/0.1%, v/v) and equilibrated with 300 μL of water. For elution of impurities and neutral monomers, the column was washed with 300 μL water. Neutral oligosaccharides were eluted by acetonitrile 25%, acidic oligosaccharides by acetonitrile/trifluoroacetic acid (25/0.05%, v/v), each 120 μL, followed by washing with 180 μL of acetonitrile/trifluoroacetic acid (75/0.1%, v/v). The eluates were dried in a SpeedVac and used for MS analysis.

3. Results and discussion

3.1. Isolation and purification of AGP

From the defatted seeds of *J. curcas L.* a raw polysaccharide (RPS) was isolated by aqueous extraction from which a water-soluble fraction JC1 was obtained (yield 1.1%). The removal of heat denaturable proteins resulted in fraction JC2 (0.28% related to the seed starting material and 26% of JC1) which was further fractionated by AEC (Fig. S1 Supplementary Data). Neutral polysaccharides were eluted with water and acidic polymers by increasing ion strength. From the main fraction R2, eluting at 0.1 mol/L SPB, an AGP fraction named Y2 was selectively precipitated by β-D-glucosyl Yariv reagent (Kreuger & van Holst, 1995) at a yield of 12%, related to R2.

3.2. Analytical characterization of Y2

3.2.1. Composition of Y2

Y2 contained 81 ± 2% carbohydrate (resorcinol sulphuric acid assay), 5.3 ± 1% uronic acids (Blumenkrantz & Asboe-Hansen, 1973) and 4.8 ± 0.04% polypeptides (HPAEC-PAD quantification of amino acid after hydrolysis against external standard calibration) as determined by 3 independent quantifications. Radial diffusion test with β-glucosyl Yariv reagent for semiquantitative analysis of the AGP content (van Holst & Clarke, 1985) resulted, as expected, in ~100%. Elemental analysis by TXRF indicated the presence of 0.52% calcium, 0.21% sulphur and 0.02% phosphorus ($n = 3$ independent experiments).

3.2.2. Molecular weight of Y2

On GPC with Superose™6 stationary phase under low pressure conditions Y2 eluted as a single peak with a mean MW of ~75 kDa. Also HP-SEC resulted in one single peak (Fig. S2 Supplementary Data). Depending on the calibration method different MWs were calculated: For the standard calibration with pullulans a molecular weight of 81 kDa was determined, the universal calibration by viscosimetry gave 130 kDa and the application of the data of multi angle light scatter measurement for the calculation by Zimm equation resulted in a MW range between 87 and 142 kDa, depending on the refractive index increment *dn*/*dc*.

The MW dispersity D_M (Gilbert et al., 2009) was calculated for the standard calibration measurement to be 1.48, by viscosimetry to be 1.50, and by light scattering a D_M to be 1.09.

SDS-PAGE using either silver staining or Yariv reagent revealed a major band from 260 to 290 kDa for untreated Y2. Additionally, a very weak band at 45 kDa was visible by silver staining; it was however not stained by Yariv reagent (Fig. 1). It was obvious that there is a great discrepancy between the molecular weight determination by SDS-PAGE and the determination be HP-SEC. Interestingly, no improvement in the separation could be achieved by two-dimensional gel electrophoresis of Y2.

3.2.3. Sugar composition and linkage analysis

The sugar composition and the respective linkage parameters of Y2 carbohydrates are summarized in Table 1A, indicating the main structural features of type II arabinogalactans with a (1,3)(1,3,6)-galactose backbone, bearing single branching units of 1-arabinose or short (1,6)-galactose side chains with 1,5-linked arabinose residues attached. The presence of 4-O-methyl-GlcA was excluded after preparation of PMAA with CD_3I as methylating agent.

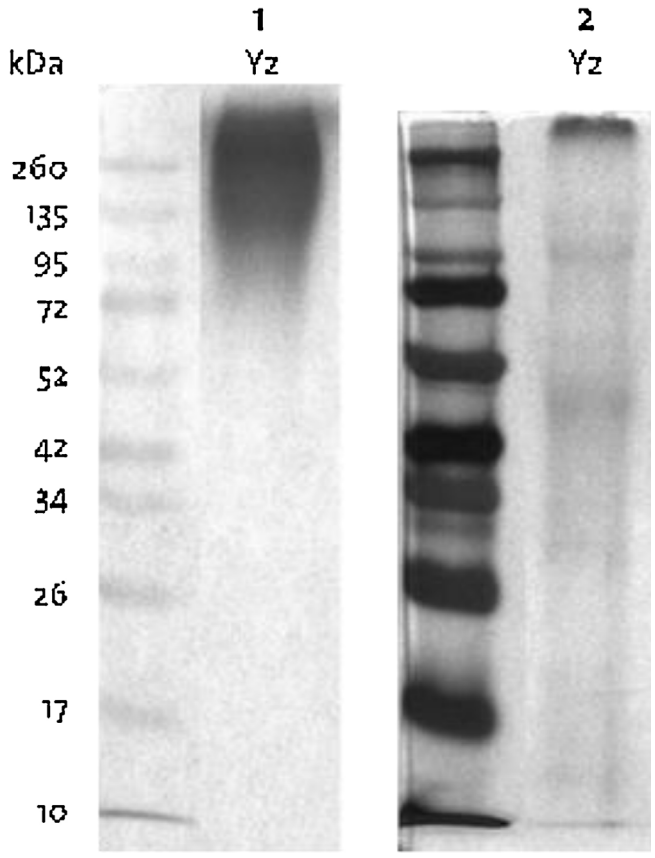

Fig. 1. SDS PAGE (10% polyacrylamide) of Y2 stained with Yariv reagent (**1**) and silver (**2**); protein ladder Spectra®Multicolor proteins 10–260 kDa.

