

Comprehensive glycan analysis of recombinant *Aspergillus niger* endo-polygalacturonase C

Bryan Woosley¹, Min Xie^{1,2}, Lance Wells, Ron Orlando, Derek Garrison³, Daniel King³, Carl Bergmann^{*}

Complex Carbohydrate Research Center and Department of Chemistry, University of Georgia, Athens, GA 30602, USA

Received 20 December 2005
Available online 21 February 2006

Abstract

The enzyme PGC is produced by the fungus *Aspergillus niger* during invasion of plant cell walls. The enzyme has been homologously overexpressed to provide sufficient quantities of purified enzyme for biological studies. We have characterized this enzyme in terms of its posttranslational modifications (PTMs) and found it to be both N- and O-glycosylated. The glycosyl moieties have also been characterized. This has involved a combination of matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF), liquid chromatography (LC)–ion trap, and LC–electrospray ionization (ESI) mass spectrometries in conjunction with trypsin degradation and β -elimination, followed by Michael addition with dithiothreitol (BEMAD). This is the first demonstration of the ability of BEMAD to map glycosylation sites other than O-GlcNAc sites. The complete characterization of all PTMs on PGC allows us to model them on the peptide backbone, revealing potential roles played by the glycans in modulating the interaction of the enzyme with other macromolecules.

© 2006 Elsevier Inc. All rights reserved.

Keywords: endo-Polygalacturonases; BEMAD; Mass spectrometry; O-Mannose; Neutral loss; Site mapping; Glycosylation

The fungus *Aspergillus niger* secretes a wide variety of plant polysaccharide-modifying enzymes that are capable of degrading plant cell wall polysaccharides, such as pectin and xylan, and that are commonly used in the food industry to clarify juice and isolate the essential oils and pigments from citrus [1–3]. Pectinases are also of interest biologically because they play an important role in the infection process of plants by degrading the plant cell wall and facilitating the entry of pathogens into the plant [4]. The pectic polysaccharides rhamnogalacturonan I

(RGI)⁴ and rhamnogalacturonan II (RGII) are complex branched molecules, whereas homogalacturonan is unbranched and substituted with methyl and acetyl

⁴ Abbreviations used: RGI, rhamnogalacturonan I; RGII, rhamnogalacturonan II; MS, mass spectrometry; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight MS; ESI-MS, electrospray ionization-MS; LC-MS, liquid chromatography-MS; BEMAD, β -elimination followed by Michael addition of dithiothreitol; CID, collision-induced dissociation; ELISA, enzyme-linked immunosorbent assay; BSA, bovine serum albumin; TBS, Tris-buffered saline; PNPP, *para*-nitrophenyl phosphate; TFA, trifluoroacetic acid; EndoH, *endo*-glycosidase H; MQ H₂O, Milli-Q water; Q-TOF, quantitative TOF; MS/MS, tandem MS; AmBic, ammonium bicarbonate; PBS, phosphate-buffered saline; EDTA, ethylenediaminetetraacetic acid; HPAEC-PAD, high-performance anion exchange chromatography with pulsed amperometric detection; BMC, BioMedCache; PTM, posttranslational modification; PNGaseA, peptide N-glycanase A; PNGaseF, peptide N-glycanase F; TIC, total ion chromatogram; GC, gas chromatography; PGIP, polygalacturonase-inhibiting protein.

^{*} Corresponding author. Fax: +1 706 542 4412.

E-mail address: cberg@ccrc.uga.edu (C. Bergmann).

¹ These authors contributed equally to this work.

² Present address: Biotech Analytics and Technology Resources, Baxter Healthcare Corporation, Round Lake, IL 60073, USA.

³ Present address: Department of Chemistry and Biochemistry, Taylor University, Upland, IN 46989, USA.

groups. An array of interacting enzymes is required for a complete degradation of the pectin network [5,6], yet only a fraction of the pectin degrading enzymes have been fully characterized to date.

A. niger is able to use pectin as its sole carbon source for growth in vitro and produces a broad spectrum of polygalacturonases with a variety of patterns of action on homogalacturonan. Currently, seven endo-polygalacturonase encoding genes—PGI, PGII, PGA, PGB, PGC, PGD, and PGE—have been cloned and individually over-expressed. The role of these enzymes in pectin degradation was investigated by analyzing the basic biochemical properties such as the pH optimum and kinetic parameters for each EPG isozyme. A number of studies have analyzed the specific activities of the various *A. niger* EPGs on partially methyl esterified pectins [7–11]. It has been suggested that PGA and PGB are scouting enzymes that help the fungus to sense the presence of pectin by generating low-molecular weight inducers while other EPGs subsequently are expressed. PGII has the highest specific activity and is primarily responsible for generating oligomers of polygalacturonic acid (the non-methyl esterified form of homogalacturonan), and these oligomers are quickly converted to monomers by the combined action of PGI and PGII. PGC and PGE have low specific activities and most likely attack parts of the homogalacturonan, but the exact locations of attack have yet to be identified [12]. A fundamental understanding of the functionalities of these enzymes will be of great value in improving their existing industrial applications and in developing new commercial applications as their commercial potential is fully used in the food and juice processing industries.

All *A. niger* EPGs investigated have been identified as glycoproteins [13–15]. Protein glycosylation has been shown to be important in maintaining protein structure and function and is also important during protein–protein interactions [16,17]. However, the effects (if any) of the carbohydrate substituents on EPGs during interaction with the substrate are not well understood.

In this article, the recombinant *A. niger* isozyme PGC is the subject of study. Its carbohydrate structures and glycosylation site heterogeneities were analyzed by mass spectrometry (MS). Matrix-assisted laser desorption/ionization time-of-flight MS (MALDI-TOF MS) was used to analyze the molecular mass of the intact protein as well as that of its deglycosylated form. The N-linked glycopeptides were identified by electrospray ionization–MS (ESI–MS) by precursor ion scanning using an orifice potential stepping technique [18–20]. The presence of O-linked glycosylation was detected based on carbohydrate heterogeneity observed by liquid chromatography–MS (LC–MS). The sites of O-linked glycosylation on PGC were identified by β -elimination followed by Michael addition with dithiothreitol (BEMAD), a method that replaces the labile glycan with a stable covalently linked dithiothreitol, which can be used for purification using thiol chromatography and is stable

for site mapping by collision-induced dissociation (CID) approaches [21]; and by pseudo-neutral loss scanning [22] on a linear ion trap.

Materials and methods

Recombinant PGC sample

The gene encoding PGC was cloned, sequenced, and expressed in *A. niger* as described previously [23]. The purified recombinant PGC (concentration 2.68 mg/ml) was stored in sodium acetate buffer (pH 5.0) and kept frozen until further use.

ELISA detection of glycosylation

The DIG Glycan differentiation kit (Roche Molecular Biochemical, Mannheim, Germany) was used to detect the carbohydrate present on PGC. The manufacturer's protocol was followed with the exception of adapting the method from a dot-blot assay to an enzyme-linked immunosorbent assay (ELISA). An ELISA plate was used to bind the proteins. First, 0.1 μ g recombinant PGC diluted to 150 μ l in ELISA coating buffer as well as the positive controls (diluted standards provided in the kit) and a negative control (ELISA coating buffer) were loaded onto different spots on the plates and incubated at 4 °C for 24 h. Then 250 μ l of blocking buffer (1% bovine serum albumin [BSA] in Tris-buffered saline [TBS]) was loaded into the sample wells, and the plate was incubated at room temperature for 3 h to ensure complete blocking and eliminate nonspecific binding. Five digoxigenin-labeled lectins with different terminal carbohydrate residue binding specificities were used: GNA, SNA, MAA, PNA, and DSA. After removing the unbound lectins by extensive washing, 200 μ l of alkaline phosphatase-labeled anti-digoxigenin solution was added to each well to form the carbohydrate–lectin–anti-lectin complex. The complex subsequently was detected by the presence of a chromophore generated by the addition of 200 μ l of 1 mg/ml *para*-nitrophenyl phosphate (PNPP) solution.

