Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 341 (2006) 2370–2378
Carbohydrate RESEARCH

Glycan analysis of recombinant *Aspergillus niger endo*-polygalacturonase A

Bryan D. Woosley,[a,b] Young Hwan Kim,[a,b] V. S. Kumar Kolli,[a,b] Lance Wells,[a,b] Dan King,[c] Ryan Poe,[c] Ron Orlando[a,b] and Carl Bergmann[a,b,*]

[a]*Complex Carbohydrate Research Center, Department of Biochemistry and Molecular Biology, University of Georgia, 315 Riverbend Road, Athens, GA 30602-4712, USA*
[b]*Complex Carbohydrate Research Center, Department of Chemistry, University of Georgia, 315 Riverbend Road, Athens, GA 30602-4712, USA*
[c]*Department of Chemistry, Taylor University, 236 Reade Avenue, Upland, IN 46989, USA*

Received 10 March 2006; received in revised form 1 June 2006; accepted 6 June 2006
Available online 18 July 2006

Abstract—The enzyme *endo*-polygalacturonase A, or PGA, is produced by the fungus, *Aspergillus niger*, and appears to play a critical role during invasion of plant cell walls. The enzyme has been homologously overexpressed in order to provide sufficient quantities of purified enzyme for structural and biological studies. We have characterized this enzyme in terms of its post-translational modifications (PTMs) and found it to be both N- and O-glycosylated. Additionally, we have characterized the glycosyl moieties using MALDI-TOF and LC–ESI mass spectrometry. The characterization of all PTMs on PGA, along with molecular modeling, allows us to reveal potential roles played by the glycans in modulating the interaction of the enzyme with other macromolecules.

Keywords: Mass spectrometry; Fungus; Glycosylation; Post-translational modification; Molecular modeling

1. Introduction

endo-Polygalacturonase A (PGA), an enzyme secreted by *Aspergillus niger*, is one of the pectin-degrading enzymes (PDEs) that are involved in the degradation of plant cell-wall materials through the degradation of the homogalacturonan part of the pectin network.[1] PDEs are important pathogenicity factors, but they also have a number of industrial uses including extensive applications in the food industry.[2–4] Among the seven *A. niger* EPGs, all of which have been expressed, PGI and PGII are the two most extensively characterized with regard to structure and functionality.[5–7] To understand the varying properties of EPGs, the other *A. niger* isozymes are being fully characterized. PGA is of special interest because it and PGB are constitutively expressed, in contrast with the major PGs expressed by *A. niger*, which must be induced by the presence of pectin.[1] While the exact role of PGA during pathogenesis is unknown, PGA and PGB are likely the enzymes required during the early stages of pathogenicity and are thus of critical importance to successful attack.

Many of the PDEs produced by fungi have been identified as being glycosylated. N-Linked glycosylation in *A. niger* is of the high-mannose type.[8–10] There are reports for the species *Aspergillus awormi* of *O*-mannose structures containing two, three, and four mannose residues,[11] but O-linked structures reported to date in *A. niger* contain only single mannose residues.[12–14] The glycosylation state of the various PDEs may play important roles in pathogenesis, as there is growing evidence that glycosylation can have dramatic effects on the structure and function of proteins, as well as on protein–protein interactions.[15,16] Unfortunately, the effects of the carbohydrate side chains on the properties of most wild-type and recombinant PDEs are not known. In

* Corresponding author. Tel.: +1 706 542 4487; fax: +1 706 542 4412; e-mail: cberg@ccrc.uga.edu

0008-6215/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2006.06.006

addition, the various conditions and hosts that are chosen for overexpression when such enzymes are used for commercial or basic research purposes may induce variation in the post-translational modifications of the recombinant proteins. Therefore, a thorough understanding of the glycosylation of each PDE is essential, and the enzymes resulting from overexpression must be characterized to ensure the validation of the product.[4]

The complete characterization of a glycoprotein requires the identification of all potential glycosylation sites, and the characterization of any carbohydrate at each site including the type and number of residues present. This is complicated by the heterogeneity of the carbohydrate portion of a glycoprotein, as multiple carbohydrates can be found at any single site of glycosylation. No single mass spectrometry (MS) technique is able to determine all of these parameters, but several techniques can be used in combination to deduce the complete structure. A three-step process is typically employed in the characterization of the N-linked carbohydrate chains of glycoproteins. First, the molecular weight of the intact glycoprotein is determined. Second, the peptide backbone of the glycoprotein is cleaved enzymatically, and the resulting glycopeptides are identified and isolated. Third, a series of *endo*- and *exo*-glycosidase digestions are performed to determine the primary structure of the carbohydrate side chains.

Matrix-assisted laser desorption/ionization (MALDI) and electrospray-ionization (ESI) mass spectrometry provide a rapid and convenient method for the characterization of glycopeptides and the oligosaccharide side chains released from a glycoprotein.[10,17,18] MALDI-MS has been reported to have several advantages over other mass spectrometric ionization techniques, particularly its high sensitivity (low pmol to fmol quantities) and compatibility with most biological salts and buffers. ESI combined with HPLC can be used to isolate and identify the glycopeptides out of the mixture resulting from the tryptic digestion of glycoproteins. This is accomplished by scanning for the characteristic carbohydrate fragment ions.[19]

The identification and characterization of any O-linked carbohydrate proves to be more challenging because there is no single enzyme capable of releasing a wide variety of *O*-glycans. Nonetheless, techniques such as β-elimination, followed by Michael addition with DTT (BEMAD), which was originally developed for mapping O-GlcNAc sites, have also recently been demonstrated to be useful for mapping other types of *O*-glycans, as would be expected based on the chemistry employed.[12,20] LC–nano ESIMS with stepped-orifice potential scanning has also proven useful for identifying specific sites of O-glycosylation.[12]

In this work, we have used a combination of MALDI-MS and LC–ESIMS with enzymatic digestion to identify the sites of N- and O-glycosylation. Additionally, detailed information on the structures and sequences of the O- and N-linked carbohydrate chains attached to the recombinant PGA from *A. niger* has been provided. A characterization of the glycosylation of PGA is thus obtained.