3.2.4. NMR analysis of Y2

For ^{13}C NMR analysis of Y2 comparative investigations with polysaccharides, described in literature were performed, in order to define the NMR signals for the respective anomeric signals unambiguously: gum arabic and sugar beet araban served as reference in ^{13}C NMR. Gum arabic contains a β-(1,3)-D-Gal*p* core with alternate β-(1,6)-Gal*p* branching. α-(1,3)-L-Ara*f* is linked to C-3 of the galactose core and α-L-Rha*p*-(1,4)-β-D-GlcA*p* or 4-O-methyl-β-D-GlcA*p* are attached to C-6 of D-Gal*p*. Sugar beet arabinan consists of an α-(1,5)-L-Ara*f* backbone with α-(1,3)-linked L-Araf residues. This standard was used for the identification of the chemical shift of the anomeric centre of α-(1,5)-L-Ara*f* residues. In literature differing data for the C-1 of terminal α-L-Ara*f* have been published. Sims and Furneaux (2003) have assigned the C-1 resonance of terminal α-L-Ara*f* from gum arabic to 108.7 ppm and the C-1 resonance of α-(1,3)-linked L-Ara*f* to 110.1 ppm. In the same publication yields of the respective monosaccharides were given with 1.38 times more terminal Ara*f* than (1,3)-linked Ara*f*. In our experiments with the reference sample GaF, both signals (110.1 and 108.7 ppm) were detected; the signal area at 110.1 ppm was 1.33 times larger than the signal at 108.7 ppm. Steinhorn, Sims, Carnachan, and Carr (2011) have identified the chemical shift of 1-α-L-Ara*f*, isolated from an AGP from kanuka honey at 110.0 ppm and that of α-(1,3)-L-Ara*f* at 109.0 ppm. Westphal et al. (2010) showed that the chemical shift of terminal α-Ara*f* residues of branched arabinan oligosaccharides is between 109.9 and 110.2 ppm, depending on the molecular size and the linkage position to the attached arabinose residue (position 3 or 5). With respect to the obtained data of GaF and the recent literature, the chemical shift of 110.1 ppm was assigned to the C-1 of terminal α-L-Ara*f*. From these data the anomeric position of α-Ara, β-Gal, β-GlcA, and α-Rha additionally to α-, as well as β-configurated hexoses at the reducing end were determined; the respective chemical shifts found in relevant literature and Y2 are given in Table S1 of the Supplementary Data.

Table 1
(A) Carbohydrate composition (mol%) of Y2 from *J. curcas*: monosaccharide composition as determined by HPAEC-PAD after TFA hydrolysis, results of linkage analysis after methylation of carboxyl-reduced material and GC–MS identification; tr.: traces, l.d.: below limit of detection. (B) Arabinose-, galactose and rhamnose composition of carbohydrate composition of Y2 after partial enzymatic digest with *exo*- and *endo*-arabinase and rhamnosidase. Values represent relative peak area percent of Ara, Gal, Rha-peaks in GC–MS. (C) Quantitative amino acid distribution in Y2 as determined by HPAEC-PAD (mol%).

(A)

Carbohydrate	Y2 [mol%, HPAED]	Linkage type	Y2 [mol%, GC–MS]
L-Arabinose	39	1-Ara*f*	29
		1,5-Ara*f*	10
L-Rhamnose	4	1-Rha*p*	4
D-Galactose	48	1-Gal*p*	0
		1,3-Gal*p*	11
		1,6-Gal*p*	2
		1,3,6-Gal*p*	35
D-Glucuronic acid[a]	6	1-GluA*p*	1
		1,4-GluA*p*	5
D-Mannose	2	1,2-Man*p*	tr.
D-Glucosamin	1	l.d.	l.d.
Fucose	tr.	l.d.	l.d.

(B)

Carbohydrate	Y2 native	*exo*-Arabinase	*endo*-Arabinase	Rhamnosidase
1-Ara*f*	20.4	0	11.9	17.2
1,5-Ara*f*	6.9	0	1.9	7.9
1-Rha*p*	2.5	2.5	2.3	1.7
1-Gal*p*	0	9.1	0	0
1,3-Gal*p*	16.3	17.8	16.8	15.6
1,6-Gal*p*	0.7	7.3	7.7	2.8
1,3,6-Gal*p*	53.2	36.0	45.8	51.8

(C)

Amino acid	[mol%]
Alanine	17.7
Serine	12.9
Glycine	11.5
Hydroxyproline	9.2
Glutamic acid/glutamine	9.0
Lysine	7.3
Threonine	6.7
Leucine	6.7
Aspartic acid/asparagine	6.2
Valine	5.6
Proline	3.7
Isoleucine	3.5
Tyrosine	tr.
Cysteine	tr.
Phenylalanine	tr.
Histidine	tr.
Methionine	tr.

[a] Determined as C-6-reduced Glu.

3.2.5. Enzymatic degradation of Y2 and analytical characterization of fragments

For evaluation of the fine structure, enzymatic partial degradations by *exo*-a-L-arabinofuranosidase, *endo*-arabinase (cleaves α-1,5-L-arabinose chains) and rhamnosidase were performed. The remaining polysaccharides were subjected to methylation analysis. Respective distributions of Ara-, Gal- and Rha-contents in comparison to native Y2 are displayed in Table 1B. Treatment with *exo*-arabinosidase released all Ara residues with strong increase of terminal Gal, 1,6-Gal and 1,3-Gal. The amount of 1,3,6-Gal decreased accordingly. This indicates that all α-(1,5)-Ara*f* chains are terminated also by Ara*f*. Short arabinose-containing side chains

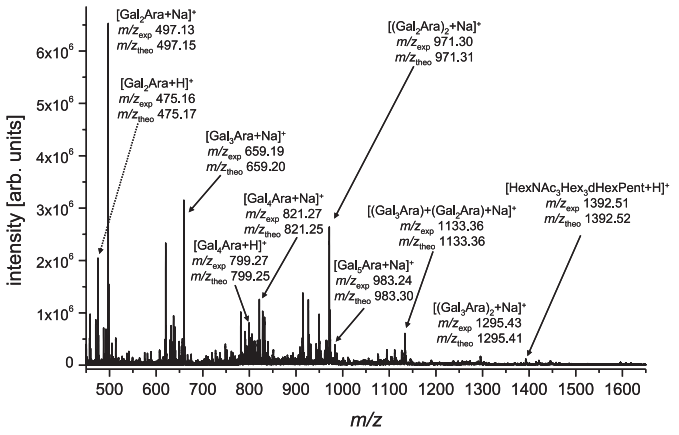

Fig. 2. Exemplary +ESI-QTOF-MS spectrum of neutral oligosaccharide fraction obtained from Y2 after treatment of the polymer for 6 h with an *exo*-(1,3)-galactanase-containing enzyme fraction; the enzymatic digest was purified on graphitized carbon (Carbograph®) and eluted with acetonitril 25%.

linked *via* position 6 to the β-(1,3)-Gal*p* backbone are present. Enzymatic digest by *endo*-arabinase affirmed this result: α-(1,5)-Ara*f* signals were significantly reduced whereas comparable signals for 1,6-Gal residues were observed as for the *exo*-arabinase digest. This indicates that most of the α-(1,5)-Ara*f* branches are located at the (1,6)-Gal*p* chains, which again are linked to the galactan backbone. 1-Rhamnose residues seem to be mainly acting as branching points of the 1,6-Gal*p* side chains.

An *exo*-β-(1,3)-galactanase-containing fraction was partially purified from Driselase from Basidomycetes sp. and galactanase activity was monitored by using gum arabic. According to Tsumuraya et al. (1990) and Nergard et al. (2006) activity of *exo*-β-(1,3)-galactanase should not be dependent on unsubstituted (1,3)-Gal chains and should cleave also β-(1,3)-Gal chains which are branched in position 6. Therefore, this enzyme can be used for degrading (1,3)-Gal backbones which are branched in position 6; this should lead to the specific release of substituted (1,6)-galactose oligomers. This enzymatic digest of Y2 resulted in nine oligosaccharide fractions. Additionally partial acid hydrolysis of Y2 with TFA (0.5 M, 100 °C, 60 min) was performed.