MALDI-TOF MS of intact and deglycosylated recombinant PGC

The MALDI-TOF MS analysis was carried out on a Hewlett–Packard G2025A time-of-flight mass spectrometer (Palo Alto, CA, USA). A 4 μ g/ μ l solution of sinapinic acid (Aldrich) was used as matrix and was made by dissolving the sinapinic acid in a solution of 50:50 (v/v) acetonitrile and 0.1% trifluoroacetic acid (TFA) in water. The recombinant PGC was mixed with the matrix in a 1:1 ratio and vacuum-dried on the MALDI target.

The deglycosylation was performed directly on the MALDI probe in order to reduce the amount of sample and enzyme required. This was accomplished by combining 0.2 μ l of endo-glycosidase H (EndoH, Prozyme, San Leandro, CA, USA) with 0.5 μ l of recombinant PGC on the

MALDI target [24]. The digestion proceeded at room temperature for 30 min. Milli-Q water (MQ H₂O) was added during the digestion to keep the reaction environment moist. After incubation, 0.5 μ l of sinapinic acid matrix solution was added to the spot and dried for MALDI-TOF analysis.

A nitrogen laser ($\lambda = 337$ 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.

Trypsin digestion of recombinant PGC

For those experiments to determine N-linked structure, recombinant PGC was dissolved in 0.2 M Tris-HCl in 8 M urea, and the disulfide bonds were reduced with 40 mM dithiothreitol at 55 °C for 30 min. The resulting free -SH groups were then subjected to carboxyamidomethylation with 80 mM iodoacetamide. This reaction was performed in the dark at room temperature for 30 min. MQ H₂O was then added to the mixture to dilute the concentration of Tris-HCl to 50 mM and of urea to 2 M. The pH value of the resulting mixture was adjusted to 8.0 with 10 mM HCl. Trypsin digestion was performed by the addition of sequencing-grade trypsin at an enzyme/substrate ratio of 1:25 (w/w), and the digestion was carried out at 37 °C for 24 h with gentle shaking.

LC-nano ESI-MS with stepped orifice potential scanning technique

Approximately 5 pmol of a tryptic digest of recombinant PGC was injected onto a 180 \times 1-mm C18 column (LC Packing, PepMap, 3 μ m particle size). A Waters CapLC system was used to deliver solvents at a rate of 1 μ l/min. The mobile phases for gradient elution were 0.1% formic acid (v/v) in MQ H₂O (eluent A) and 0.1% formic acid (v/v) in methanol (eluent B). The sample was first desalted with a 90% eluent A wash for 20 min, and the separation was achieved by increasing the eluent B concentration from 10 to 80% over a 60-min gradient time. The column effluent was interfaced to a quantitative TOF (Q-TOF) instrument and directly analyzed in the positive ion mode.

The mass spectrometer was a hybrid quadrupole/TOF instrument equipped with a nanospray source (Q-TOF-2, Micromass, Milford, MA, USA). To detect glycopeptides selectively, a stepped orifice voltage technique was used, whereby the mass spectrometer was operated alternatively in both normal- and high-orifice voltage modes. In the high-voltage mode, the orifice voltage was set at a value of 80 V to generate “in-source” fragmentation and produce carbohydrate oxonium ions such as those at *m/z* 163 (hexose) and 204 (*N*-acetylhexosamine). The corresponding mass range of this function was set at *m/z* 50–400. A second function with an *m/z* range of 400–2500 was acquired at an orifice voltage of 35 V to avoid in-source fragmentation.

The instrument was set to collect data for both voltages by alternating between the two orifice 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. CID tandem MS (MS/MS) was performed to provide data for peptide sequencing of the resulting peptide fragments. The ESI voltage was set at 3100 V, and the desolvation temperature was set at 225 °C. The instrument was externally calibrated using 1.5 pmol [Glu¹]-fibrinopeptide B prior to sample analysis.

O-Linked glycopeptide identification and purification

For O-linked site mapping, 80 μ g recombinant PGC was dissolved in 40 mM ammonium bicarbonate (AmBic) and 8 M urea, the disulfide bonds were reduced with 10 mM dithiothreitol at 55 °C for 1 h, and the resulting free -SH groups were then subjected to carboxyamidomethylation with 120 μ l of 55 mM iodoacetamide. This reaction was performed in the dark at room temperature for 45 min with vortexing every 15 min. Then 1 μ g Lys-C was added to the mixture and allowed to digest at 37 °C for 16 h with gentle shaking. After that, 40 mM AmBic was added to the mixture to dilute the concentration of urea to 1 M. The trypsin digestion was performed by the addition of sequencing-grade trypsin at an enzyme/substrate ratio of 1:25 (w/w), and the digestion was carried out at 37 °C for 24 h with gentle shaking. The resulting peptides were desalted using a C18 reverse-phase spin column (Nest Group, Southborough, MA, USA) and eluted with 0.1% TFA (v/v) and 75% acetonitrile (v/v) in MQ H₂O. The resulting fractions were then evaporated to dryness using a SpeedVac evaporator.

Approximately 20 pmol of a tryptic digest of recombinant PGC (as described previously) was subjected to the BEMAD protocol [21]. Briefly, the peptides were resuspended in 1.0% triethylamine (v/v), 0.1% sodium hydroxide (v/v), and 10 mM dithiothreitol and were incubated at 42 °C for 3 h. Peptides were acidified with an equal volume of 1% TFA (v/v) and then desalted using C18 reverse-phase spin columns and dried down in a SpeedVac as in the digestion protocol. Peptides were resuspended in degassed phosphate-buffered saline (PBS) containing 1 mM ethylenediaminetetraacetic acid (EDTA) and were subjected to thiol chromatography (Thiolpropyl Sepharose 6B, Amersham Biosciences, Uppsala, Sweden) as described previously [21]. Dithiothreitol-eluted peptides were desalted and subjected to LC-MS/MS analysis.

A nitrogen bomb was used to load the peptides onto a self-packed 75- μ m \times 10.5-cm C8 reverse-phase column (PicoFrit PicoTip Emitter, New Objective, Woburn, MA, USA) using 5 μ l 120A C8 packing material (Vydac, Hesperia, CA, USA). Peptides were eluted at a flow rate of approximately 400 nl/min using a linear gradient of increasing acetonitrile in 0.1% formic acid (v/v) directly into the nanospray source of a linear ion trap (LTQ, ThermoFinnigan, Waltham, MA, USA). Data-dependent scanning was

performed such that a full mass scan was followed by fragmentation analysis of the eight most abundant peaks with a dynamic exclusion value of 2. For pseudo-neutral loss scanning, MS³ was performed on the resulting ion if one of the two major fragments observed in the MS/MS spectra was equal to a loss of *m/z* 81, 162, or 243 from the parent mass. All data were analyzed using TurboSequest software (ThermoFinnigan). Allowed modifications of −18.00, +136.15, and +162.05 on serine and threonine; +57.02 and +120.15 on cysteine; and +15.99 on methionine were selected in the data analysis. The database searched was the nonredundant *A. niger* database (downloaded from the National Center for Biotechnology Information).

