Phytochemistry 72 (2011) 1173–1179

Contents lists available at ScienceDirect

Phytochemistry

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

Different site-specific N-glycan types in wheat (*Triticum aestivum* L.) PAP phytase

Giuseppe Dionisio [a], Henrik Brinch-Pedersen [a,*], Karen Gjesing Welinder [b], Malene Jørgensen [b,1]

[a] Aarhus University, Faculty of Agricultural Sciences, Dept. of Genetics and Biotechnology, Research Centre Flakkebjerg, DK-4200 Slagelse, Denmark
[b] Section for Biotechnology, Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Sohngaardsholmsvej 49, DK-9000 Aalborg, Denmark

ARTICLE INFO

Article history:
Available online 15 February 2011

Keywords:
Triticum aestivum L. cv. bobwhite
PAP phytase
Plant glycans
Grain protein

ABSTRACT

Phytase activity in grain is essential to make phosphate available to cell metabolism, and in food and feed. Cereals contain the purple acid phosphatase type of phytases (PAPhy). Mature wheat grain is dominated by TaPAPhy_a which, in the present work, has been characterized by extensive peptide and glycopeptide sequencing by mass spectrometry. Seven N-linked glycosylation sites were found. Three of these sites were dominated by variant forms of the $XylMan_3FucGlcNAc_2$, i.e. the HRP-type of glycan. Complex-type glycans with one or two additional GlcNAc were observed, however in trace amounts only. At four sites the glycan consisted of a single GlcNAc residue. The mature protein is ca. 500 residues in size and appears to be truncated at the N- and C-termini.

1. Introduction

Phytases (*myo*-inositol hexakisphosphate phosphohydrolase) [EC 3.1.3.26 and EC 3.1.3.8] are phosphatases which can initiate the sequential hydrolysis of orthophosphate groups from phytate ($InsP_6$, *myo*-inositol 1,2,3,4,5,6-hexakisphosphate), thus providing phosphate, inositol phosphates and inositol required for a range of cellular activities (Brinch-Pedersen et al., 2002). Phytases are known from animals, microorganisms and plants (Dvorakova, 1998). They are classified according to their catalytic mechanism as histidine acid phosphatases (HAPs), cysteine phosphatases (CPs), β-propeller phytases BPPhys, and purple acid phosphatases (PAPs) (see Lei et al., 2007 for a review). In plants only HAP and PAP phytases have been identified so far. Phytases are of particular importance during seed germination because they are the primary determinants for the mobilization of phosphate from phytate, the major reserve of phosphorus (P) in plant seeds accounting for ~70% of the total P (Lott, 1984). Moreover, they have profound effects on phosphate bioavailability when dry grains are used as food or feed.

Among cereals, wheat (*Triticum aestivum and durum* L.), barley (*Hordeum vulgare* L.) and rye (*Secale cereale* L.) possess a significant phytase activity in mature grains. In a recent work, the PAPhy genes from wheat, barley, maize and rice were cloned and characterized (Dionisio et al., 2011). Pre formed PAPhy in wheat grains were localized in the vacuoles of the aleurone layer. For wheat and barley, *PAPhy* genes could be divided into two isogenes, designated *PAPhy_a* and *PAPhy_b*. For each isogene up to two variants were identified. In wheat they were named *TaPAPhy_a1* (FJ973998), *TaPAPhy_a2* (FJ973999), *TaPAPhy_b1* (FJ974000) and *TaPAPhy_b2* (FJ974001). Peptide mapping of TaPAPhy purified from bran and from germinating grains confirmed that pre-formed phytase activity in mature grains is constituted largely by the TaPAPhy_a1 and a2 isoforms, whereas phytases synthesized *de novo* during wheat grain germination are dominated by the TaPAPhy_b1 and b2 phytases (Dionisio et al., 2011).

Plant protein N-glycosylation has long been recognized as a common post-translational modification of proteins destined to the vacuole and the extra-cellular environment (Wilson, 2002). The protein N-glycosylation in plants has a great impact on their physico-chemical properties and often on their biological activity. In the present study, we report a very high protein sequence coverage and observed glycan structures of wheat TaPAPhy_a1 and _a2 derived from mass spectrometric sequencing of tryptic, chymotryptic and AspN peptides. Our data shows that wheat phytase contains different N-glycan structures, however the glycan type appears to be site-specific indicating that N-glycan processing depends on the surrounding peptide sequence or protein structure. Glycan heterogeneity is common in plants, though normally represented by variant forms only of one glycan type, such as variants of the most abundant $(Xyl)Man_3(Fuc)GlcNAc_2$ identified at all eight N-glycan sites in horseradish peroxidase (Gray et al., 1998). We discuss the implications of our data on the mechanism of recognition of the specific glycosyl transferases and hydrolases processing N-glycans in the ER and Golgi (Wilson, 2002).