The sugar composition of all fractions was investigated by use of nanoESI-Q-TOF MS and MS/MS experiments. A typical spectrum of the neutral fraction of Y2 after treatment with *exo*-(1,3)-galactanase is displayed in Fig. 2. Intact gaseous ions of Gal_2Ara, Gal_3Ara, Gal_4Ara, Gal_5Ara, and $HexNac_3Hex_3PentdHex$ have been observed. The structures were confirmed by MS/MS experiments.

Table 2 summarizes the most relevant data obtained for the neutral and acidic fractions, respectively. Rha is always linked to GlcA. Gal containing oligosaccharides (DP 2–7) are linked to GlcA- and Ara residues. This correlates with the data obtained from enzymatic arabinase degradation.

Three neutral fractions from the enzymatic digest showed signals corresponding to HexNAc-containing species, pointing to the presence of *N*-glycans (Table 2, Fig. 3). Longer incubation times result in an increased amount of glycans comprising two or more *N*-acetylhexosamine residues. Three ionic species were detected and their sugar building block composition was assigned according to their *m*/*z* values: $[HexNAc_2Hex_3dHexPent+H]^+$(*m*/*z* 1189.43), $[HexNAc_3Hex_3dHexPent+H]^+$ (*m*/*z* 1392.51), and $[HexNAc_4Hex_3dHexPent+H]^+$ (*m*/*z* 1189.43). Additionally to the MS investigations all oligosaccharide fractions were further investigated by HPAEC after TFA hydrolysis for unambigious identification of sugar composition.

MS/MS analysis (for a representative MS/MS spectrum see Fig. S3 in Supplementary Data) of the fragments and investigation of

Table 2
Ions detected by +/−ESI-QTOF-MS and assignments to oligosaccharide composition in an analyte mixture obtained from Y2 after treatment of the polymer with an *exo*-(1,3)-galactanase containing enzyme fraction and non-enzymatic partial acid hydrolysis (TFA 0.5 M, 1 h, 100 °C).

m/z_{exp}	m/z_{theo}	Ionic species
Neutral fractions after enzymatic cleavage		
475.13	475.15	$[Gal_2Ara+H]^+$
497.13	497.15	$[Gal_2Ara+Na]^+$
659.19	659.20	$[Gal_3Ara+Na]^+$
799.27	799.25	$[Gal_4Ara+H]^+$
821.27	821.25	$[Gal_4Ara+Na]^+$
971.30	971.31	$[(Gal_2Ara)_2+Na]^+$[a]
983.24	983.30	$[Gal_5Ara+Na]^+$
1133.36	1133.36	$[(Gal_3Ara)(Gal_2Ara)+Na]^+$[b]
1295.43	1295.41	$[(Gal_3Ara)_2+Na]^+$[a]
Acidic fractions after enzymatic cleavage		
339.11	339.09	$[RhaGlcA-H]^-$
355.10	355.09	$[GalGlcA-H]^-$
649.23	649.18	$[Gal_2AraGlcA-H]^-$
781.26	781.23	$[Gal_2Ara_2GlcA-H]^-$
795.28	795.24	$[Gal_3AraGlcARha-H]^-$
927.32	927.28	$[Gal_2Ara_2GlcARha-H]^-$
Y2-TFA (non-enzymatic cleavage)		
1165.47	116.35	$[Gal_6GlcA-H]^-$
1327.54	1327.40	$[Gal_7GlcA-H]^-$
1503.62	1503.44	$[Gal_7GlcA_2-H]^-$
GlcNAc containing fragments (*N*-glycans)		
1189.43	1189.44	$[Man_3GlcNAc_2XylFuc+H]^+$
1392.51	1392.52	$[Man_3GlcNAc_3XylFuc+H]^+$
1595.60	1595.59	$[Man_3GlcNAc_4XylFuc+H]^+$

[a] Cation-bound homodimer.
[b] Cation-bound heterodimer.

the respective monomer composition by HPAEC-PAD and GC-FID indicated additionally the existence of a complex N-glycan linker with a structural proposal as shown in Fig. 4A. The modifications of β-xylose-(1→2) to the branched β-Man and α-Fuc-(1→3) to the proximal GlcNAc are typical features of *N*-glycans in plants (Staudacher, Altmann, Wilson, & Marz, 1999). In comparison to the general building scheme (Etzler & Mohnen, 2009) of such complex type *N*-glycans (Fig. 4B), which result from highly conserved biosynthetic pathways, the *N*-glycan linker from Y2 was found to contain less fucose residues, and this carbohydrate was only linked to the first GlcNac (or GlcN). Fucosylation on the antennal GlcNAcs as described in literature (Etzler & Mohnen, 2009) has not been detected in Y2.

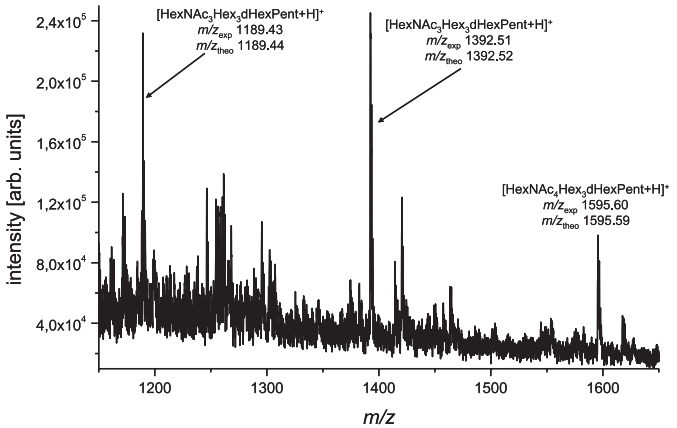

Fig. 3. Exemplary +ESI-QTOF-MS spectrum with typical masses related to HexNAc-containing oligosaccharides obtained from neutral oligosaccharide fraction obtained from Y2 after treatment of the polymer for 22 h with an *exo*-(1,3)-galactanase containing enzyme fraction; enzymatic digest was purified on graphitized carbon (Carbograph®) and eluted with acetonitril 25%.

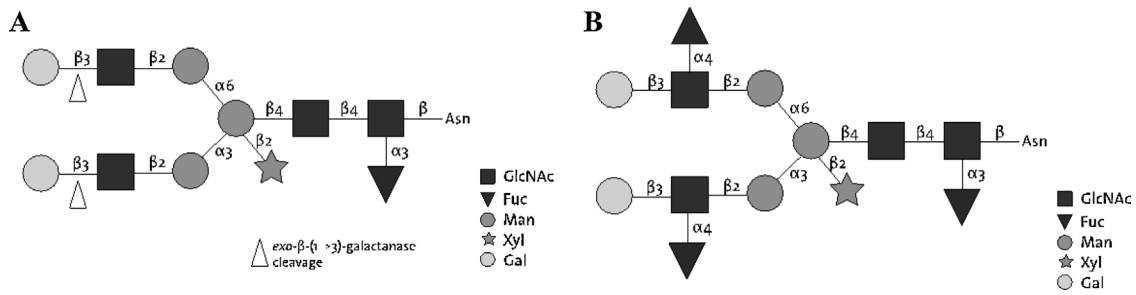

Fig. 4. Proposed structural features of an unusual N-glycan-linker identified from Y2 (A) in comparison to the general scheme of complex type N-glycan linkers from plants according to Etzler and Mohnen (2009) (B).