O-Linked carbohydrate composition analysis by high-performance anion exchange chromatography with pulsed amperometric detection

Approximately 200 µg recombinant PGC was digested for 16 h at 37 °C with PNGase-F to release all N-linked carbohydrate. The resulting de-N-glycosylated protein was purified by passing the mixture through a 30-µm × 1-mm C18 column (Oasis HLB). The N-linked carbohydrate was discarded, and the remaining protein fraction was subjected to β-elimination with 1 M sodium borohydride in 50 mM sodium hydroxide at 50 °C overnight to release the O-linked carbohydrates. The treated sample was neutralized with 5% acetic acid (v/v), and salts were removed by treatment with an ion exchange resin (Dowex 50WX8-100). The sample was dried, and the resulting borates were removed by repeated evaporation with methanol. The sample was finally treated with 4 M TFA for 4 h at 100 °C to hydrolyze the oligosaccharides into monomers.

Analysis was performed using Dionex DX500 high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC–PAD) equipped with a GP40 gradient pump, an ED40 electrochemical detector, and a Thermo-Separations AS3500 autosampler using a Dionex CarboPac PA1 (4 × 250-mm) analytical column with amino trap and borate trap columns at a flow rate of 0.5 ml/min [25]. Gradient elution was achieved by setting eluent A (water) at 90% and eluent B (1 M NaOH) at 10%. After an initial hold time of 5 min at this condition, eluent B increased to 70% over a 10-min gradient and was held constant for an additional 10 min to complete the separation. Instrument control and data processing were accomplished using Dionex PeakNet software (version 5.01).

Homology modeling

To create a homology model for PGC of *A. niger*, the mature amino acid sequence of PGC [26] was threaded onto the crystal structure of PG2 from *A. niger* [27] using the Swiss PDB viewer DeepView (version 3.7). The homology structure was then optimized using the protein modeling suite BioMedCache (BMC, version 6.1). A structure optimization was performed using molecular mechanics

MM3. Single mannose monomers and one high-mannose glycosylation (GlcNAc₂Man₁₁) were created using BMC, and their structures were optimized using the semiempirical method PM5. Seven O-linked α-1-mannose glycans were attached to residues Thr2, Thr3, Thr5, Ser7, Ser9, Ser70, and Ser114 of the optimized PGC model. The N-linked high-mannose glycosylation was then attached at Asn220. The structure of the glycosylated PGC model was then optimized using molecular mechanics MM3. The optimized structure was then allowed to move using molecular dynamics, simulating 300 K for 1 ps, before it was again optimized with molecular mechanics.

Results and discussion

Among seven *A. niger* EPGs expressed to date, PGI and PGII are the two most extensively characterized with regard to structure and functionality [28–30]. To understand the different properties of EPGs, other isozymes are being fully characterized. For PGC, the pH range for activity falls between 3 and 5 (with a pH optimum at 4.1) and is similar to that for other EPGs [8]. PGC shows much lower activity when compared with PGI and PGII and prefers the non-methyl-esterified substrate, as indicated by a decrease in activity when the degree of methyl esterification is increased [8]. The role of PGC during pathogenesis is unknown, but studies such as the current one will provide valuable information leading to a full understanding of the effect of glycosylation on EPG activities. There is growing evidence that glycosylation can have dramatic effects on intermolecular recognition and the activity of proteins [16,17]. A complete characterization of the glycosylation of PGC is the subject of the current work.

Lectin and ELISA results

The presence of carbohydrate on PGC was detected by modified lectin blotting. All five of the lectins used have different carbohydrate binding specificities, and only GNA showed a strong positive reaction with recombinant PGC. This indicated the presence of terminally linked mannose resulting from either high-mannose or hybrid-type N-linked glycans. An examination of the amino acid sequence of PGC indicated one potential N-linked site at Asn220. (Amino acid numbering in this work is based on the mature protein that begins at Ala41 of the preproprotein, as shown in Fig. 1.)

MALDI-TOF MS of intact and deglycosylated protein

The molecular mass of intact PGC was determined by MALDI-TOF MS to be approximately 39,600 Da (Fig. 2A), which is approximately 3300 Da larger than the molecular mass calculated from the amino acid sequence. The large mass discrepancy indicated the presence of post-translational modifications (PTMs) on PGC, and the observed broad peak and poor mass resolution in the

1	** * * *	11	21	31	41
ATTCTfSGSE	GASKASKSKT	SCSTIYLSDV	AVPSGTTLDL	SDLNDGTHVI	
FQGETTFGYE	EWEGPLVRV S *	GTDITVEGES	DAVLNGDGSR	WWDGEGGNGG	
KTKPKFFYAH	DLT S TIKSI	YIENSPVQVF	SIDGSTDLTM	TDITVDNTDG	
DTDDLAANTD	GFDIGESTYI	TITGAEIYNQ	DDCVAINSGE	NIYFSASVCS	
GGHGLSIGSV	GGRDDNTV K N@	VTFYDVNVLK	SQQAIRIKTI	YGDTGSVSEV	
TYHEIAFSDA	TDYGIVIEQN	YDDTSKTPTT	GVPIIDFVLE	NIVGTCEDDD	
CTEVIACGD	GSCSDWTWTG	VSVTGGSVSD	DCLNVPSGIS	CDL	

Fig. 1. Amino acid sequence of the mature PGC protein (which begins at Ala41 of the preproprotein) containing an N-linked glycosylation site at Asn220 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 symbol @ or *, respectively.

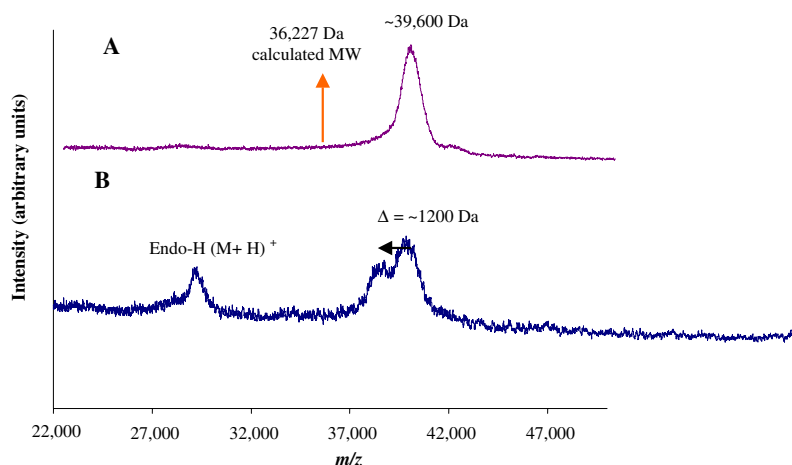


Fig. 2. MALDI-TOF MS spectra of intact PGC protein (A) and the EndoH-digested PGC (B). The molecular weight of intact PGC was approximately 39,600 Da, which is 3300 Da larger the molecular weight calculated from the amino acid sequence. After EndoH on-target digestion for 30 min, a new peak of approximately 1200 Da less appeared. However, the intact protein still was the major peak, suggesting that the digestion was not complete. The additional peak at approximately 29,000 Da shows the EndoH molecular ion peak.

MALDI-TOF spectrum was indicative of heterogeneous glycoforms. Therefore, the intact protein was deglycosylated by EndoH directly on the MALDI target to confirm the presence of N-linked glycosylation. After a 30-min digestion, a new peak appeared at approximately 1200 Da less than the intact protein peak (Fig. 2B), indicating that the recombinant PGC was indeed modified by N-linked glycans having either a high-mannose or hybrid structure. Therefore, Asn220 must be the site of glycosylation given that this is the sole N-linked consensus sequence present in the PGC amino acid sequence. This result was also consistent with the results of the lectin test.

Although N-linked glycans can be released from the protein by either EndoH, peptide *N*-glycanase A (PNGaseA), or peptide *N*-glycanase F (PNGaseF), detection of O-linked glycosylation sites is not as straightforward because there is no single enzyme capable of releasing a wide variety of *O*-glycans. Thus, determination of the presence of O-linked glycosylation using on-target digestion is not feasible.