* Corresponding author. Tel.: +45 89993651; fax: +45 89993501.
E-mail addresses: Giuseppe.dionisio@agrsci.dk (G. Dionisio), Henrikbrinchpedersen@agrsci.dk (H. Brinch-Pedersen), kgw@bio.aau.dk (K.G. Welinder), maljoe@rn.dk (M. Jørgensen).
[1] Present address: Department of Clinical Immunology, Aalborg Hospital, Aarhus University Hospital, Urbansgade 32-36, DK-9000 Aalborg, Denmark

0031-9422/$ - see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.phytochem.2010.12.023

2. Results

2.1. TaPAPhy_a sequencing

Fig. 1 shows the amino acid sequences of TaPAPhy_a1 (ACR23326) and a2 (ACR23327) translated from cDNA clones. This figure also provides a summary of peptides and glycans documented in this study by MS analyses of peptides derived from wheat PAPhy purified from bran. Peptides identified in bran PAPhy are shown in the Supplementary Data, chymotryptic peptides in Tables S1a and S1b (deglycosylated protein), tryptic peptides in Table S2, AspN peptides in Table S3. The data confirms that bran contained mainly phytase TaPAPhy_a isoforms. The good sequence coverage of TaPAPhy_a was obtained from peptides of reduced and S-carboxymethylated protein in both of the naturally glycosylated form and after deglycosylation with glycopeptidase A (Fig. 1). Together, chymotrypsin, trypsin, and AspN provided overlapping peptides covering the mature protein sequence up to 10 times (Tables S1–S3). Some proline-containing sequences were not cleaved, such as the sequence 441–449 which contains an amazing five prolines out of nine residues. Small and very large peptides are often lost in standard LC-MS. Peptides from LysC digestion added only few and confirmatory peptides (not shown).

SignalP prediction (Bendtsen et al., 2004; www.cbs.dtu.dk/services/SignalP) suggested very little ER-signalpeptidase specificity at the N-terminus due to an abundance of small size residues. Thus the phytase molecules during folding in the ER most likely have differential N-termini. An N-terminal AspN-peptide was sequenced twice in TaPAPhy_a1 (Table S3) indicating that Thr29 is the N-terminus of the mature phytase. The same N-terminal sequence derived from chymotryptic peptides was also sequenced twice in TaPAPhy_a1 and the corresponding sequence from _a2 once (Table S1a). However, chymotrypsin can cleave after Leu28, so only the AspN peptide can support Thr29 as the likely N-terminus of mature TaPAPhy_a1 and an N-terminal truncation. At the C-terminus six chymotryptic peptides terminated after Leu529 (Table S1b) and no peptides were seen after Leu529 in wt or deglycosylated phytase. For that reason the glycan status of Asn534 and Asn544 is unknown. There might be C-terminal processing and a C-terminal propeptide; but the data does not prove this.

2.2. TaPAPhy_a glycan sequencing

None of the remaining seven putative N-linked Asn residues, 137, 202, 233, 289, 411, 462 and 497, were seen as non-glycosylated residues in chymotryptic, tryptic and AspN digests (Tables S1a, S2, and S3). This indicates that all are fully glycosylated, taking into account the high sequence coverage and that the glycopeptides will not be recognized by the standard MS-analysis software (glycopeptide data summarized in Table 1). Because chymotrypsin provided the best sequence coverage, an additional chymotryptic digestion of glycopeptidase A deglycosylated phytase was carried out as a control. Deglycosylated peptides covered three sites, 137, 289 and 462, showing free Asp residues (Table S1b). This confirms the presence of N-linked glycans at these residues because glycopeptidase A removes N-glycans and leaves an aspartate residue in place of the original asparagine (Fig. 2; Figs. S1A, S1B). Glycopeptidase A should be capable of removing also complex N-linked glycans (in contrast with glycopeptidase F). However, only the simple GlcNAc glycans were removed (see Table 1).