3.2.6. Search for GPI-anchors in Y2

Classical AGPs are known to contain GPI-anchors (Albersheim, Darvill, Roberts, Sederoff, & Staehelin, 2006, Chap. 3). GC–MS analysis of a hexane extract of acetic acid/acetic anhydride-treated Y2 did not give any indications for the existence of lipids typical for GPI-anchors. HPAEC-PAD and GC–MS analysis did not find evidence for inositol. Based on these results, no experimental indication for the presence of GPI-anchors in Y2 was found. Summarizing these data a structure for the polymeric carbohydrate part of Y2 is proposed (Fig. 5).

3.3. Analytical investigation of the Y2 protein part

The overall content of amino acids in Y2 detected after hydrolysis was 4.8 ± 0.04% with alanine, serine, glycine, hydroxyproline and glutamate being the major monomers (Table 1C). This result indicated a hydroxyproline-rich AGP (Sommer-Knudsen, Bacic, & Clarke, 1979). The presence of ~12% Gly was not considered to be typical for AGPs, but recent reports have described such Gly-rich AGPs (Gleeson & Clarke, 1979; Hermann et al., 2012; Serpe & Nothnagel, 1995). In general, amino acid hydrolysis reflected only partially protein identification (see below and Table S7 of Supplementary Data).

In order to elucidate the type of glycosylation, various methods were employed. Alkaline treatment ($NaOH/NaBH_4$) degraded the AGP only to a very limited extend. Total deglycosylation using the GlycoProfile™ β-Elimination Kit and subsequent SDS-PAGE of the reaction mixture showed a time-dependent increase of a cloudy band at 30 kDa whereas the band typically seen for Y2 (135–260 kDa) was still present. This indicated that some glycans in Y2 are *O*-glycosidically linked to serine or threonine and that these *O*-glycosides account for only a minor part of the Y2 glycosylation pattern. The co-existence of *O*-glycans *via* Ser/Thr/Hyp as well asas *N*-glycans can therefore be assumed.

Deglycosylation by TFMS should remove carbohydrates linked to Ser/Thr/Hyp, whereas e the N-glycosidic linkage between GlcNAc and asparagine should be stable so that the proximal GlcNAc residues of N-linked carbohydrate chains remain at the protein (Edge, 2003). After 6 h of TFMS/anisole treatment of Y2, bands at 20 and 40 kDa appeared in 1D-PAGE, although the original band remained at about the same intensity. Prolongation of the incubation to 24 h diminished the intensity of the AGP band substantially with coincident increase of the intensity of the bands at 20 and 40 kDa (Fig. S4 of Supplementary Data) indicating that Y2 contains peptide–carbohydrate bonds that can be cleaved by prolonged TFMS treatment. The usual time for a complete deglycosylation of heavily glycosylated proteins like AGPs is three to 4 h at 0 °C (Bosch et al., 2001) As the deglycosylation by alkaline β-elimination did reveal only little cleavage, most of the O-linked sugars most likely are connected to hydroxyproline.

For investigation of potential *N*-glycosylation, Y2 was treated with PNGase F (EC 3.5.1.52) and N-glycosidase A (EC 3.5.1.52).

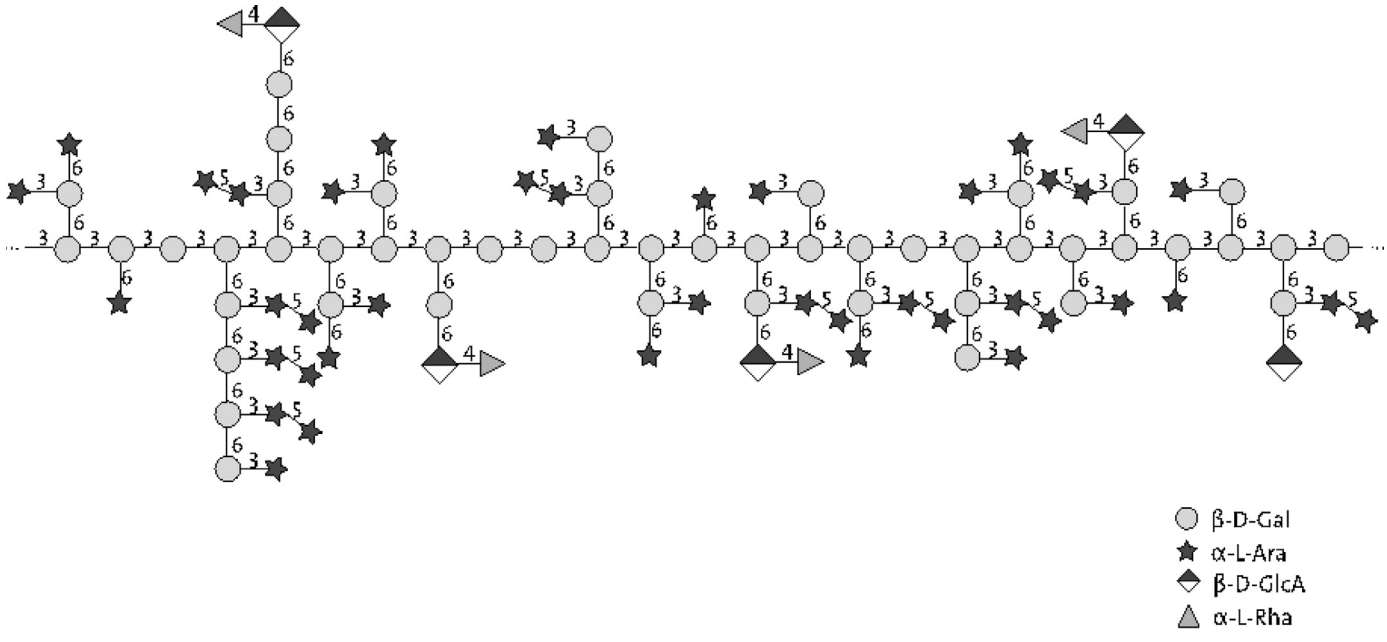

Fig. 5. Proposed structure of the carbohydrate moiety in Y2.