2. Results

2.1. N-Linked glycopeptide and N-linked glycan structure identification

The MALDI-TOF mass spectrum taken of recombinant PGA from *A. niger* (Fig. 1a) revealed that the experimental molecular mass of 37,226 Da was much larger than the theoretical molecular mass of 35494.2 Da calculated from the amino acid sequence.[1] The 1730 Da difference was suspected to be due to the presence of glycosylation. Although potential N-linked glycosylation sites can be identified by the consensus sequence of Asn-X-Ser/Thr (where X is any amino acid other than Pro), there is no consensus sequence for O-linked glycosylation. Consequently, an examination of the amino acid sequence of PGA showed one potential site for N-linked glycosylation at Asn214, but O-linked glycosylation can potentially occur at any serine or threonine residue. The MALDI mass spectrum of the deglycosylated PGA produced by on-probe *endo*-glycosidase H (Endo H) digestion on the MALDI target indicates an experimental molecular mass of 36,203 Da as shown in Figure 1b. Endo H is an enzyme specific for the glycosidic bond between the two *N*-acetylglucosamine residues of a high-mannose N-linked carbohydrate core

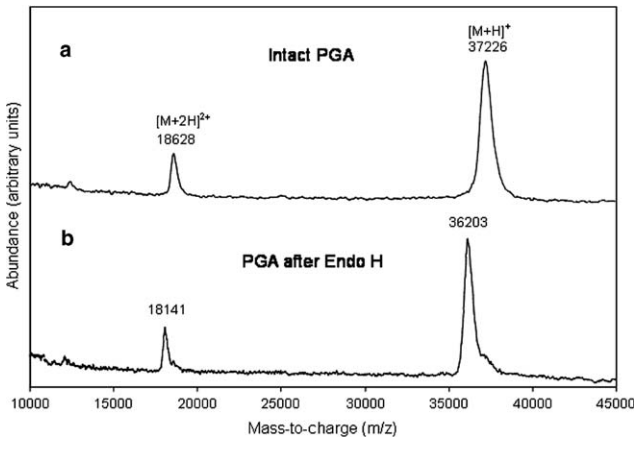

Figure 1. The MALDI mass spectra of (a) intact PGA and (b) deglycosylated PGA. The mass shift following Endo-H treatment indicates that PGA contains an N-linked carbohydrate with a mass of 1023 Da. The resulting protein mass of 36,203 Da indicates that there are other PTM's on PGA (containing an additional 505 Da) as the theoretical protein mass is 35,494 Da. The $[M+2H]^{2+}$ peak shown lies within the mass accuracy of the instrument.

structure, and a successful enzymatic cleavage results in a single *N*-acetylglucosamine residue (+203 Da) attached to the peptide backbone. The shift in mass of 1023 Da between the intact and Endo-H treated PGA is consistent with a high-mannose N-linked structure. From the mass of the Endo-H treated PGA, it can be concluded that the protein lacking *N*-glycan modification has a molecular mass of 35,999 Da. However, a difference of 505 Da between this mass and the theoretical value of 35494.2 Da is still present and indicates that other modifications are likely present on PGA.

To obtain information on the structure of the N-linked carbohydrate moiety, MALDI mass analysis was employed. Generally, the *N*-glycan chain was heterogeneous and resulted in a series of peaks with a mass shift of *m*/*z* 162 (Fig. 2), which corresponded to the sodium adducts of six glycoforms having molecular masses of 1098, 1259, 1421, 1583, 1745, and 1907 Da. These peaks corresponded to the masses expected of the $[M+Na]^+$ ions of a high-mannose N-linked glycan series, $Man_{4-9}GlcNAc_2$, with the predominant species containing five mannoses.

We compared the MALDI-TOF data from proteolytic digests of the glycoprotein before and after treatment with peptide-*N*-glycosidase F (PNGase F) to identify putative N-linked glycopeptides. The MALDI-TOF mass spectra of the tryptic digests of intact PGA and PGA following PNGase F digestion (Fig. 3a and b, respectively) showed that the two peaks observed at *m*/*z* 3058 and 3730 prior to enzymatic cleavage with PNGase F completely disappeared, and two new peaks appeared at *m*/*z* 1841 and 2513. According to the theoretical masses of the tryptic digest fragments listed (Table 1), *m*/*z* 1840 and 2512 corresponded to T14 and T13–14, respectively. (The single mass unit difference between the predicted and observed mass is because of the transformation of Asn to Asp that occurs with the removal of the N-linked carbohydrate side chain by PNGase F.) This result was consistent with the presence of an N-glycosylation consensus sequence found in the T14 tryptic peptide. The difference between the molecular weights of the glycosylated and deglycosylated tryptic peptides at T14 and T13–14 indicated that T14 was modified with an *N*-glycan consisting of $GlcNAc_2Man_5$ (+1216 Da). This was consistent with the results shown in Figures 1 and 2 and was also supported by the MALDI-MS analysis of the LC fractions eluting at 23.2 and 23.8 min (Figs. 4a and 5a, respectively), which were determined to be glycosylated T14 and T13–14.