The MS2 spectra not assigned by the standard MS software were searched manually for protonated GlcNAc at 204.09 and GlcNAc-Man at 366.14, indicative of glycopeptide fragmentation. Independent of CID-MS instrument, the predominant fragment of

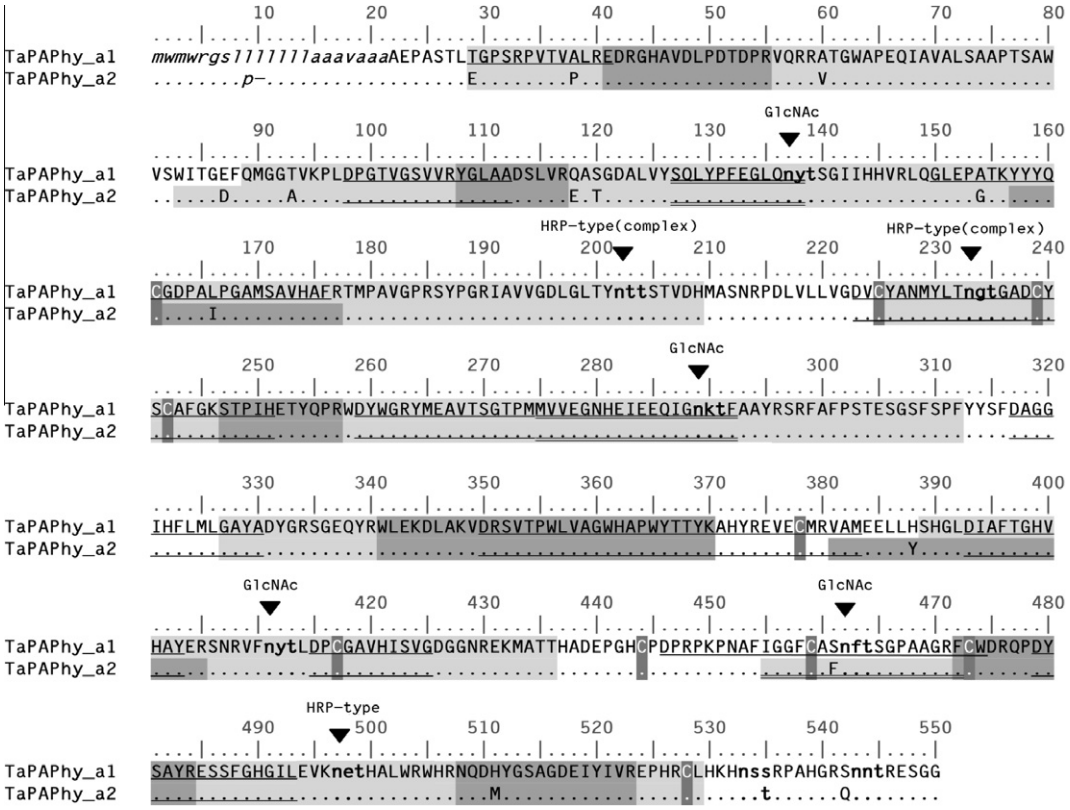

Fig. 1. Summary of amino acid sequences and glycans of wheat bran phytases. Identified peptides are marked, chymotryptic (grey shade), tryptic (dark gray shade), and AspN (underscored); chymotryptic peptides from deglycosylated phytase are shown double underscored. Predicted glycan sites are shown in small capital letter; observed glycans are shown by arrow head. Amino acids in TaPAPhy_a2 identical to a1 are shown as dots. A predicted ER signal peptide is shown in small italics and Cys in reversed print.

Table 1
Mass spectrometric analyses of N-linked glycans in TaPAPhy_a1 and TaPAPhy_a2.