N-linked glycopeptide and *N*-linked glycan structure identification

The trypsin-digested PGC peptides were analyzed by MALDI-TOF MS and LC-MS. At the higher mass ranges of the MALDI-TOF MS spectrum, a series of signals of moderate intensity were observed with mass differences of 162 Da, a characteristic mass of a peptide modified by a heterogeneous N-linked glycan (data not shown). The 162-Da spacing indicated each peak in this series contained an additional mannose residue, as found in high-mannose glycans, because the monosaccharide residues present within the branches of the other two classes of N-linked structures result in peak spacing other than m/z 162. The composition of the N-linked glycan is consistent with the high-mannose structures that we have previously characterized on other PDEs from *A. niger* using *exo*- and *endo*-glycosidases [31,32]. The experimental peptide masses were then compared with the *in silico* digestion fragments using Masslynx (version 3.4, Micromass) to identify discrepancies

indicative of possible PTM. The mass of the peptide containing the potential N-linked site at Asn220 (T12) is determined to be 1310.7 Da. However, the lack of a peak at m/z 1310.7 Da and the presence of experimental peaks at m/z 2800–3000 Da indicate that Asn220 possibly contains a PTM with an m/z on the order of 1500–1700 Da. The remaining identifiable peaks observed in the MALDI-TOF spectra were identified as unmodified tryptic fragments T6, T11, T13, and T13–14.

The specific structure of the attached glycan was determined by LC-MS on a Q-TOF2 using a stepped orifice potential. Cone voltage settings were optimized based on standard glycopeptide fragmentations observed under variable cone voltage conditions. Although no carbohydrate fragments are observed at a low cone voltage of 35 V, a series of strong carbohydrate fragments can be generated when the cone voltage is increased to 80 V. Approximately 3 pmol of trypsin-digested peptide mixture was analyzed. The selected ion currents for three carbohydrate fragment ions extracted from the high-cone voltage total ion chromatogram (TIC) are shown in Fig. 3. A single ion chromatographic peak centered around 25 min contained all three marker ions, and its summed spectrum is shown in Fig. 4. One series of doubly charged peaks is present from m/z 859.5–1752.0 and shows a peak spacing of m/z 81. The mass for each glycopeptide was calculated and compared with peptide masses calculated from *in silico* digestion, and the composition of the attached glycan was calculated accordingly. For example, the ion peak at m/z 1508.3 corresponds to an $[M+2H]^{2+}$ peptide with a molecular mass of 3014.6 Da. Although this mass is not a match with any of the calculated tryptic peptide masses, the mass difference

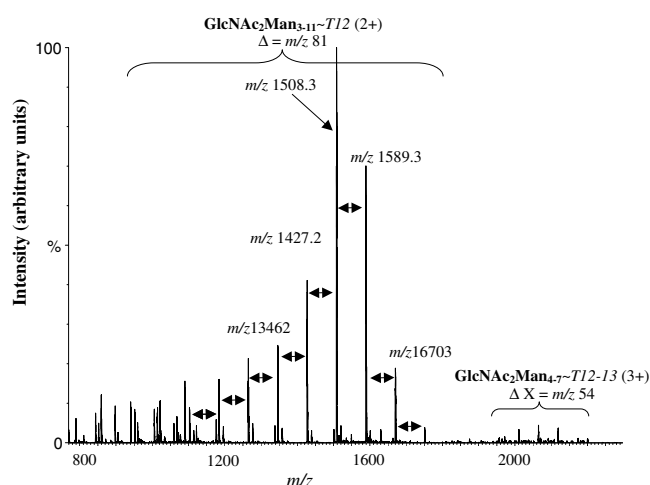


Fig. 4. Averaged spectrum of N-linked glycopeptide. The 2+ charged peaks had incremental spacing of m/z 81 between peaks, and the 3+ charged peaks had incremental spacing of m/z 54. Each species in both spectra was separated by a single hexose residue.

between this ion and T12 (1703.9 Da) corresponds to the mass of a high-mannose glycan, $\text{GlcNAc}_2\text{Man}_8$. Therefore, this ion at m/z 1508.3 is determined to be the doubly charged T12 + $\text{GlcNAc}_2\text{Man}_8$. The rest of the glycopeptides were analyzed in a similar manner, and all were determined to be modifications of the T12 peptide ranging from T12 + $\text{GlcNAc}_2\text{Man}_3$ to T12 + $\text{GlcNAc}_2\text{Man}_{11}$. In addition, a series of triply charged peaks appeared at slightly higher m/z values with a peak spacing of 54, and those peaks correspond to the T12–13 peptide (one missed cleavage) with the glycan structure ranging from T12–13 + $\text{GlcNAc}_2\text{Man}_4$ to T12–13 + $\text{GlcNAc}_2\text{Man}_7$.

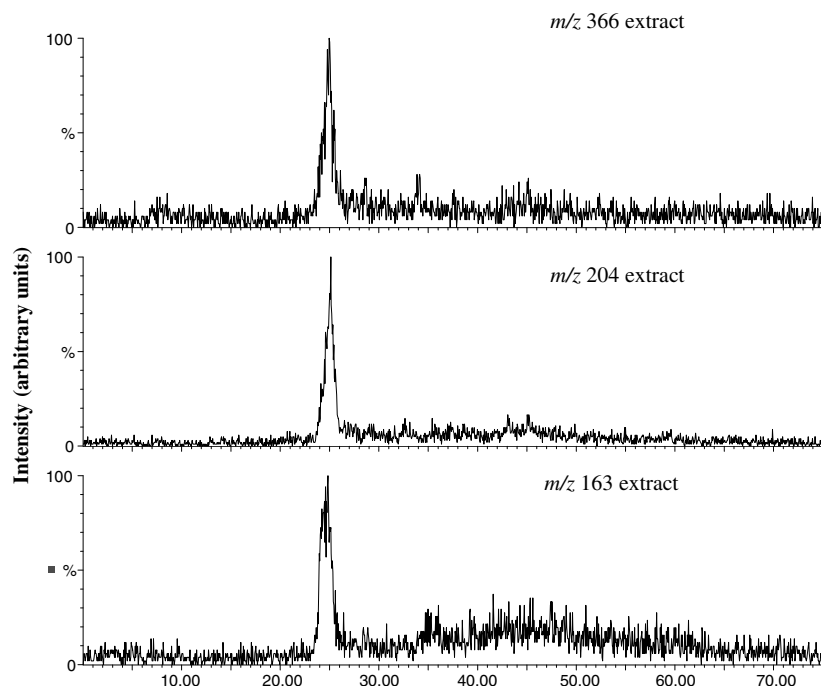


Fig. 3. Ion currents of three carbohydrate fragment ions extracted from TIC under high-cone voltage acquisition.

The molecular mass of $\text{GlcNAc}_2\text{Man}_{11}$, the largest glycan detected, was calculated to be 2190 Da; thus, N-glycosylation can account for only part of the molecular mass discrepancy observed in the intact protein analysis (~3300 Da). The remaining mass difference (~1110 Da) indicated that there should be other PTMs present in the protein, but peptide fingerprint mapping using both MALDI-TOF MS and LC-MS data revealed only one N-linked glycopeptide (T12). Therefore, the unexplained experimental mass was suspected to be due to O-glycosylation on one or more of the unmatched peptides (T1, T4, T5, T9, T10, T15, and T16), and O-linked site mapping was pursued as a result.