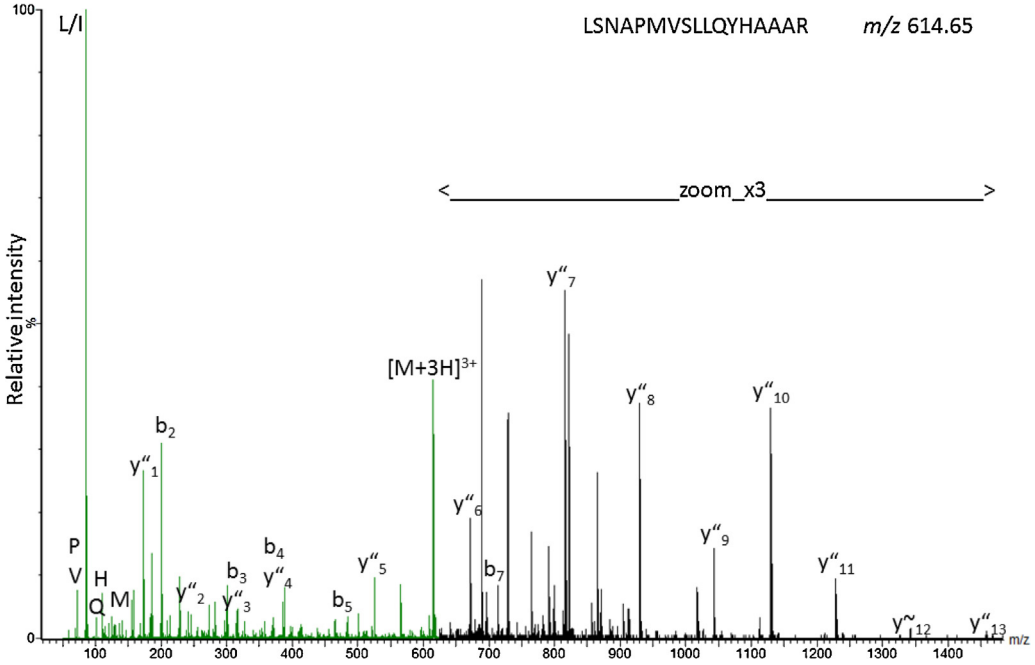

Fig. 6. MS/MS spectrum of *m/z* 614.65 detected in the tryptic digest of the Y2 high MW band. The MS BLAST search with a sequence tag led to sequence LSNAPMVSLLQYHAAAR which has homologous expressions in FLA proteins (see Table S2 of Supplementary Data).

PNGase F cleaves between asparagine and GlcNAc of oligomannose or hybrid N-glycans, but does not affect α-(1,3)-fucosylated innermost GlcNAc residue. In contrast, *N*-glycosidase A cleaves between asparagine and GlcNAc of *N*-glycans and thus in cases where GlcNAc carries an α-(1,3)-linked fucose.

The digest of Y2 with PNGase F did not lead to a MW change of the Y2 band in 1D-PAGE. This result indicated that no *N*-glycan substrates for PNGase F were present.

N-glycosidase A activity is improved, when the glycoprotein is submitted to proteolytic digestion, *e.g.* by trypsin first. In a comparative analysis of *N*-glycosidase A-treated and non-treated Y2 highly complex peptide profiles were measured by LC–MS. In case of trypsin-digested Y2 many more signals were detected in the mass range of *m/z* 1200–1400 in contrast to the *N*-glycosidase A treated sample, which exhibited a drastic increase of ionic species in the *m/z*-range between 600 and 800. Following improved protein separation (see below) detailed analysis of specific glycopeptides will provide more specific information, but already this crude profiling experiment indicated the presence of fucose-bearing GlcNAc residues in *N*-glycans (see Section 3.2). So far, evidence for *N*- and *O*-glycosidic bonds has been provided.

For protein identification, the high molecular part of Y2 (140–250 kDa band in 1D-PAGE) was trypsinized and analysed by LC–MS/MS. Using *de novo* sequencing (König et al., 2013), gas phase fragmentation spectra of non-glycosylated peptides were assigned to representatives of three protein groups, fasciclin-like arabinogalactan (FLA), xylogen-like proteins (XLP) and LysM domain-containing proteins (LYSM). *J. curcas* AGP sequences had not been described before so that the protein assignment relied on the search for homologous peptides from other plant species. Proteins were only considered when consecutive sequence stretches of at least 10 amino acid residues were matched (for MS detection limits see Hermann et al., 2012; König et al., 2013). An exemplary spectrum is shown in Fig. 6; further details on FLA, XLP and LYSM matching to known proteins from other genera such as Arabidopsis, Gossypium, Medicago, Ricinus, Populus or Vitis are given in Tables S2–S5 of Supplementary Data and additional MS/MS exemplary fragmentation spectra can be found in Figures S5–6 of Supplementary Data. Having taken into account possible sequence variations among species, best matches were obtained from Ricinus and Populus (FLA: GI 255541352, 224130034; XLP: 255578847, 118486833; LYSM: 25554572). Multiple FLA proteins seemed to be present as suspected by spectra fitting specific peptides such as FLDYLLSTK (Vitis FLA7, GI 225446995; Table S3 of Supplementary Data). However, fasciclin-domain (FAS1/BIgH3, see UniProt/Prosite)-containing proteins have been described to contain two or four copies of the domain possibly with slight sequence variations. The FAS1 domain (~140 amino acid residues) is reported to be an extracellular cell adhesion domain common in plants and animals and the proteins are often GPI anchored (Huber & Sumper, 1994). The crystal structure of two Drosophila FAS1 domains has been solved (Clout, Tisi, & Hohenester, 2003) showing a seven-stranded wedge and at least five α-helices in each. A conserved asparagine in the interface region between the two FAS1 domains carries two *N*-acetylglucosamine moities. A Prosite scan for known sequence features on the Ricinus FLA described here, revealed more possible glycosylation sites (Table S6 of Supplementary Data) in addition to other features of high probability (CK2 and PKC phosphorylation and myristoylation sites), but they are not necessarily truly modified. Moreover, the *C*-terminally located 54-amino acid residue long stretch of a proline-rich region in FLA has been highlighted, although no particular function is yet known for it.

More proteins may exist in Y2, but they could not be assigned with equal confidence and further identification was complicated until protein separation was improved. Analyses were further complicated by protein modifications as exemplified for XLP peptide TVLKTDAQCNCEAFK (Fig. 7). Spectral evidence suggested the successive loss of 80 Da upon collision-induced dissociation which is typically observed for phosphorylated, but also *O*-sulphonated peptides (Fig. S7 of Supplementary Data). The presence of phosphorus had already been proven by TXRF for Y2. The addition of 80 Da was also seen in a LYSM peptide containing the sequence tag GPQEFPPLLAP. In both cases no phosphorylation site was predicted so that more detailed site analysis is called for. Other modifications could be artefactual such as oxidation, deamidation and the