Residue	Peptide/protease[a]	MS	Obs ion or parent ion	Glycan, and glycan mass (Da)
Asn137	SQLYPFEGLQNY/c	MS2	MH2+ 730.32	Yes[b] (Table S1b, Fig 2B)
	SQLYPFEGLQNY/c	MS2	MH2+ 831.38	GlcNAc 204.09 (Fig 2A)
	EGLQNY/c	MS1	MH2+ 464.42	GlcNAc 204.09
	QASGDALVYSQLYPFEGLQNYTSGIIHHVR/t	MS1	MH3+ 1190.63	GlcNAc 204.09
	QASGDALVYSQLYPFEGLQNYTSGIIHHVR/t	MS1	MH4+ 893.23	GlcNAc 204.09
Asn202	NTTSTVDHM/c	MS1		Glycan variants in MS1 scan (Fig. S2A):
			MH2+ 934.35	$XylMan_2GlcNAc_2$ 862.3
			MH2+ 1015.88	$XylMan_3GlcNAc_2$ 1024.36
			MH2+ 1088.89	$XylMan_3FucGlcNAc_2$ 1170.4
			MH2+ 1116.91	$GlcNAcXylMan_3GlcNAc_2$ 1227.4
	NTTSTVDHM/c	MS2	MH2+ 1015.88	$XylMan_3GlcNAc_2$ 1024.36 (Fig. S2B)
	DLGLTYNTTSTV/a	MS1		Glycan variants in MS1 scan (Fig. S3A):
			MH2+ 1073.95	$XylMan_2GlcNAc_2$ 862.3
			MH2+ 1088.96	$Man_3GlcNAc_2$ 892.3
			MH2+ 1155.61	$XylMan_3GlcNAc_2$ 1024.36
			MH2+ 1256.53	$GlcNAcXylMan_3GlcNAc_2$ 1227.4
	DLGLTYNTTSTV/a	MS2	MH2+ 1155.61	$XylMan_3GlcNAc_2$ 1024.36 (Supplementary Fig. S3B)
Asn233	LTNGTGADCYSCAF/c	MS1		Glycan variants in MS1 scan (Fig 3A):
			MH3+ 757.02	$XylManFucGlcNAc_2$ 846.3
			MH3+ 762.35	$XylMan_2GlcNAc_2$ 862.3
			MH3+ 767.03	$Man_2FucGlcNAc_2$ 876.3
			MH3+ 811.05	$XylMan_2FucGlcNAc_2$1008,36
			MH3+ 865.07	$XylMan_3FucGlcNAc_2$ 1170.4
			MH3+ 932.76	$GlcNAcXylMan_3FucGlcNAc_2$ 1373.5
			MH3+ 1000.46	$GlcNAc_2XylMan_3FucGlcNAc_2$ 1576.6
	LTNGTGADCYSCAF/c	MS2	MH2+ 1216.73	$XylMan_2FucGlcNAc_2$ 1008.4 (Fig 3B)
	DVCYANMYLTNGTGA/a	MS1		Glycan variants in MS1 scan (Fig. S4A):
			MH2+ 1614.09	$GlcNAc_2XylMan_3FucGlcNAc_2$ 1576.6
			MH2+ 1513.07	$GlcNAcXylMan_3FucGlcNAc_2$ 1373.5
			MH2+ 1439.55	$GlcNAcXylMan_3GlcNAc_2$ 1227.4
			MH2+ 1411.02	$XylMan_3FucGlcNAc_2$ 1170.4
			MH2+ 1345.02	$Man_3FucGlcNAc_2$ 1038.4
			MH2+ 1338.00	$XylMan_3GlcNAc_2$ 1024.4
	DVCYANMYLTNGTGA/a	MS2	MH2+ 1513.42	$GlcNAcXylMan_3FucGlcNAc_2$ 1373.5 (Fig. S4B)
Asn289	MVVEGNHEIEEQIGNKTF/c	MS2	MH3+ 692.32	Yes[b] (Table S1b, Fig. S1A)
	MVVEGNHEIEEQIGNKTF/c	MS1	MH3+ 760.57	GlcNAc 204.09
	MVVEGNHEIEQQIGNKTF/c[c]	MS2	MH2+ 1138.52	GlcNAc 204.09 (Fig 4)
	WDYWGRYMEAVTSGTPMMVVEG-NHEIEEQIGNKTFAAYR/t	MS1	MH3+ 1643.79	GlcNAc 204.09
Asn411	NYTLDPCGAVH/c	MS1	MH2+ 697.70	GlcNAc 204.09 (Fig. S5A)
	ERSNRVFNY/c	MS1	MH2+ 694.32	GlcNAc 204.09 (Fig. S5C)
	SNRVFNYTLDPCGAVHISVGDGGNR/t	MS1	MH3+ 951.99	GlcNAc 204.09 (Fig. S5B)
Asn462	TaPAPhy_a2: IGGFCAFNFTSGPAAGRF/c	MS2	MH2+ 939.91	Yes[b] (Table S1b, Fig. S1B)
	DPRPKPNAFIGGFCASNFTSGPAAGRFCW/a	MS1	MH3+ 1093.51	GlcNAc 204.09
Asn497	CWDRQPDYSAYRESSFGHGILEV-KNETHALW/c	MS2	MH4+ 1232.25	$XylMan_3FucGlcNAc_2$ 1170.4; MS1 scan[d] (Fig 5A)
	RESSFGHGILEVKNETHALW/c	MS2	MH3+ 922.42	$XylMan_3FucGlcNAc_2$ 1170.4; MS1 scan[d] (Fig 5B)
Asn533	Absent			
Asn543	Absent			