O-Linked glycopeptide identification and *O*-linked site mapping

The trypsin- and PNGaseF-digested peptide mixtures were separated on a microbore C18 column that was interfaced with a conventional ESI source. The Q-TOF parameters were optimized to accommodate the relatively high flow rate of 20 $\mu\text{l}/\text{min}$ (vs. 1 $\mu\text{l}/\text{min}$ as used in nanospray techniques). Fig. 5 shows the TIC obtained. Mass spectra

corresponding to each ion peak were obtained and compared with *in silico* digests. Two summed spectra corresponding to the ion peaks at 15.2 and 15.8 min (Figs. 6 and 7, respectively) showed multiple peaks spaced at intervals of m/z 81 and indicate that there is a possible presence of *O*-mannose on the peptide, as expected. The mass spectra obtained from the effluent at 15.2 min showed a dominant ion at m/z 1107.5 and a series of peaks with much weaker signals in the lower mass region. The ion peaks were also observed in the spectra obtained from the effluent at 15.8 min. The ion peak at m/z 702.4 corresponds to the doubly charged T1 peptide backbone (ATTCTFSGSEGASK, 1403.6 Da) after carboxyamidomethylation of the Cys4 residue, indicating that the glycopeptides eluting at these two retention times actually originate from the same peptide. The fraction at 15.2 min corresponds to the T1 peptide with a maximum of five hexoses attached, and the fraction at 15.8 min corresponds to the T1 peptide with a maximum of four hexoses attached.

O-Linked glycosylation site mapping was thoroughly performed using pseudo-neutral loss scanning and BEMAD. The pseudo-neutral loss scanning experiment clearly established five hexoses (out of six potential sites for

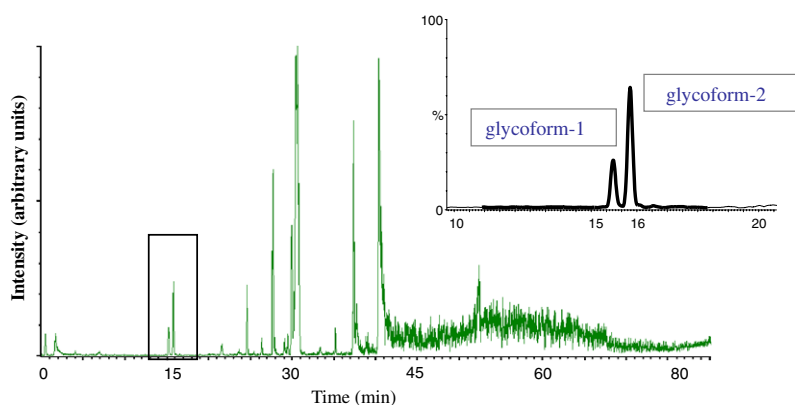


Fig. 5. Total ion current of PGC tryptic peptides obtained under conventional ESI LC-MS. Glycopeptides elute at 15.2 and 15.8 min. The purified *O*-linked peptide was collected over a 1-min window of 15–16 min.

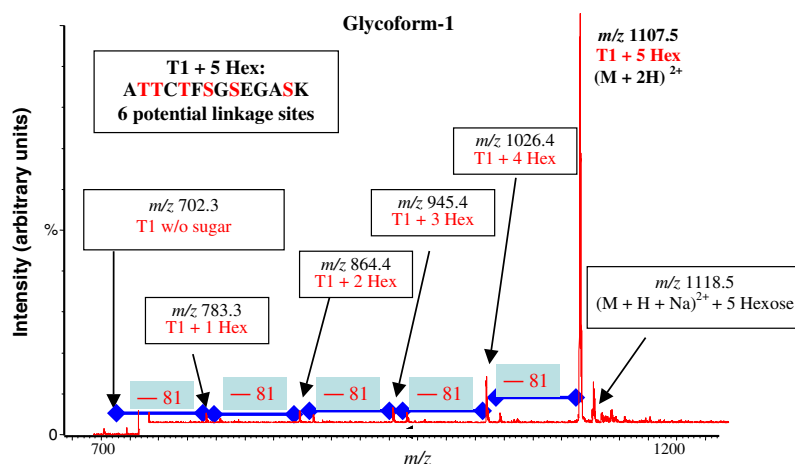


Fig. 6. Mass spectra of the T1 *O*-glycopeptide (modified with five hexoses) with retention time at 15.2 min. Sequential shift of m/z 81 indicates the presence of hexose for the $(M + 2H)^{2+}$ ions. w/o, without.

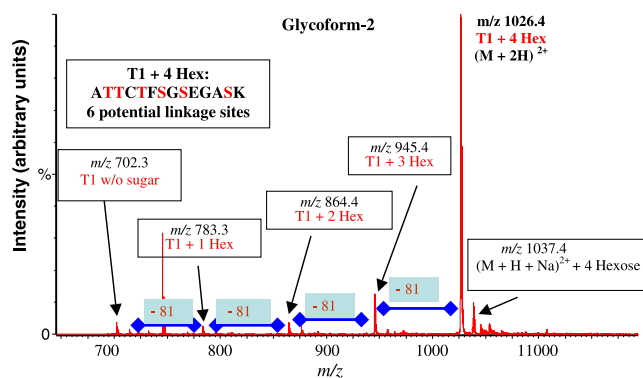


Fig. 7. Mass spectra of the T1 *O*-glycopeptide (modified with four hexoses) with retention time at 15.8 min. Sequential shift of m/z 81 indicates the presence of hexose for the $(M + 2H)^{2+}$ ions. w/o, without.

O-glycosylation on the T1 tryptic peptide sequence) identified by a sequential loss of m/z 81 (due to the doubly charged peptide) in the MS/MS spectra (Fig. 8) and provided support for the results of the previous Q-TOF MS analysis. MS³ analysis of the m/z 702.2 ion identified the peptide sequence to be ATTCTFSGSEGASK, as expected ($X_{\text{corr}} = 4.2$). Neutral loss also identified a hexose modification on the tryptic peptide T9 containing the glycosylated Ser114 residue ($X_{\text{corr}} = 4.1$). MS³ experiments occasionally can map sites of modification directly, and MS³ fragmentation of the m/z 864.1 ion, which had lost three of the five sugars, provided information that allowed us to map the sites of hexose modification (+162.05 Da) specifically onto Thr3 and Thr5 of the same T1 peptide. Direct identification by MS³ sequencing also allowed us to map an additional hexose modification on the tryptic peptide T5 containing the glycosylated Ser70 residue ($X_{\text{corr}} = 5.25$), as shown in Fig. 9. Although the neutral loss/MS³ approach is very useful for narrowing sites of modification to a single peptide, the exact site cannot always be determined unless there is only one serine or threonine in the peptide (which then allows the specific site to be mapped by inference) because

the loss of the *O*-linked glycan results in no net change to the site of modification (unlike the case of Ser/Thr phosphorylation, where a dehydro-amino acid is formed on loss of the phosphate modification) [33]. In addition, the neutral loss/MS³ approach does not allow for the enrichment of peptides of interest. Therefore, we were unable to assign the sugar modification to the exact site and were forced to pursue a different strategy (BEMAD), as described below, to map the exact location of the hexose modifications on the peptide.

A method has been developed recently for the mapping of *O*-GlcNAc sites on peptides that relies on BEMAD [21]. This approach has several promising features, including the replacement of a labile moiety (glycan) with a CID stable tag, and the dithiothreitol tag can be used to enrich the peptides of interest. This procedure can also be adapted for use as a quantitative technique, providing more flexibility for quantitative analysis of glycopeptides [34]. The β -elimination chemistry is applicable to any glycan type, so we applied this method to the analysis of *O*-mannose sites. Following protein digestion, the resulting peptides were subjected to the BEMAD chemistry to replace any *O*-mannose sites with covalent dithiothreitol and the peptides were enriched using thiol chromatography. LC-MS/MS analysis of the thiol-enriched BEMAD modified peptides provided data used to identify seven sites of *O*-glycosylation: Thr2, Thr3, Thr5, Ser7, and Ser9 on the T1 peptide and two other modifications at Ser70 ($X_{\text{corr}} = 4.9$) and Ser114 ($X_{\text{corr}} = 4.2$) on T5 (Fig. 10) and T9, respectively.