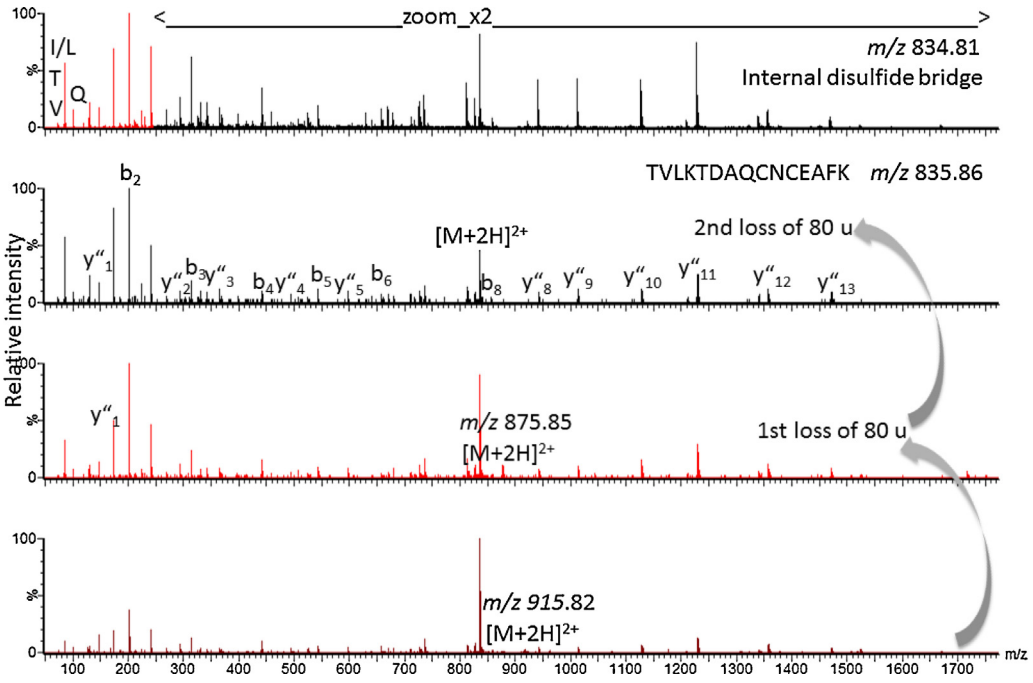

Fig. 7. MS/MS spectra corresponding to sequence TVLKTDAQCNCEAFK indicating twice phosphorylation as well as the formation of an internal disulphide bridge (for oxidation and deamidation see Table S4 of Supplementary Data).

formation of an internal disulphide bridge (Fig. S7 of Supplementary Data). They may occur spontaneously due to sample handling. On the other hand, plant lipid transfer proteins (LPT), a group of protein XLP (see domain descriptions for IPR013770, IPR016140, which is described as a bifunctional inhibitor of LPT/seed storage helical domain, IPT000528, which is described asPAR allergen), are known to contain eight conserved cysteine residues forming four disulphide bridges. The LPT protein family transfers phospholipids, glycolipids, fatty acids and sterols between lipid membranes and is thought to be involved in defence, pollination and germination (UniProt).

To our knowledge, LYSM has not been described as AGP so far, but is known as peptidoglycan-binding lysine domain (IPR018392, IPR002482) found in enzymes involved in bacterial cell wall degradation. However, the AGP-complex precipitated by Yariv reagent may contain a dominant glycoprotein dragging other proteins along.

FLAs are key players in the early stages of embryo development. XLPs act within the intracellular communication and stimulate xylem cell differentiation. LYSM proteins trigger plant immune response against fungi or symbiotic bacteria. These physiological activities are congruent with the presence of the respective AGPs in seed embryos. Yariv staining of sagittal and transversal sections visualized AGPs in *J. curcas* seeds (Fig. 8). Unspecific colouration can be excluded by respective negative controls of unstained embryos. No AGPs were detected in the endosperm, while intense staining was observed in the vessels of the cotyledons and within the procambium ring; especially the embryogenetic vessels, forming the vascular bundle, contain AGP.

Summarizing, Y2 can be described as a cluster of non GPI-anchored *N*- and *O*-glycosidic linked hydroxyproline-rich AGPs, which may be present in a modified 'wattle blossom' system, as the glycan chains of Y2 are assumed to be large polysaccharides with a high degree of branching. The presence of a complex *N*-glycan linker in Y2 suggests linkage between GlcNAc and Asn. The connection of Ara or Gal residues to hydroxyproline is most likely; it is an often described O-glycan type in AGPs (Albersheim et al., 2006, Chap. 3; Kieliszewski, Lamport, Tan, & Cannon, 2011; Showalter,

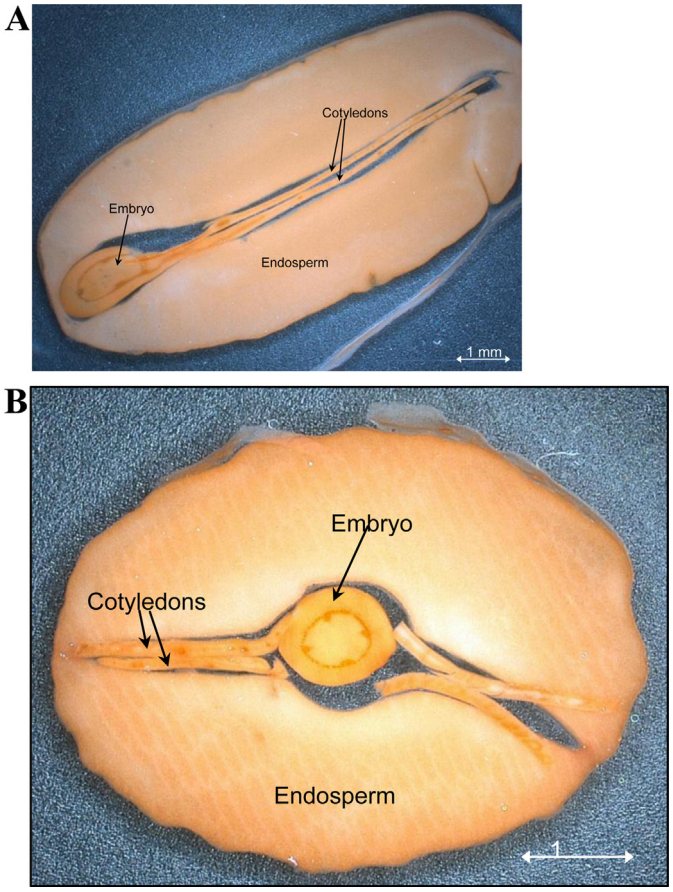

Fig. 8. *J. curcas* seed sections after staining with β-D-glucosyl Yariv: (A) sagital and (B) transversal section.

2001). Hyp occurs in Y2 in high concentrations. The Hyp contiguity hypothesis of Kieliszewski and Lamport (1994) predicts arabinosylation of contiguous Hyp residues and addition of arabinogalactan polysaccharides *via* linkage of Gal to clustered, non-contiguous Hyp. Later it was shown that a gum arabic glycoprotein, which contained both contiguous and non-contiguous Hyp residues, directed both modes of glycosylation (Kieliszewski, 2001). Based on the present findings, the glycosylation pattern of Hyp in Y2 is likely to follow the Hyp contiguity hypothesis.