[a] Chymotrypsin (c), trypsin (t), AspN (a).
[b] Deglycosylated peptide sequence showing Asp not Asn at position N.
[c] E-residue substituted by Q, indicating the presence of a variant isophytase.
[d] MS1 scan of glycopeptide nano-LC peak showed one glycan size only.

a MH3+ glycopeptide precursor ion is the peptide-GlcNAc (Wuhrer et al., 2007). The Bruker micrOTOF-MS instrument fragments glycopeptides similarly from MH2+, MH3+ and MH4+ precursor ions (Welinder and Jørgensen, 2009). This search provided glycopeptides covering three sites, Asn202 (Figs. S2 and S3), Asn233 (Fig. 3 and Fig. S4), and Asn497 (Fig. 5). These peptides carry variants of different glycan types, (i) typical plant glycan or HRP-type: $(Xyl)Man_3(Fuc)GlcNAc_2$, and (ii) traces of complex type at Asn202, and Asn233. Analyses of the corresponding MS1 spectra revealed that this type of glycopeptides eluted from the on-line nano-LC column with the large glycan variants a few seconds before the short glycan variants (more hydrophobic). The sum of MS1 spectra over such a heterogeneous glycopetide peak is denoted an MS1 scan (Table 1; Fig. 3A; Figs. S2A, S3A, S4A). In contrast to Asn202 and Asn233, MS 1 scans of Asn497 peptides only showed one glycan form, the HRP-type.

Searches for missing peptides among unassigned MS1 spectra from the three digests indicated that glycans consisting of a single GlcNAc were present at several potential N-glycan sites. This was confirmed by a Mascot search allowing Asn-GlcNAc as variable modification at Asn. At four sites the glycan consisted of a single GlcNAc residue. This is documented for Asn202 (Fig. 2A), Asn289, Asn411 (Fig. S5), and Asn462 (Table 1). In general MS2 spectra of GlcNAc peptides were of low quality. The best provided both evidence of Asn-linked GlcNAc and the peptide sequence around Asn289 as shown in Fig. 4.

3. Discussion

Table 1 lists the N-linked glycan variants identified at each site. MS is not quantitative and data analyses might also introduce some ambiguity. Still, the present study does show a remarkable tendency toward a distinct preference for one glycan type at each of the seven glycan site. Only the small GlcNAc glycan was observed at Asn137, Asn289, Asn411, and Asn462. HRP- or complex-type were observed at Asn202, Asn233, and Asn497. At