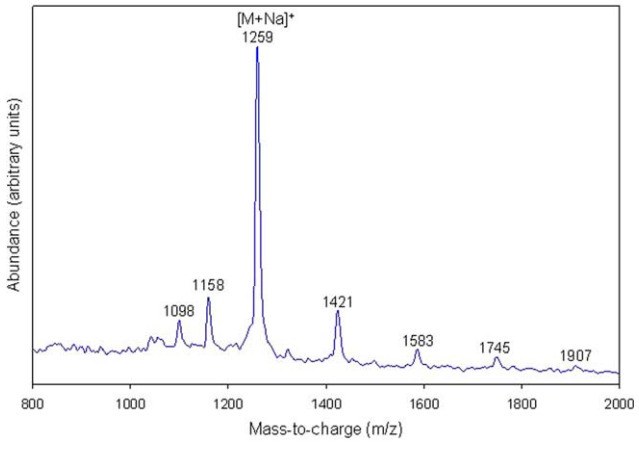

Figure 2. N-Linked glycan distribution of PGA released by PNGase F digestion. The heterogeneous *N*-glycan chain results in a series of peaks with a mass shift of *m*/*z* 162 corresponding to the sodium adducts of six glycoforms having masses of 1098, 1259, 1421, 1583, 1745, and 1907 Da. The peak at *m*/*z* 1259 corresponds to the expected mass of the $[M+Na]^+$ ion of a $GlcNAc_2Man_5$ N-linked glycan.

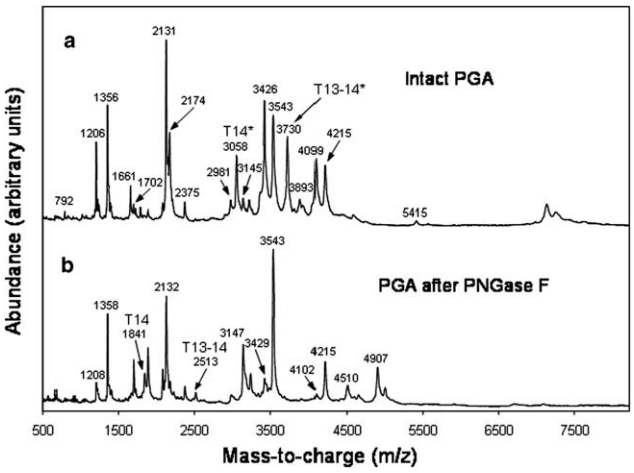

Figure 3. The MALDI mass spectra of tryptic digests of (a) intact PGA and (b) PGA after PNGase F digestion. T14* and T13–14* are glycopeptides, which contain an N-linked glycosylation site at Asn^{214}. T14 and T13–14 are deglycosylated peptides.

On-probe glycosidase experiments combined with MALDI-MS were performed on the LC fractions eluting at 23.2 and 23.8 min in order to obtain further detailed information on the structure of the N-linked glycan. Each fraction was digested with Endo-H, and the MALDI-TOF mass spectra (Fig. 4b) did not contain any of the six original species found in the LC fractions. Only one molecular ion was seen in each of the T14 and T13–14 fractions (*m*/*z* 2044 and 2717, respectively). Each of these ions had an *m*/*z* equal to the mass of the corresponding tryptic peptide, plus the mass of a single *N*-acetylglucosamine residue as seen previously, and confirmed that Asn214 was the site of N-glycosylation. The digestion of the T14 glycopeptide with α-mannosidase, an enzyme that cleaves α-linked mannose residues from the non-reducing termini of carbohydrate structures, yielded peaks at *m*/*z* 2406, 2568, and 2730 corresponding to the loss of four, three, and two mannose residues, respectively, from an ion species of *m*/*z* 3056 (Fig. 4c). In the case of α-mannosidase digestion of the T13–14 glycopeptide, the same result was also obtained (data not shown), confirming the

Table 1. Theoretical and observed (Fig. 3a) masses of tryptic peptides of intact PGA identified in the MALDI mass spectrum

Peptide fragment	AA residue #'s	Sequence	Calculated $[M+H]^+$	Observed $[M+H]^+$
T10	113–119	ISGLYIK	793.5	792
T7	87–97	WWDGEGTNGGK	1206.5	1206
T9	102–112	FFYAHDLDDSK	1357.6	1356
T7–8	87–101	WWDGEGTNGGKTKPK	1660.8	1661
T6	69–86	DITVTQSSDAVLDGNGAK	1790.9	1791
T14*	214–230	NVTFIDSTVSDSENGVR	1839.9	3059
T9–10	102–119	FFYAHDLDDSKISGLYIK	2132.1	2131
T5–6	65–86	ISGKDITVTQSSDAVLDGNGAK	2176.1	2174
T13–14*	208–230	DDNTVKNVTFIDSTVSDSENGVR	2512.2	3731
T6–7	69–97	DITVTQSSDAVLDGNGAKWWDGEGTNGGK	2978.4	2981
T4	38–64	DLNDGTTVTEFGTTTWEYEEWDGPLLR	3145.4	3145
T12	173–207	NQDDCVAINSGENIYFSGGTCSGGHGLSIGSVGGR	3428.5	3426
T12–13	173–213	NQDDCVAINSGENIYFSGGTCSGGHGLSIGSVGGRDDNTVK	4100.8	4099

$[M+H]^+$ is the average molecular mass. T14* and T13–14* are glycopeptides, which contain N-linked glycosylation site at Asn^{214}.