A Gal-Ser linkage is indicated as well by site-specific digestion. As neither Glc nor GalNAc were found, a linkage of Glc-Hyp and GalNAc-Thr can probably be excluded. Xylose was present only in traces and could be assigned to the complex-type *N*-glycan linker so that its connection to Thr seems unlikely. Y2 can be described as a mixture of soluble AGPs of approximately the same size, similar acidity and charge, precipitable by β-D-glucosyl-Yariv reagent. The answer to this can only be given in respect to the function of each AGP, which is not fully elucidated until now. On the other hand, it can be speculated that certain proteins get arabinogalactosylated in the golgi system for regulation of the protein function, especially for inactivation during seed dormancy.

The presence of AGP with xylogen-like properties in the embryogenic vessels is well understandable for this key-player in vascular development (for a review see Fukuda, Hirakawa, & Sawa, 2007). The determined amino acid sequence could be assigned to a 'non-proline rich' domain of xylogen proteins, which are characteristic for non-classical AGPs (Albersheim et al., 2006, Chap. 3).

3.3.1. How can this cluster of AGPs from the seeds be assessed concerning tissue physiology and what can we conclude?

FLA proteins are known as a distinct subclass of GPI-anchored AGPs (Johnson, Jones, Bacic, & Schultz, 2003), occurring besides classical AGPs, those with Lys-rich domains and AGPs with short peptide backbones. FLAs contain one to two FAS1 domains and additionally one or two AGP-like regions that are based on the presence of at least two non-contiguous Pro residues, for example (A/S)P(A/S)P. These regions are predicted to be *in vivo* Hyp-O-glycosylated, which was proven for at least three FLAs (Johnson et al., 2003). On the other hand, typical N-glycosylation motifs (Asn-X-Ser/Thr, X can be any amino acid, except proline) were identified as well (see Supplementary Data Table S6).

FAS1 domains have then been identified as putative cell-adhesion domains, found in extracellular matrix proteins of organisms from all kingdoms and play important roles in plant development by involvement in cell–cell contacts and cell adhesions due to protein–protein (FLA) as well as protein–carbohydrate (arabinogalactan) interactions (Gaspar, Johnson, McKennam, Bacic, & Schultz, 2001; Johnson, Kibble, Bacic, & Schultz, 2011) assume that FLA1 in *Arabidopsis thaliana* plays an important role in formation of roots and in shoot development.

FLAs are predicted to be originally GPI-anchored, but then may be released from the membrane. Using this background, the occurrence of FLA AGPs in *J. curcas* seeds can be explained. In its early stage of development the embryo needs cell-cell adhesion and later, when growing, FLAs may become important in inducing root and shoot development. This hypothesis correlates well with the higher concentration of AGPs in the embryo than in the endosperm. The lysin motif (LysM) first was described as a small protein domain, often present in lytic enzymes from bacteria, and was suspected to be responsible for the binding of the enzymes to cell wall substrates (Fliegmann et al., 2011). Thereafter, LysM was not only found in bacterial cell wall-degrading enzymes, but also in proteins from eukaryotes. LysM-domains from plant have been shown to interact with chitin and chitooligomers (Fliegmann et al., 2011). Furthermore, LysM domains of some plant kinases recognise symbiotic bacteria, by interaction with chitin-like compounds that are secreted by these symbiotic bacteria (Buist, Steen, Kok, & Kuipers, 2008). LysM domain-containing proteins have not been described as arabinogalactan proteins so far. But Fliegmann et al. (2011) suggest glycosylation of MtLYM, a LysM from *Medicago truncatula*, at the proline-rich regions according to the 'Hyp contiguity hypothesis', postulated from Kieliszewski (2001), which would result in addition of arabinogalactan chains to non-contiguous Hyp while contiguous Hyp residues are getting only decorated by arabinose residues. As the postulated LysM domain-containing protein of *J. curcacs* seeds was isolated in an AGP fraction by β-D-glucosyl Yariv reagent precipitation, it can be stated that this glycoprotein is definitely an arabinogalactan protein. A LysM domain-containing protein is here described as AGP for the first time. According to the described functionality of LysM by interacting with chitin or GlcNAc-oligomers and triggering thereby the cells into an immune defence against fungi or recognising symbiotic bacteria, these proteins are likely to act in the same way in *J. curcas* seeds. Especially in seeds a well-working immune response against fungi is important, because they have to defy fungous infections in humid soil and keep their germination ability.

Acknowledgements

The authors thank Dr. K. Bergander for NMR measurements, Mrs. B. Quandt for performing CE-analysis and Prof. Dr. U. Karst and coworkers for TXRF analysis. Financial support by the German Research Foundation DFG (International Research Training Group 1549 Molecular and Cellular Glyco-Sciences MCGS, Münster Hyderabad) is acknowledged.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.carbpol.2013.06.003.

References

Albersheim, P., Darvill, A., Roberts, K., Sederoff, R., & Staehelin, A. (2006). *Plant cell walls, from chemistry to biology* (1st ed.). New York: Garland Publishing.
Blumenkrantz, N., & Asboe-Hansen, G. (1973). New method for quantitative determination of uronic acids. *Analytical Biochemistry*, *54*, 484–489.
Bosch, M., Knudsen, J. S., Derksen, J., & Mariani, C. (2001). Class III pistil-specific extensin-like proteins from tobacco have characteristics of arabinogalactan proteins. *Plant Physiology*, *125*, 2180–2188.
Buist, G., Steen, A., Kok, J., & Kuipers, O. P. (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. *Molecular Microbiology*, *68*, 838–847.
Clout, N. J., Tisi, D., & Hohenester, E. (2003). Novel fold revealed by the structure of FAS1 domain pair from the insect cell adhesion molecule fasciclin I. *Structure*, *11*, 197–2013.
Defaye, J., & Wong, E. (1986). Structural studies of gum arabic, the exudate polysaccharide from Acacia senegal. *Carbohydrate Research*, *150*, 221–231.
Edge, A. S. B. (2003). Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: Elucidation of molecular structure and function. *Biochemical Journal*, *376*, 339–350.
Etzler, M. E., & Mohnen, D. (2009). Viridiplantae. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, & M. E. Etzler (Eds.), *Essentials of glycobiology* (pp. 321–332). New York: Cold Spring Harbor Laboratory Press.
Fairless, D. (2007). Biofuel: The little shrub that could – maybe. *Nature*, *449*, 652–665.
Fliegmann, J., Uhlenbroich, S., Shinya, T., Martinez, Y., Lefebvre, B., Shibuya, N., et al. (2011). Biochemical and phylogenetic analysis of CEBiP-like LysM domain-containing extracellular proteins in higher plants. *Plant Physiology and Biochemistry*, *49*, 709–720.
Fukuda, H., Hirakawa, Y., & Sawa, S. (2007). Peptide signaling in vascular development. *Current Opinion in Plant Biology*, *10*, 477–482.
Gaspar, Y., Johnson, K. L., McKennam, J. A., Bacic, A., & Schultz, C. J. (2001). The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Molecular Biology*, *47*, 161–176.
Gilbert, R. G., Hess, M., Jenkins, A. D., Jones, R. G., Kratochvil, P., & Stepto, R. F. T. (2009). Dispersity in polymer science (IUPAC recommendations 2009). *Pure and Applied Chemistry*, *8*, 351–353.