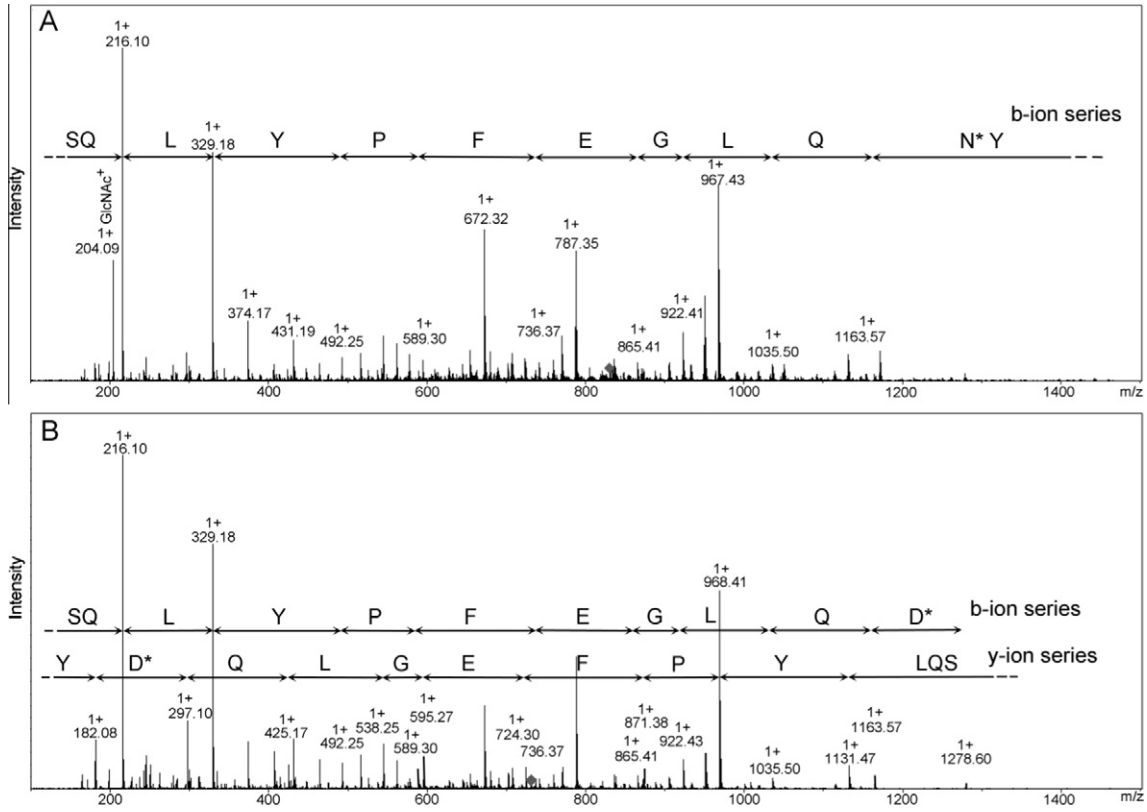

Fig. 2. Mass spectrometric collision-induced-dissociation fragmentation (MS2) of chymotryptic peptide 127–138. (A) The native peptide carrying GlcNAc at asparagine (N*) and the GlcNAc fragment at 204.09 and (B) The deglycosylated peptide with aspartate (D*).

Asn202 fucosyl was almost absent, whereas only the canonical HRP-type (Xyl)Man_3(Fuc)$GlcNAc_2$ was observed at Asn497.

Mass spectrometry determines mass/charge (m/z) of intact analytes such as peptides or glycopeptides. Some major molecules are then selected by the instrument for fragmentation and m/z measurements of fragments. The total mass of a molecule and differences in fragment size permit deduction of the sequence, i.e. amino acids in peptides and monosaccharides in a glycopeptide. The fragmentation created by the mass spectrometer used in the present study will split most of the glycan into smaller fragments leaving an intact peptide-GlcNAc fragment. The peptide is not sequenced (very low ion score), only the glycan as illustrated in Fig. 3B, Fig. 5A and B, Figs. S2B, S3b, S4B (Wuhrer et al., 2007; Welinder and Jørgensen, 2009). It is also important to notice that mass spectrometry cannot distinguish bond anomery and carbon links such as the β1–4 bond linking the two core GlcNAc. Furthermore, hexoses (Glc, Man, Gal), or pentoses, or deoxyhexoses are identified only as type because of identical masses. However, these species and the bond types have been well documented for plants, mammals and other eukaryotic species (Kornfeld and Kornfeld, 1985; Wilson, 2002). Therefore, we have labeled the glycans according to convention, but omitting linkage type.

In the plant ER, $Man_9GlcNAc_2$ is first attached to the asparagine side chain nitrogen by an oligosaccharide transferase (OST) which co-recognizes the hydroxyl group of a serine or threonine in AsnXaaSer/ThrYaa tetrapeptides, in general if Xaa and Yaa are not proline residues. Furthermore, asparagines close to the protein C-terminal are rarely glycosylated (Gavel and von Heijne, 1990). OST is part of the ER translocase complex and is situated on the luminal site. It appears to act on the nascent growing polypeptide chain 65 residues upstream of the P-site of the ribosome (Whitley et al., 1996). Furthermore, Thr appears to be more efficient than Ser as shown by systematic mutant studies in rabies virus glycoprotein (Kasturi et al., 1997). The nine putative N-glycan sites of TaPAPhy_a are not inhibited by proline resides at the Xaa or Yaa positions. In addition, all sites in both the _a1 and _a2 forms have Thr-residues, except for Asn534 in the _a1 form, which has Ser. The Asn534 and Asn544 sites were not observed as they might be part of a C-terminal propeptide. Thus our observation of glycosylation at all seven sites is in agreement with prediction of OST activity.

The next step in plant protein glycosylation is hydrolysis of Manα1–2 bonds by ER and Golgi I mannosidase leaving Man_5 (Wilson, 2002). This appears to have taken place efficiently in all glycans. In the Golgi, other specific mannosidases can hydrolyze Manα1–3 and Manα1–6 bonds thus giving rise to the common core glycan $Man_3GlcNAc_2$ which is shared by most plants and animals. In plants, this core glycan is the substrate of β1,2 xylosyltransferase, α1,3 fucosyltransferase, and GlcNAc transferase 2. In horseradish peroxidase (HRP) only β1,2 xylosyltransferase and α1,3 fucosyltransferase are active giving rise to the most abundant (Xyl)Man_3(Fuc)$GlcNAc_2$ or HRP-type glycan. In Fig. 5A and B the canonical form of the HRP-type glycan is shown. However, these two transferases rarely reach 100% completion. These incomplete transfers give rise to glycan heterogeneity at most sites carrying the HRP-type glycan, which is further increased by excessive mannosidase Manα1–3 and Manα1–6 activity (Table 1). In fact, this heterogeneity was first demonstrated by Gray et al. (1998) in HRP. The finding of trace amounts of complex glycans at Asn202 and Asn233 suggests that GlcNAc transferase 2 activity is present.