HPAEC-PAD is an established method for glycan composition analysis and was used to confirm the presence of *O*-mannose as opposed to any of the other possible *O*-hexose modifications. Compared with gas chromatography (GC)-MS, another common method for analysis of glycan composition, HPAEC is relatively simple. The method used in this work is a modified version of the Carlson degradation in which the PGC is heated for several hours in dilute NaOH with an excess of NaBH₄ [35]. The excess of NaBH₄

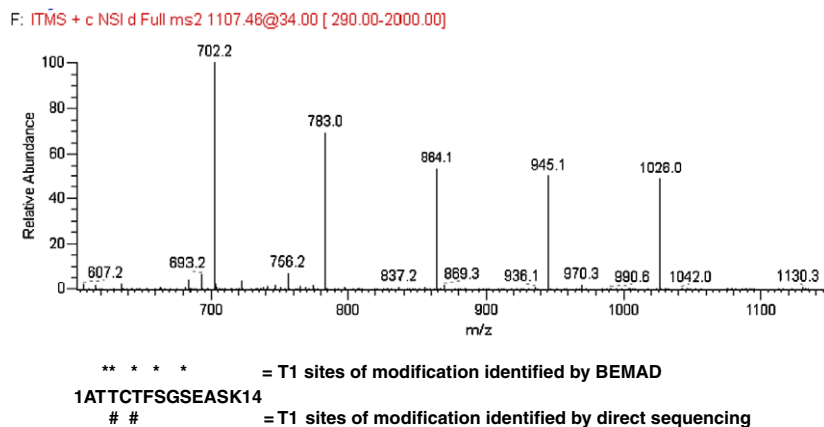


Fig. 8. Neutral loss site mapping of *O*-hexose sites on peptides of PGC. Shown is the MS/MS spectra of a tryptic peptide (T1) from PGC. MS³ sequencing of m/z 702.2 identified the peptide ATTCTFSGSEGASK. Direct sequencing of PGC peptides identified two sites (#) of modification on the T1 peptide, and BEMAD site mapping identified all five sites (*) of modification on the T1 tryptic peptide. A total of seven sites of *O*-mannose modification (Thr2, Thr3, Thr5, Ser7, Ser9, Ser70, and Ser114) were identified for PGC.

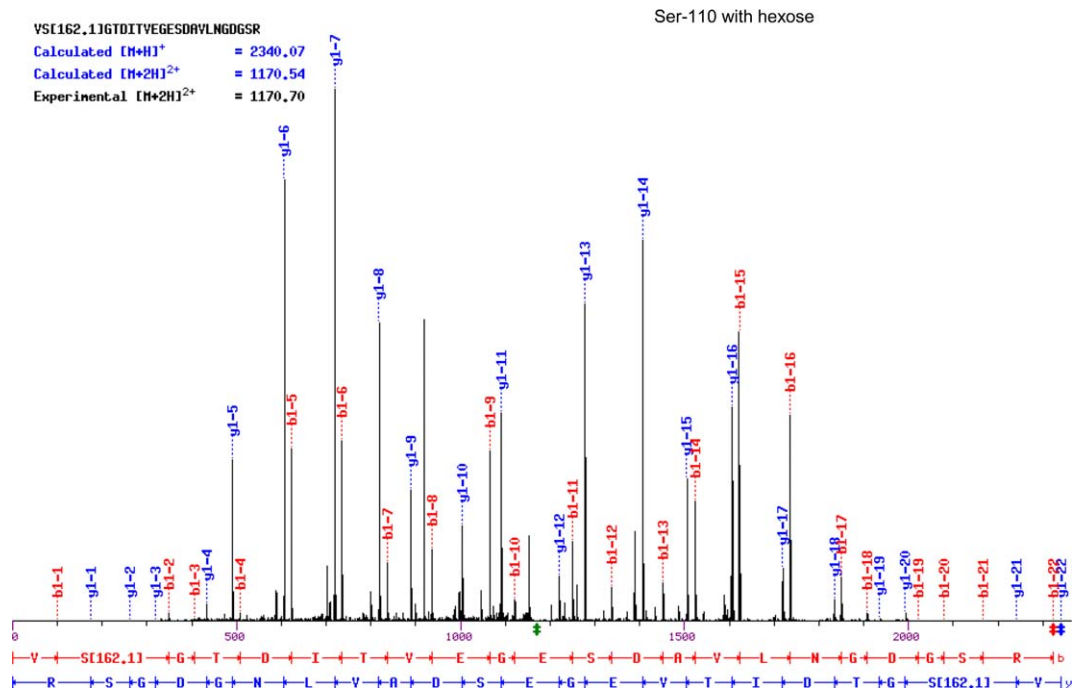


Fig. 9. Direct MS³ sequencing of *O*-hexose sites on peptides of PGC. Direct MS³ sequencing of PGC peptides identified Ser70 as a site of modification on the T5 peptide, as indicated by the +162.1 shift in mass of the serine residue at that position. Red and blue assignments represent theoretical fragmentation; the observed fragmentation pattern is in black.

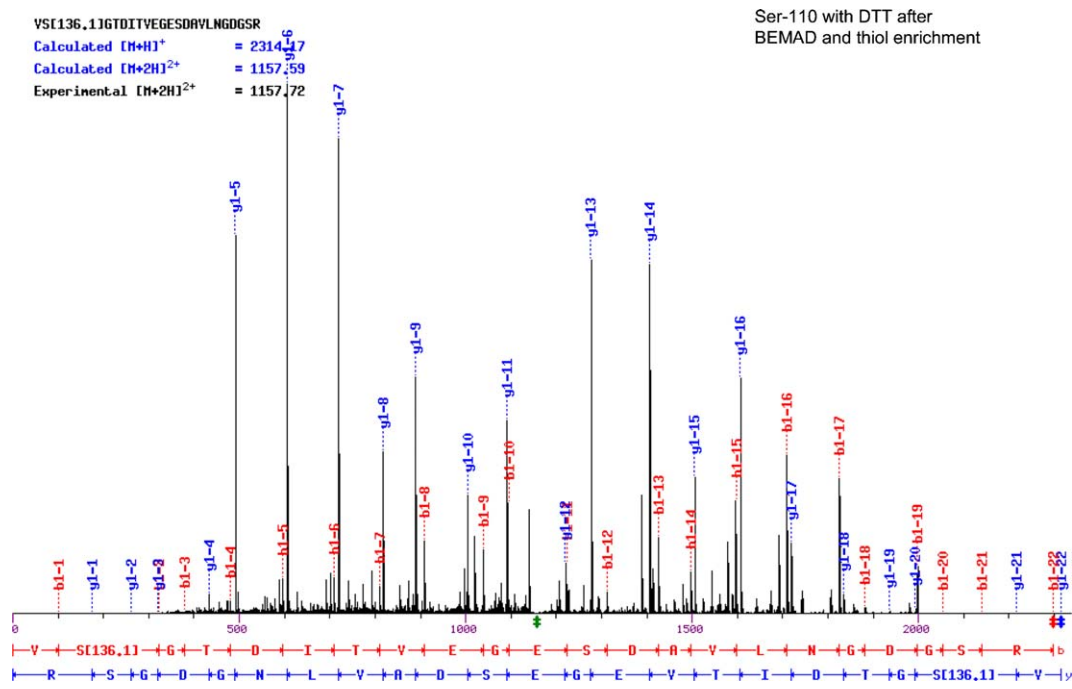


Fig. 10. BEMAD/Thiol chromatography site mapping of *O*-hexose sites on peptides of PGC. BEMAD followed by thiol chromatography identified Ser70 as a site of modification on the T5 peptide, as indicated by the +136.1 shift in mass of the serine residue at that position. In this approach, *O*-glycans are removed by β -elimination and replaced by dithiothreitol, which can be used to map specific sites of modification due to the additional mass of +136.1 on serine or threonine. Red and blue assignments represent theoretical fragmentation; the observed fragmentation pattern is in black. DTT, dithiothreitol.