Gleeson, P. A., & Clarke, A. E. (1979). Structural studies on the major component of Gladiolus style mucilage, an arabinogalactan-protein. *Biochemical Journal*, *181*, 607–621.

Goel, G., Makkar, H. P. S., Francis, G., & Becker, K. (2007). Phorbol esters: Structure, biological activity, and toxicity in animals. *International Journal of Toxicology*, *26*, 279–288.

Hermann, A., König, S., Lechtenberg, M., Sehlbach, M., Vakrushev, S., Peter-Katalinic, J., et al. (2012). Proteoglycans from *Boswellia serrata* Roxb. and *B. carteri* Birdw. and identification of a proteolytic plant basic secretory protein. *Glycobiology*, *22*, 1424–1439.

Huber, O., & Sumper, M. (1994). Algal-CAMs: Isoforms of a cell adhesion molecule in embryos of the alga Volvox with homology to Drosophila fasciclin. *EMBO Journal*, *13*, 4212–4222.

Johnson, K. L., Jones, B. J., Bacic, A., & Schultz, C. J. (2003). The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. *Plant Physiology*, *133*, 1911–1925.

Johnson, K. L., Kibble, N. A. J., Bacic, A., & Schultz, C. J. (2011). A fasciclin-like arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana*, fla1, shows defects in shoot regeneration. *PLoS ONE*, *69*, e25154.

Kieliszewski, M. J. (2001). The latest hype on Hyp-O-glycosylation codes. *Phytochemistry*, *57*, 319–323.

Kieliszewski, M. J., & Lamport, D. T. A. (1994). Extensin: Repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant Journal*, *5*, 157–172.

Kieliszewski, M. J., Lamport, D. T. A., Tan, L., & Cannon, M. C. (2011). Hydroxyproline-rich glycoproteins: Form and function. *Annual Plant Reviews: Plant Polysaccharides*, *41*, 321–342.

König, S., Herrmann, A., & Hensel, A. (2013). Assignment of unknown proteins based on mass spectrometric peptide fragmentation. *Biomacromolecular Mass Spectrometry*, *3*, 61–75.

Kreuger, M., & van Holst, G. J. (1995). Arabinogalactan proteins are essential in somatic embryogenesis of *Daucus carota* L. *Planta*, *189*, 243–248.

Kumar, G., Shivappa, B., Pohlentz, G., Mormann, M., & Kumar, N. S. (2013). Characterization of a-mannosidase from Dolichos lablab seeds: Partial amino acid sequencing and N-glycan analysis. *Protein Expression and Purification*, *89*, 7–15.

Lin, J., Zhou, X., Wang, J., Jiang, P., & Tang, K. (2010). Purification and characterization of curcin, a toxic lectin from the seed of *Jatropha curcas*. *Preparative Biochemistry and Biotechnology*, *40*, 107–118.

Montreuil, J., Bouquelet, S., Debray, H., & Lemoine, J. (1994). Glycoproteins. In M. F. Chaplin, & J. F. Kennedy (Eds.), *Carbohydrate Analysis, A Practical Approach*. New York: Oxford University Press.

Nergard, C. S., Kiyohara, H., Reynolds, J. C., Thomas-Oates, J. E., Petersen, D., Michaelsen, T. E., et al. (2006). Structures and structure–activity relationships of three mitogenic and complement fixing pectic arabinogalactans from the Malian antiulcer plants Cochlospermum tinctorium A. Rich and Vernonia kotschyana Sch. Bip. exWalp. *Biomacromolecules*, *7*, 71–79.

Noe, C. R., & Freissmuth, J. (1995). Capillary zone electrophoresis of aldose enantiomers: Separation after derivatization with S-(−)-1-phenylethylamine. *Journal of Chromatography*, *704*, 503–512.

Serpe, M. D., & Nothnagel, E. A. (1995). Fractionation and structural characterization of arabinogalactan-proteins from the cell wall of rose cells. *Plant Physiology*, *109*, 1007–1016.

Sims, I. M., & Furneaux, R. H. (2003). Structure of the exudate gum from Meryta sinclairii. *Carbohydrate Polymers*, *52*, 423–426.

Showalter, A. M. (2001). Arabinogalactan-proteins: Structure, expression and function. *Cellular and Molecular Life Science*, *58*, 1399–1417.

Sommer-Knudsen, J., Bacic, A., & Clarke, A. E. (1979). Hydroxyproline-rich plant glycoproteins. *Phytochemistry*, *47*, 483–497.

Staudacher, E., Altmann, F., Wilson, I. B. H., & Marz, L. (1999). Fucose in N-glycans: From plant to man. *Biochimical and Biophysical Acta*, *1473*, 216–236.

Steinhorn, G., Sims, I. M., Carnachan, S. M., & Carr, A. J. (2011). Isolation and characterisation of arabinogalactan-proteins from New Zealand kanuka honey. *Food Chemistry*, *128*, 949–956.

Tsumuraya, Y., Mochizuki, N., Hashimoto, Y., & Kovac, P. (1990). Purification of an exo-a-(1→3)-D-galactanase of Irpex lacteus (Polyporus tulipiferae) and its action on arabinogalactan-proteins. *Journal of Biological Chemistry*, *265*, 7207–7215.

van Holst, G. J., & Clarke, A. E. (1985). Quantification of arabinogalactan-protein in plant extracts by single radial gel diffusion. *Analytical Biochemistry*, *148*, 446–450.

Westphal, Y., Kuehnel, S., Schols, H. A., Voragen, A. G. J., & Gruppen, H. (2010). LC/CE-MS tools for the analysis of complex arabino-oligosaccharides. *Carbohydrate Research*, *15*, 2239–2251, 345

Yariv, J., Rapport, M. M., & Graf, L. (1962). The interaction of glycosides and saccharides with antibody to the corresponding phenylazoglycosides. *Biochemical Journal*, *85*, 383–388.

Zippel, J., Deters, A., Pappai, D., & Hensel, A. (2009). A high molecular arabinogalactan from *Ribes nigrum* L.: Influence on cell physiology of human skin fibroblasts and keratinocytes and internalization into cells via endosomal transport. *Carbohydrate Research*, *344*, 1001–1008.

Zippel, J., Wells, T., & Hensel, A. (2010). Arabinogalactan protein from *Jatropha curcas* L. seeds as inductor of keratinocyte *in vitro* differentiation and stimulation of GM-CSF, HGF, KGF and TGFβ1 in organotypic skin equivalents. *Fitotherapia*, *81*, 772–778.