Yet, what is the origin of the unusual N-linked GlcNAc glycans at four sites (Table 1; Fig. 1)? O-linked GlcNAc is well known, but can be excluded as glycopeptidase A deglycosylated samples had Asp at the Asn137, Asn289 and Asn462 sites in agreement with glycopeptidase A activity. Glycan degradation might occur during phytase purification or proteolytic digestions. This seems unlikely because no glycan heterogeneity was observed at the four GlcNAc sites.

This leaves us with the possibility that wheat bran contains specific glycosidases, which may truncate the OST attached glycans to

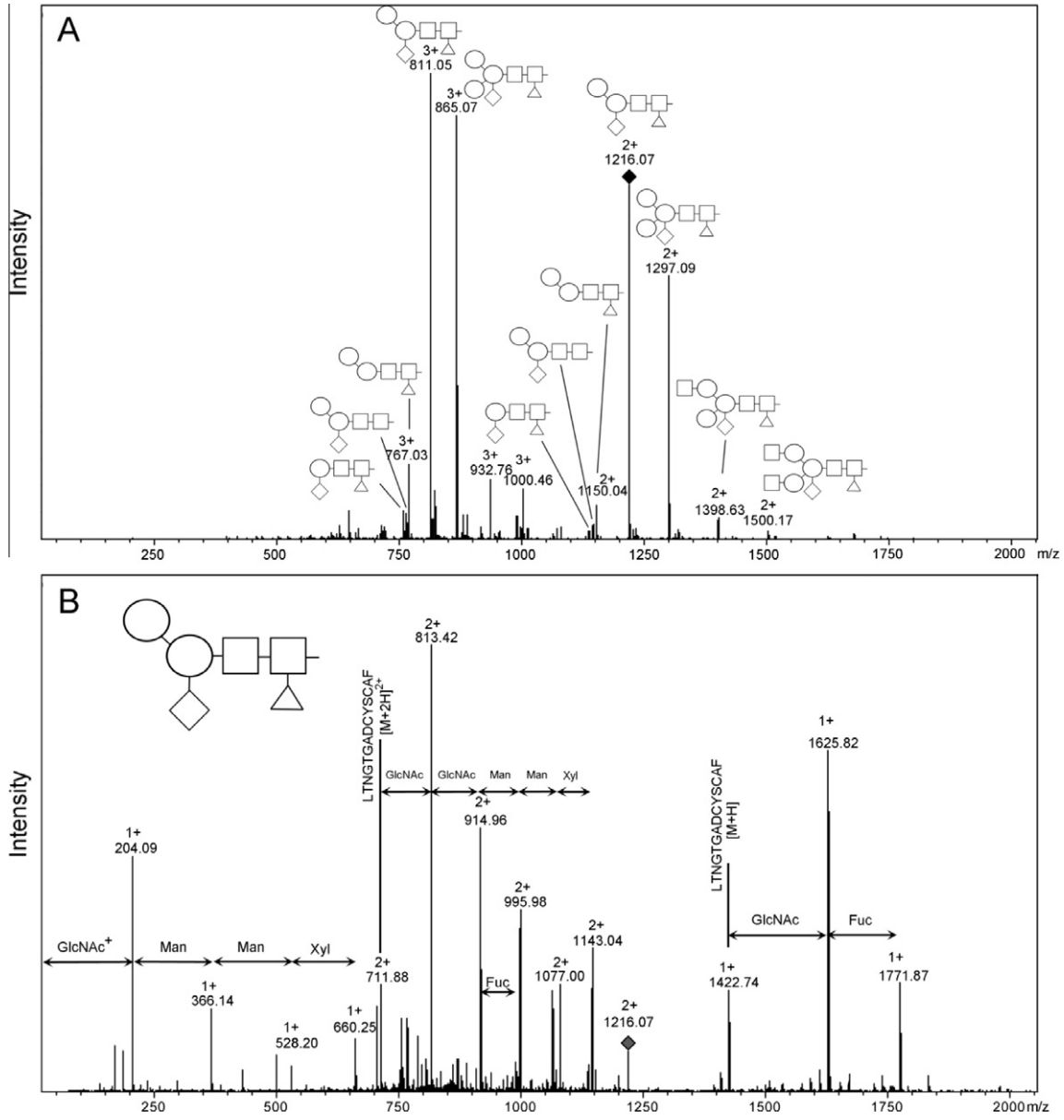

Fig. 3. (A) MS1 of chymotryptic peptide 231–244 containing Asn233 and showing seven glycan variants as both 2+ and 3+ charged ions. (B) The MS2 spectrum of the 2+ parent ion 1216.07 (diamond) and the derived HRP-type glycan structure. The 1+ and 2+ charged peptide-GlcNAc fragments are prominent at m/z 1625.82 and 813.42, respectively.

GlcNAc, depending on local peptide or protein structure. Endoglycosidase H-like activity cleaves the bond between the two core N-acetylglucosamines of simple or high-mannose type glycans. Such a glycosidase must then act in the ER or Golgi before fucosyltransferase to make a simple GlcNAc. Less likely, cereals may have a biosynthetic mechanism distinct from OST initiated N-glycosylation, such as a special GlcNAc-transferase attaching GlcNAc to Asn residues.