is used to ensure fast reduction of released oligosaccharides and to protect them from further degradation under alkaline conditions. After releasing the *O*-linked glycans, acid hydrolysis using TFA was performed to cleave any glycosidic bonds between monosaccharide residues. Only one

peak was observed in the HPAEC-PAD result (Fig. 11), suggesting that recombinant PGC contains only one type of monosaccharide. A mannose standard was analyzed in the same manner to confirm the identity of this monosaccharide, and the peaks present in both chromatograms

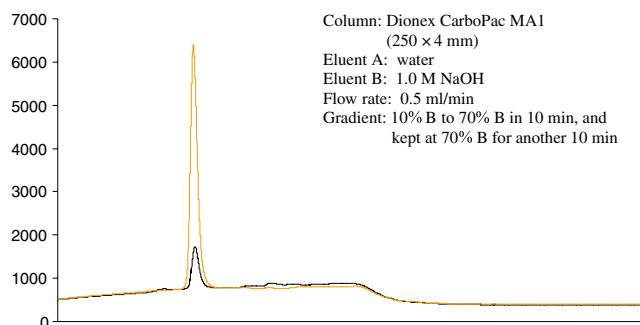


Fig. 11. HPAEC–PAD of O-linked glycan composition analysis. HPAEC–PAD data showed that all of the O-glycosylation sites on PGC are *O*-mannose. The shorter peak in the spectrum (PGC O-linked glycan) matches perfectly to the taller peak (mannose standard) and confirms that all O-glycosylation sites on PGC contain a single *O*-mannose modification.

overlap (Fig. 11). Therefore, the monosaccharide in PGC was determined to be entirely mannose, and the results presented here indicate that *O*-mannose is one type of fungal glycosylation.

Peptide fingerprint mapping was performed based on MALDI-TOF and LC-MS/MS data, and seven tryptic peptides with m/z values greater than 600 Da were unable to be identified. The N- and O-linked glycosylation sites on PGC and the structures of the carbohydrate chain were determined by various MS methods. A single N-linked glycosylation site (Asn220) was found in the recombinant PGC at T12, and the attached glycan was proposed to have a high-mannose structure (GlcNAc₂Man_{3–11}) that was consistent with that observed for other fungal enzymes [13,15]. The T1 peptide was identified as an O-linked glycopeptide from LC-MS/MS data based on the heterogeneity of glycoforms, and the attached glycan was further identified as mannose by HPAEC–PAD analysis. Overall, seven specific sites of *O*-mannose modification (Thr2, Thr3, Thr5, Ser7, Ser9, Ser70, and Ser114) were identified using a combination of neutral loss and BEMAD experiments.

The combined experimental mass from N- and O-linked glycans, 3323 Da, accounts for the mass discrepancy (~3300 Da) between the experimental and calculated molecular masses for PGC. The lack of experimental evidence for any other type of PTM (i.e., phosphorylation) combined with the mass comparison of the individual modified fragments with the intact glycoprotein mass from MALDI-TOF experiments indicates that the glycopeptides found in this research account for all of the PTMs on *A. niger* PGC.

The locations of all PTMs of PGC are shown in Fig. 12, and the locations of *O*-mannose modification are of particular interest in understanding the biological role of this enzyme. As mentioned earlier, enzymes such as PGC are released by the fungus during pathogenesis in an attempt to degrade the plant cell wall. The plant responds to this by producing proteinaceous inhibitors known as polygalacturonase-inhibiting proteins (PGIPs) [36]. The mannoses are placed in the region where, according to one model, the PGIP is proposed to interact with the EPG [37]. The role of

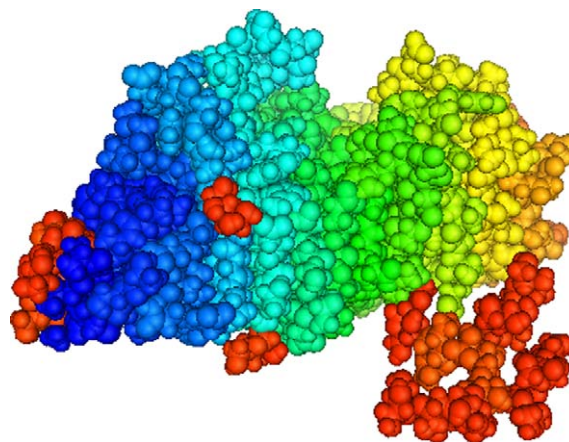


Fig. 12. Threaded representation of *O*-mannose and N-linked GlcNAc₂Man₁₁ glycosylation (both shown in orange) of PGC.

O-linked glycosylation has been an area of increasing interest, and recently *O*-linked mannose has been shown to be important in protein–protein interactions [16,17]. A systematic site mutagenesis of the mannose-containing serines and threonines will allow us to understand what roles each of the mannoses plays (if any) in EPG–PGIP interactions and will lead us to new strategies to promote plant resistance of pathogens. Because the BEMAD methodology originally developed for the mapping of *O*-GlcNAc sites is dependent on the glycan being susceptible to β -elimination, it was expected that it would work for site mapping any type of O-glycosylation. The mapping of simple *O*-mannose sites here serves as proof of principle that the BEMAD methodology does indeed work for *O*-glycans besides *O*-GlcNAc. Having demonstrated the utility of BEMAD for mapping fungal, nonextended *O*-mannose sites, this study sets the stage for mapping complex *O*-glycans, including highly medically relevant *O*-mannose sites such as those found on mammalian α -dystroglycan [16,17].

Acknowledgments

This work was supported in part by 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 (MDA0474). Special appreciation goes to Jacques A. E. Benen for providing the recombinant PGC sample, without which this work would not have been possible.

References

- [1] R.I. Barense, M.A. dos, S.C. Chellegatti, M.J.V. Fonseca, S. Said, Partial purification and characterization of exopolygalacturonase II and III of *Penicillium frequentans*, *Braz. J. Microbiol.* 32 (2001) 327–330.
- [2] C. Grassin, P. Fauquembergue, Application of pectinases in beverages, in: J. Visser, A.G.J. Voragen (Eds.), *Pectins and Pectinases*, Elsevier Science, Amsterdam, 1996, pp. 453–462.
- [3] H.P. Heldt-Hansen, L.V. Kofod, G. Budolfsen, P.M. Nielsen, S. Hützel, T. Bladt, Application of tailor-made pectinases, in: J. Visser, A.G.J.