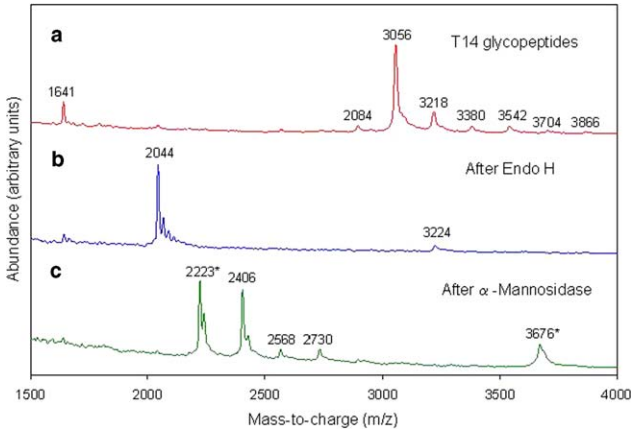

Figure 4. MALDI mass spectrometric analysis of (a) the T14 glycopeptides from PGA, (b) the glycopeptides after Endo-H digestion, and (c) the glycopeptides after α-mannosidase digestion. Peaks marked with an asterisk represent ions found in the enzyme control. The *m*/*z* values of 3056 and 2044 Da correspond to the T14 peptide containing the N-linked glycan and the T14 peptide plus the mass of a single *N*-acetylglucosamine residue after treatment with Endo-H, respectively. The mass discrepancy observed at 3056 Da for the glycan and at 2044 for the deglycosylated T14 tryptic fragment containing the *N*-acetylglucosamine residue (compared to the previous experimental mass of 3058.6 Da shown in Figure 2a and the calculated value of 1839.9 Da for the T14 tryptic peptide, respectively) is due to the limited mass accuracy of the instrument.

composition of the glycans. Thus, the N-linked glycan attached at PGA Asn214 contained a high-mannose structure of $GlcNAc_2Man_{4-9}$ similar to findings for other *endo*-polygalacturonase enzymes.[9,10]

2.2. O-Linked glycopeptide identification and O-linked site mapping

The trypsin-digested peptide mixtures were separated on a microbore C_{18} column that was interfaced with a conventional nanospray source. The Q-TOF parameters were optimized to accommodate the flow rate of 2 μL/min (vs 1 μL/min as used in traditional nanospray techniques). Mass spectra corresponding to each ion peak in the total ion chromatogram were obtained and compared to the expected tryptic peptide masses calculated from the in silico digest. Beginning at an elution time of 47.6 min, ion peaks began showing multiple peaks spaced at intervals of *m*/*z* 81, indicating the possible presence of *O*-mannose on the peptide as expected. The mass spectra obtained from the effluent at 48.0 min showed a dominant ion at *m*/*z* 1194.5 and a series of peaks with much weaker signals in the lower mass region in addition to weaker peaks in the higher mass region also spaced at intervals of *m*/*z* 81. The weaker series of ions evident in the mass spectrum were the result of incomplete carboxyamidomethylation. Thus, the ion peak at *m*/*z* 1194.5 corresponded to the doubly charged T1 peptide backbone (STCTFTDAATASESK, 1519.7 Da) after carboxyamidomethylation of the Cys3 residue (+57 Da) and with *O*-glycan modification equivalent to five hexose modifications. Similarly, the adjacent ion peaks eluting at 49.4 and 50.7 min had dominant ion *m*/*z* values of 1113.5 and 1031.9 Da, respectively, and showed the same spacing of *m*/*z* 81 corresponding to *O*-glycan modifications equivalent to four hexose modifications at 49.4 min (Fig. 6) and three hexose modifications at 50.7 min (Fig. 5). Analysis of the major ion peaks by MS/MS confirmed the preliminary conclusion that this was the T1 peptide containing *O*-glycan modification as shown for the parent ion having a *m*/*z* value of 1113.5 (Fig. 6).

The additional mass due to the addition of five *O*-hexose modifications explained the mass previously unexplained in the *N*-glycan analysis. Although the glycoform containing three hexoses matched very closely (486 Da) with the MALDI data (which show the addition of 505 Da to the protein mass), the low ion intensity (as shown in Fig. 5) for glycoforms with greater than three hexose residues indicated that a modification of three hexoses is the major contributor to the extra mass as seen in the MALDI spectrum.

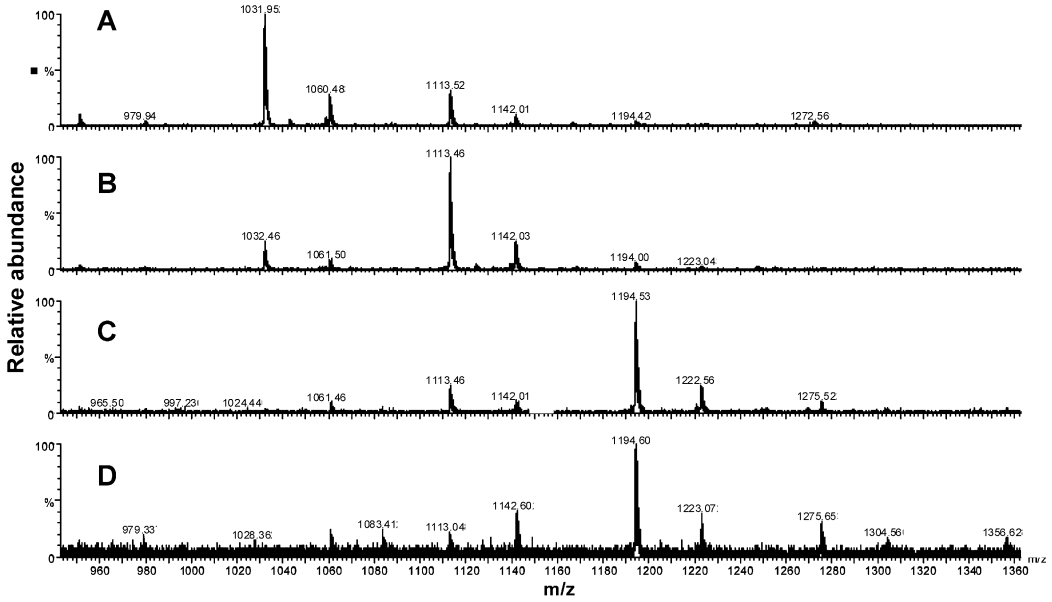

Figure 5. Mass spectrum of peptides eluting at (A) 50.71, (B) 49.44, (C) 48.01, and (D) 47.64 min. Peptides eluting at later retention times show lower m/z ratios, and in all spectra the major peaks are spaced by 81 m/z units, the expected shift for doubly charged peptides containing *O*-hexose. The fact that the smaller masses elute later follows the pattern that glycopeptides generally elute before their respective non-glycosylated peptides, and the data show this as the intensities of more massive peaks diminish as a function of longer elution time.

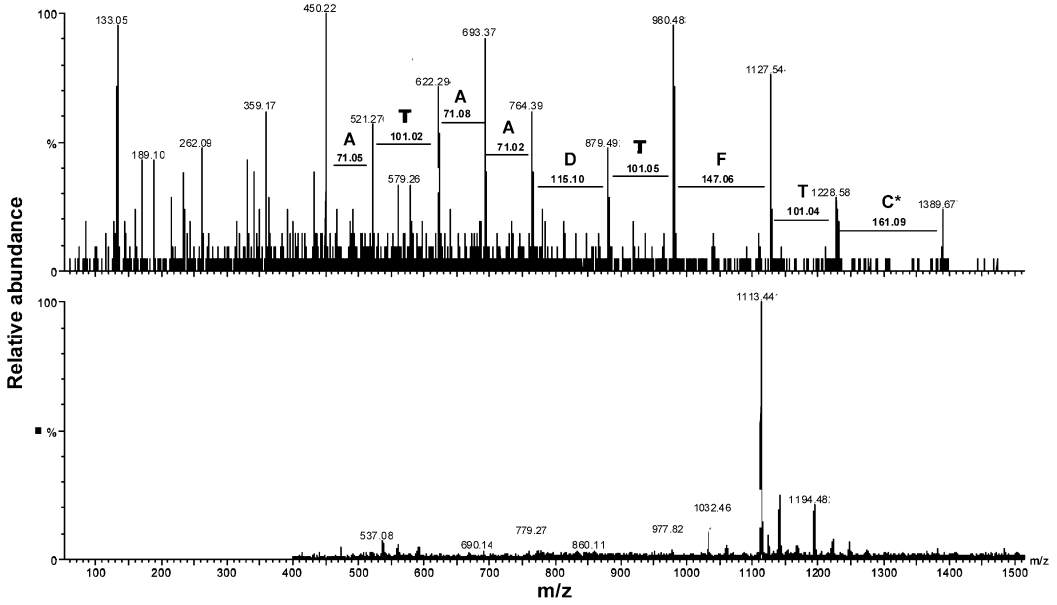

Figure 6. MS/MS analysis (upper trace) of $[M+H]^{2+}$ glycopeptide 1113.5 (lower trace) identifying the amino acid sequence to be T1 of PGA (1577.8 Da) plus four *O*-hexose units (648 Da) resulting in the parent ion mass of 2226 Da. The other peptide fragments mentioned in Figure 5 were identified in the same fashion, and all of the peptides were verified to contain the sequence consistent with PGA T1 as presumed.

Following the confirmation of five hexose residues on the T1 peptide, other glycopeptides were analyzed in a similar manner. Peptides eluting at 47.6 and 48.0 min showed evidence of weak ions having m/z values of 1275.6 and 1356.7 Da. These two ions showed the same spacing of m/z 81 as many of the other ions eluting in this window, and they also corresponded to the doubly charged T1 peptide backbone after carboxyamidomethylation containing *O*-glycan modifications equivalent to six and seven hexose modifications, respectively. The elution pattern of the glycopeptides followed what would be expected for a peptide with heterogeneous

modifications. Thus, glycoforms containing the highest degree of glycosylation eluted before glycoforms with lesser degrees of glycosylation or peptides of the same sequence containing no modification. Although the intensity of the ion peaks at these higher levels of glycosylation were too low to allow for sequencing using MS/MS techniques, all evidences pointed to the fact that there is some percentage of the glycoform containing the T1 peptide with seven hexose modifications equal to the number of serine and threonine residues in the peptide. While the data did not absolutely preclude the possibility of a mannose chain at one single site, the presence of single mannose residues at each serine or threonine was consistent with previous results obtained by X-ray crystallography for *Aspergillus* enzymes, as well as analysis based on BEMAD and HPAEC methodologies for the *A. niger* isoform PGC.[12–14] No other peptides showed evidence of glycosylation when analyzed by this method.