- Voragen (Eds.), Pectins and Pectinases, Elsevier Science, Amsterdam, 1996, pp. 463–474.
- [4] C. Riou, G. Freyssinet, M. Fèvre, Purification and characterization of extracellular pectinolytic enzymes produced by *Sclerotinia sclerotiorum*, Appl. Environ. Microbiol. 58 (1992) 578–583.
- [5] A.G.J. Voragen, W. Pilnik, J.-F. Thibault, M.A.V. Axelos, C.M.G.C. Renard, 10 pectins, in: A.M. Stephen (Ed.), Food Polysaccharides and Their Applications, Marcel Dekker, New York, 1995, pp. 287–339.
- [6] M.A. O'Neill, P. Albersheim, A. Darvill, The pectic polysaccharides of primary cell walls, in: D.M. Dey (Ed.), Methods in Plant Biochemistry, vol. 2, Academic Press, London, 1990, pp. 415–441.
- [7] R.P. de Vries, J. Visser, *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides, Microbiol. Mol. Biol. Rev. 65 (2001) 497–522.
- [8] J.A.E. Benen, H.C.M. Kester, J. Visser, Kinetic characterization of *Aspergillus niger* N400 endopolygalacturonases I, II, and C, Eur. J. Biochem. 259 (1999) 577–585.
- [9] L. Parenicová, J.A.E. Benen, H.C.M. Kester, J. Visser, pgaE encodes a fourth member of the endopolygalacturonase gene family from *Aspergillus niger*, Eur. J. Biochem. 251 (1998) 72–80.
- [10] J.A.E. Benen, H.C.M. Kester, L. Parenicová, J. Visser, Kinetics and mode of action of *Aspergillus niger* polygalacturonases, in: J. Visser, A.G.J. Voragen (Eds.), Pectins and Pectinases, Elsevier Science, Amsterdam, 1996, pp. 221–230.
- [11] L. Pařenicová, H.C.M. Kester, J.A.E. Benen, J. Visser, Characterization of a novel endopolygalacturonase from *Aspergillus niger* with unique kinetic properties, FEBS Lett. 467 (2000) 333–336.
- [12] J.A.E. Benen, G.J.W.M. van Alebeek, A.G.J. Voragen, J. Visser, Mode of action analysis and structure–function relationships in *Aspergillus niger* pectinolytic enzymes, in: F. Voragen, H. Schols, R. Visser (Eds.), Advances in Pectin and Pectinase Research, Kluwer Academic, Dordrecht, Netherlands, 2003, pp. 235–256.
- [13] J. Colangelo, V. Licon, J. Benen, J. Visser, C. Bergmann, R. Orlando, Characterization of the N-linked glycosylation site of recombinant pectate lyase, Rapid Commun. Mass Spectrom. 13 (1999) 2382–2387.
- [14] D.B. Archer, J.F. Peberdy, The molecular biology of secreted enzyme production by fungi, Crit. Rev. Biotechnol. 17 (1997) 273–306.
- [15] Y. Yang, C. Bergmann, J. Benen, R. Orlando, Site-specific structural characterization of the glycans of recombinant endo-polygalacturonase II (EPG-II) by mass spectrometry, Rapid Commun. Mass Spectrom. 11 (1997) 1257–1262.
- [16] R.S. Haltiwanger, J.B. Lowe, Role of glycosylation in development, Annu. Rev. Biochem. 73 (2004) 491–537.
- [17] P.T. Martin, Dystroglycan glycosylation and its role in matrix binding in skeletal muscle, Glycobiology 13 (2003) 55R–65R.
- [18] M.J. Huddleston, M.F. Bean, S.A. Carr, Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: Methods for selective detection of glycopeptides in protein digests, Anal. Chem. 65 (1993) 877–884.
- [19] M.A. Ritchie, A.C. Gill, M.J. Deery, K. Lilley, Precursor ion scanning for detection and structural characterization of heterogeneous glycopeptide mixtures, J. Am. Soc. Mass Spectrom. 13 (2002) 1065–1077.
- [20] A.P. Hunter, D.E. Games, Evaluation of glycosylation site heterogeneity and selective identification of glycopeptides in proteolytic digests of bovine α_1 -acid glycoprotein by mass spectrometry, Rapid Commun. Mass Spectrom. 9 (1995) 42–56.
- [21] L. Wells, K. Vosseller, R. Cole, J. Cronshaw, M. Matunis, G. Hart, Mapping O-GlcNAc sites using affinity tags for serine and threonine post-translational modifications, Mol. Cell. Proteomics 1 (2002) 791–804.
- [22] G. Neubauer, M. Mann, Parent ion scans of large molecules, J. Mass Spectrom. 32 (1997) 94–98.
- [23] H.J.D. Bussink, H.C.M. Kester, J. Visser, Molecular cloning, nucleotide sequence, and expression of the gene encoding prepro-polygalacturonase of *Aspergillus niger*, FEBS Lett. 273 (1990) 127–130.
- [24] J. Colangelo, R. Orlando, On-target *exo*-glycosidase digestions/MALDI-MS for determining the primary structures of carbohydrate chains, Anal. Chem. 71 (1999) 1479–1482.
- [25] M.R. Hardy, R.R. Townsend, High-pH anion-exchange chromatography of glycoprotein-derived carbohydrates, Methods Enzymol. 230 (1994) 208–225.
- [26] H.J. Bussink, F.P. Buxton, B.A. Fraaye, L.H. deGraff, J. Visser, The polygalacturonases of *Aspergillus niger* are encoded by a family of diverged genes, Eur. J. Biochem. 208 (1991) 83–90.
- [27] Y. Van Santen, J.A.E. Benen, K.-H. Schroer, K.H. Kalk, S. Armand, J. Visser, B.W. Dijkstra, 1.68-Å Crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis, J. Biol. Chem. 274 (1999) 30474–30480.
- [28] P. Biely, J.A.E. Benen, K. Heineichova, H.C.M. Kester, J. Visser, Inversion of configuration during hydrolysis of α -1,4-galacturonidic linkage by three *Aspergillus* polygalacturonases, FEBS Lett. 382 (1996) 249–255.
- [29] A. Armand, M.J.M. Wagemaker, P. Sánchez-Torres, H.C.M. Kester, Y. van Santen, B.W. Dijkstra, J. Visser, J.A.E. Benen, The active topology of *Aspergillus niger* endopolygalacturonase II as studied by site-directed mutagenesis, J. Biol. Chem. 275 (2000) 691–696.
- [30] G. Limberg, R. Körner, H.C. Buchholt, T.M.I.E. Christensen, P. Roepstorff, J.D. Mikkelsen, Analysis of different de-esterification mechanisms for pectin by enzymatic fingerprinting using endopectin lyase and endopolygalacturonase II from *A. niger*, Carbohydr. Res. 327 (2000) 293–307.
- [31] J. Colangelo, V. Licon, J. Benen, J. Visser, C. Bergmann, R. Orlando, Characterization of the glycosylation of recombinant endo-polygalacturonase I from *Aspergillus niger*, Rapid Commun. Mass Spectrom. 13 (1999) 1448–1453.
- [32] J. Colangelo, V. Licon, J. Benen, J. Visser, C. Bergmann, R. Orlando, Characterization of the N-linked glycosylation site of recombinant pectate lyase, Rapid. Commun. Mass Spectrom. 13 (1999) 2382–2387.
- [33] S.A. Beausoleil, M. Jedrychowski, D. Schwartz, J.E. Elias, J. Villen, J. Li, M.A. Cohn, L.C. Cantley, S.P. Gygi, Large-scale characterization of HeLa cell nuclear phosphoproteins, Proc. Natl. Acad. Sci. USA 101 (2004) 12130–12135.
- [34] K. Vosseller, K.C. Hansen, R.J. Chalkley, J.C. Trinidad, L. Wells, G.W. Hart, A.L. Burlingame, Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol, Proteomics 5 (2005) 388–398.
- [35] D.M. Carlson, Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins, J. Biol. Chem. 243 (1968) 616–626.
- [36] G. Kemp, L. Stanton, C.W. Bergmann, R.P. Clay, A. Darvill, P. Albersheim, Polygalacturonase-inhibiting proteins can function as activators of polygalacturonase, Mol. Plant–Microbe Interact. 17 (2004) 888–894.
- [37] D. King, C. Bergmann, R. Orlando, J.A.E. Benen, H.C.M. Kester, J. Visser, The use of amide exchange–mass spectrometry to study conformational changes within the endopolygalacturonase II/homogalacturonans/polygalacturonase-inhibiting protein system, Biochemistry 41 (2002) 10225–10233.