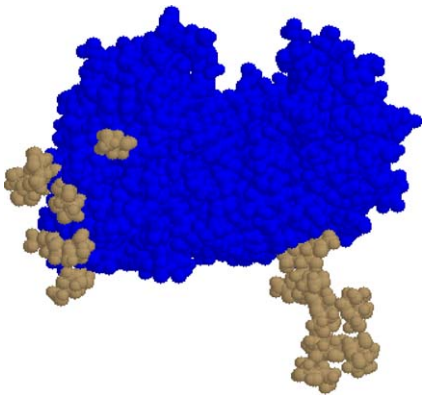

Figure 8. Threaded representation of *O*-mannose and N-linked $GlcNAc_2Man_5$ glycosylation (both shown in tan) of PGA.

3. Discussion

An N-linked high-mannose carbohydrate structure of PGA from *A. niger* was identified using MALDI-MS comparative mapping, HPLC–ESIMS techniques, and on-probe glycosidase digestion. The presence of an *N*-glycan at Asn214, $GlcNAc_2Man_5$, increased the mass of the protein by 1236 Da from the calculated theoretical mass. We also detected the presence of up to seven *O*-hexose modifications on the N-terminal peptide (Fig. 7). Their identification as single mannose residues was consistent with the identification of individual O-linked mannose residues on the PGC isoform.[12] It was evident that the glycoform with three hexose modifications was the major contributor to the additional mass, adding a further 486 Da to the theoretical mass. This resulted in a calculated mass of 37,216 Da, accounting for 99.97% of the observed mass of 37,226 Da for intact PGA (Fig. 1a), well within the mass accuracy (1000 ppm) of the MALDI instrument at this molecular weight.

The locations of all post-translational modifications of PGA are shown (Fig. 8). The locations of *O*-mannose modification are of particular interest in understanding the biological role of this enzyme. As mentioned earlier, enzymes such as PGA are thought to be released by the fungus in the initial phases of pathogenesis in an attempt to degrade the plant cell wall. The plant responds to this by the production of proteinaceous inhibitors known as polygalacturonase-inhibiting proteins (PGIPs).[21] The mannoses are placed in the region where, according to one model, the PGIP is proposed to interact with the EPG.[22] O-Linked mannose has been shown to be important in protein–protein interactions.[15,16] A systematic site mutagenesis of the mannose containing serines and threonines will allow us to understand what roles each of the mannoses play, if any, in EPG–PGIP interactions and will lead us to new strategies to promote plant resistance of pathogens.

1	11	21	31	41
* * * * *	* *			
STC**T**F**T**DAA**T**	A**S**E**S**KTSCSD	IVLKDITVPA	GETLNLKDLN	DGTTVTFEGT
TTWEYEEWDG	PLLRISGKDI	TVTQSSDAVL	DGNGAKWWDG	EGTNGGKTKP
KFFYAHDLDD	SKISGLYIKN	TPVQAISVES	DNLVIEDVTI	DNSDGDSEGG
HNTDGFDISE	STYITITGAT	VKNQDDCVAI	NSGENIYFSG	GTCSGGHGLS
	@			
IGSVGGRDDN	TVK**N**VTFIDS	TVSDSENGVR	IKTVYDATGT	VEDITYSNIQ
LSGISDYGIV	IEQDYENGDP	TGTPSNGVTI	SDVTLEDITG	SVDSDAVEIY
ILCGDGSCSD	WTMSGIDITG	GETSSDCENV	PSGASCDQ	

Figure 7. The amino acid sequence of the mature PGA protein (which begins at Ser33 of the preproprotein) containing an N-linked glycosylation site at Asn214 and seven O-linked sites. Site mapping of glycosylation is indicated with larger, bold letters at the site of modification and is designated N-linked glycosylation or O-linked glycosylation by the symbols @ and *, respectively.

4. Experimental

4.1. Recombinant PGA sample

The *A. niger* pgaA gene under control of the glycolytic pkiA promoter was used to prepare recombinant PGA in *A. niger*[1] and was a kind gift from Dr. Jaap Visser (Wageningen Agricultural University, Wageningen, The Netherlands). The purified recombinant PGA (2.2 mg/mL) was stored in NaOAc buffer (pH 5.0) and kept frozen until further use. Endo-H from *Streptomyces plictus* and PNGase F from *Chryseobacterium meningosepiticum* were purchased from Prozyme (San Leandro, CA, USA). α-Mannosidase from *Canavalia ensiformis* was obtained from Boehringer-Mannheim (Indianapolis, IN, USA). Sequencing grade trypsin was from Promega (Madison, WI, USA), 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), 2,5-dihydroxybenzoic acid (DHB), and α-cyano-4-hydroxycinnamic acid (α-CHC) were purchased from Sigma–Aldrich (St. Louis, MO, USA).

4.2. MALDI-TOF mass spectrometry

The MALDI-TOF MS analysis was carried out on a Hewlett–Packard (Palo Alto, CA, USA) G2025A time-of-flight mass spectrometer. A nitrogen laser (λ 337 nm) was used to ionize the sample. The instrument operated in the positive mode with an accelerating voltage of 28 kV, an extractor voltage of 7 kV, and a pressure below 10^{-6} Torr. A mixture of standard proteins was used as an external standard to calibrate the instrument prior to sample analysis. Sinapinic acid, DHB, and α-CHC were used as the matrices, and the instrument was calibrated externally with mixtures of known peptides and proteins.

4.3. On-probe glycosidase digestion

Intact recombinant PGA(0.5 μL) or HPLC-isolated tryptic peptides were applied to the MALDI target. Endo-H (0.3 μL, 0.05 U/mL) or α-mannosidase (1.25 U/mL) with 0.5 μL of water was deposited over the region that contained the sample. The digestion was allowed to proceed for 15 min at room temperature. The sample region was monitored carefully to ensure that the sample remained solvated during the entire digestion period. After incubation, 0.5 μL of sinapinic acid (or α-CHC for peptides) was added to each digestion mixture, and the mixture was dried. The digestion products were then analyzed by MALDI-TOF MS.

4.4. Deglycosylation of recombinant PGA by PNGase F

Approximately 40 μg of PGA was added to 10 μL of 250 mM sodium phosphate (pH 7.5) and heated at 100 °C for 5 min. The sample was cooled to room temperature before adding 2 μL of PNGase F (5.0 U/mL) and then incubated overnight at 37 °C. The glycans were separated from the intact deglycosylated PGA using a 10,000 MW cut-off micron filter. A 0.5-μL aliquot of this reaction mixture was subjected to MALDI-TOF MS analysis in order to obtain the N-linked glycan profile of PGA. The deglycosylated PGA was then buffer exchanged into 40 mM ammonium bicarbonate and 8 M urea.

4.5. Trypsin digestion of recombinant PGA

Recombinant PGA (80 μg) was dissolved in 40 mM NH_4HCO_3 and 8 M urea. This intact PGA and the deglycosylated PGA prepared above had their disulfide bonds reduced with 10 mM dithiothreitol at 55 °C for 1 h, and the resulting free –SH groups were then subjected to carboxyamidomethylation with 120 μL of 55 mM iodoacetamide. This reaction was performed in the dark at room temperature for 45 min with vortexing every 15 min. Lys-C (1 μg) was added to each sample and allowed to digest at 37 °C for 16 h with gentle shaking. Sufficient 40 mM NH_4HCO_3 was then added to the mixture to dilute the concentration of urea to 1 M. Trypsin digestion was performed by the addition of sequencing grade trypsin at a ratio of 1:25 (w/w enzyme to substrate), and the digestion was carried out at 37 °C for 24 h with gentle shaking. The resulting peptides were desalted using a C_{18} reversed-phase spin column (The Nest Group, Inc., Southborough, MA, USA) and eluted with 0.1% TFA (v/v) and 75% CH_3CN (v/v) in H_2O. The resulting fractions were then evaporated to dryness using a centrifugal evaporator. Samples were resuspended in water and an aliquot of each was subjected to MALDI-TOF analysis.

4.6. HPLC separation of the tryptic digest of recombinant PGA

Approximately 15 μg of the tryptic digest of recombinant PGA was separated on a 50 × 1.0 mm LUNA 5 μm C_{18} column (Phenomenex, Torrance, CA, USA) using a Hewlett–Packard (Palo Alto, CA, USA) HP 1100 series system. The mobile phases for gradient elution were 0.1% HCO_2H (v/v) in H_2O (eluent A) and 0.1% HCO_2H (v/v) in an 80% CH_3CN–water (v/v) mixture (eluent B). Following a 2 min isocratic flow of 5% eluent B, separation was achieved by increasing the concentration of eluent B from 5% to 40% over a 48 min gradient before increasing the concentration of eluent B from 40% to 100% over a 10 min gradient. A flow rate of 100 μL/min was used for the separation. The effluent from the column was monitored at 214 nm by a UV detector. Fractions were collected every 2 min during the gradient and later analyzed by MALDI-TOF MS and on-probe glycosidase digestion.

4.7. LC–nano ESIMS with stepped-orifice potential scanning technique

A trypsin-digested PGA sample (10 μg) was injected onto a 0.3×150 mm C_{18} capillary column (LC Packings, Netherlands). Reversed-phase separation was achieved by mixing 0.1% HCO_2H (v/v) in MQ-H_2O (eluent A) and 0.1% HCO_2H (v/v) in 80% CH_3CN–water (eluent B) from 10% to 55% eluent B at a rate of increase of 0.8% eluent B/min. A Waters CapLC system delivered the solvents at a flow rate of 1 μL/min. The effluent was sprayed into a hybrid quadrupole/time-of-flight instrument equipped with a nanospray source (Q-TOF-2™, Micromass, Milford, MA, USA) operating in the positive-ion mode. A voltage applied to the orifice of the instrument, the cone voltage, creates a potential difference between the sampling cone and extraction cone on the Q-TOF. Two separate scan functions were programmed so that, depending on the cone voltage setting (30 and 70 eV, respectively), the result was either an intact ionized species or a species fragmented within the ionization source. A detector range suitable for both peptides and glycopeptides, m/z 500–2300, was selected for the 30-eV cone voltage setting. A detector range of m/z 50–400 was selected for the 70-eV cone voltage setting to monitor the ions observed at m/z values of 163 (hexose) and 204 (*N*-acetylhexosamine) resulting from fragmentation of carbohydrate-containing species.[6,23,24] The instrument was set to collect data for both voltages by alternating between the two cone voltage settings and scanning 1 s for each. The glycopeptides were identified in the low-energy condition based on the detection of the carbohydrate fragments generated during the corresponding high-energy condition. Collision-induced dissociation (CID) MS/MS was then performed to provide data for peptide sequencing from the resulting peptide fragments. The ESI voltage was set at 3100 V, and the desolvation temperature was 225 °C. For these experiments, the instrument was externally calibrated using 1.5 pmol [Glu^1]-Fibrinopeptide B prior to sample analysis.

4.8. Homology modeling

To create a homology model for PGA, the mature amino acid sequence of PGA[1] was threaded onto the crystal structure of PG2 from *A. niger*[25] using the Swiss PDB Viewer, DeepView v. 3.7. The homology structure was then optimized with molecular mechanics, MM3, using the protein modeling suite, BioMedCache v. 6.1 (BMC). The optimized structure was then allowed to move using molecular dynamics, simulating 300 K for 2 ps, before it was again optimized with molecular mechanics. Single mannose monomers and one high-mannose glycosylation ($GlcNAc_2Man_5$) were created using BMC, and their structures were optimized using the semi-empirical method, PM5. The O-linked α,1-mannose glycans and the N-linked high-mannose glycosylation were attached to the appropriate residues of the optimized PGA model. The glycosylated PGA model was optimized by alternating molecular mechanics and molecular dynamics computations as before.

Acknowledgments

This work is supported in part by the grants from the National Institutes of Health (1 P41 RR018502), the National Science Foundation (MCB-0115132), the Department of Energy (DE-FG02-96ER20221; DE-FG02-93ER20097), and the Muscular Dystrophy Association. The authors express special appreciation to Drs. Jaap Visser and Jacques A. E. Benen (Wageningen University, The Netherlands) for providing the recombinant PGA sample without which this work would not have been possible.

References

1. Pařenicová, L.; Benen, J. A. E.; Kester, H. C. M.; Visser, J. *Biochem. J.* **2000**, *345*, 637–644.
2. Sakai, T.; Sakamoto, T.; Hallaert, J.; Vandamme, E. J. *Adv. Appl. Microbiol.* **1993**, *39*, 213–294.
3. Cooper, R. M. The Mechanisms and Significance of Enzymic Degradation of Host Cell Walls by Parasites. In *Biochemical Plant Pathology*; Callow, J. A., Ed.; Wiley: New York, 1983; pp 101–135.
4. Archer, D. B.; Peberdy, J. F. *Crit. Rev. Biotechnol.* **1997**, *17*, 273–306.
5. Armand, S.; Wagemarker, M. J. M.; Sanchez-Torres, P.; Kester, H. C. M.; van Santen, Y.; Dijkstra, B. W.; Visser, J.; Benen, J. A. E. *J. Biol. Chem.* **2000**, *275*, 691–696.
6. Van Santen, Y.; Benen, J. A. E.; Schroer, K.-H.; Kalk, K. H.; Armand, S.; Visser, J.; Dijkstra, B. W. *J. Biol. Chem.* **1999**, *274*, 30474–30480.
7. Pagès, S.; Heijne, W. H. M.; Kester, H. C. M.; Visser, J.; Benen, J. A. E. *J. Biol. Chem.* **2000**, *275*, 29348–29353.
8. Xie, M.; Bergmann, C.; Benen, J.; Orlando, R. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3389–3397.
9. Colangelo, J.; Licon, V.; Benen, J.; Visser, J.; Bergmann, C.; Orlando, R. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1448–1453.
10. Yang, Y.; Bergmann, C.; Benen, J.; Orlando, R. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1257–1262.
11. Oka, T.; Sameshima, Y.; Koga, T.; Kim, H.; Goto, M.; Furukawa, K. *Microbiology* **2005**, *151*, 3657–3667.
12. Woosley, B. D.; Xie, M.; Wells, L.; Orlando, R.; Bergmann, C. *Anal. Biochem.* **2006**, *354*, 43–53.
13. Van Pouderoyen, G.; Snijder, H. J.; Benen, J. A.; Dijkstra, B. W. *FEBS Lett.* **2003**, *554*, 462–466.
14. Petersen, T. N.; Kauppinen, S.; Larsen, S. *Structure* **1997**, *5*, 533–544.
15. Haltiwanger, R. S.; Lowe, J. B. *Ann. Rev. Biochem.* **2004**, *73*, 491–537.
16. Martin, P. T. *Glycobiology* **2003**, *13*, 55R–65R.

17. Colangelo, J.; Orlando, R. *Anal. Chem.* **1999**, *71*, 1479–1482.
18. Haslam, S. M.; Coles, G. C.; Morris, H. R.; Dell, A. *Glycobiology* **2000**, *10*, 223–229.
19. Carr, S. A.; Huddleston, M. J.; Bean, M. F. *Protein Sci.* **1993**, *2*, 183–196.
20. Wells, L.; Vosseller, K.; Cole, R.; Cronshaw, J.; Matunis, M.; Hart, G. *Mol. Cell. Proteom.* **2002**, *1*, 791–804.
21. Kemp, G.; Stanton, L.; Bergmann, C. W.; Clay, R. P.; Darvill, A.; Albersheim, P. *Mol. Plant–Microbe Interact.* **2004**, *17*, 888–894.
22. King, D.; Bergmann, C.; Orlando, R.; Benen, J. A. E.; Kester, H. C. M.; Visser, J. *Biochemistry* **2002**, *41*, 10225–10233.
23. Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal. Chem.* **1993**, *65*, 877–884.
24. Charlwood, J.; Langridge, J.; Camilleri, P. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1522–1530.
25. Van Santen, Y.; Benen, J. A. E.; Schroer, K.-H.; Kalk, K. H.; Armand, S.; Visser, J.; Dijkstra, B. W. *J. Biol. Chem.* **1999**, *274*, 30474–30480.