In the present study we have observed an unusual pattern of peptide sequence or protein structure dependent glycans. This requires that specific transferases or hydrolases exist, which recognize the substrate glycoprotein structure. A crystal structure of TaPAPhy may assist in answering this question.

4. Experimental

4.1. Plant material and TaPAPhy purification

Bobwhite SH 98 26 (*T. aestivum* L.) plants were grown in the greenhouse as described before (Brinch-Pedersen et al., 2000). Mature seeds were harvested at 12% relative humidity and milled. TaPAPhy was purified and characterized as previously described (Dionisio et al., 2011). In short, protein was extracted at 40 °C in presence of xylanase and β-glucanase, which will solubilize the cell wall, but not influence protein-linked glycans. The pellet of 60% ammoniumsulfate precipitation was dialyzed and phytase activity fractions from SP-Sepharose were further purified by Q-Sepharose chromatography. Coomassie brilliant blue stained SDS–PAGE and Western blotting, using polyspecific rabbit antibody recognizing all TaPAPhy a and b-forms, indicated >90% phytase purity. The TaPAPhy used for the present study was not fractionated by ConA lectin chromatography.

4.2. Sequence and glycan analyses of TaPAPhy

Proteolytic digestions and nano-LC MS/MS analyses were as previously described (Welinder and Jørgensen, 2009). Purified phytase (10 μg) was reduced, S-carboxymethylated and digested by either sequencing grad modified porcine trypsin (Promega,

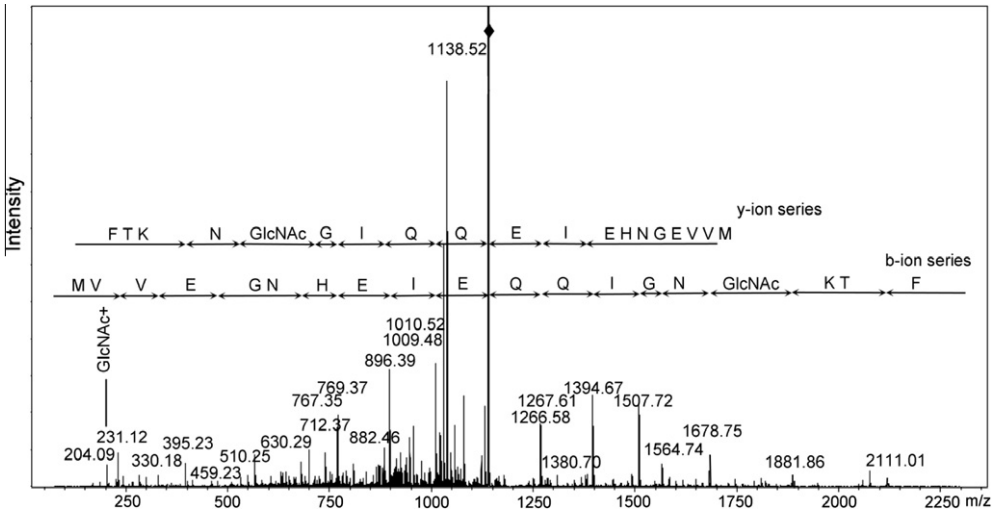

Fig. 4. MS2 sequencing of chymotryptic glycopeptide 275–292. Both the b- and y-ion series demonstrate GlcNAc attachment to Asn289.

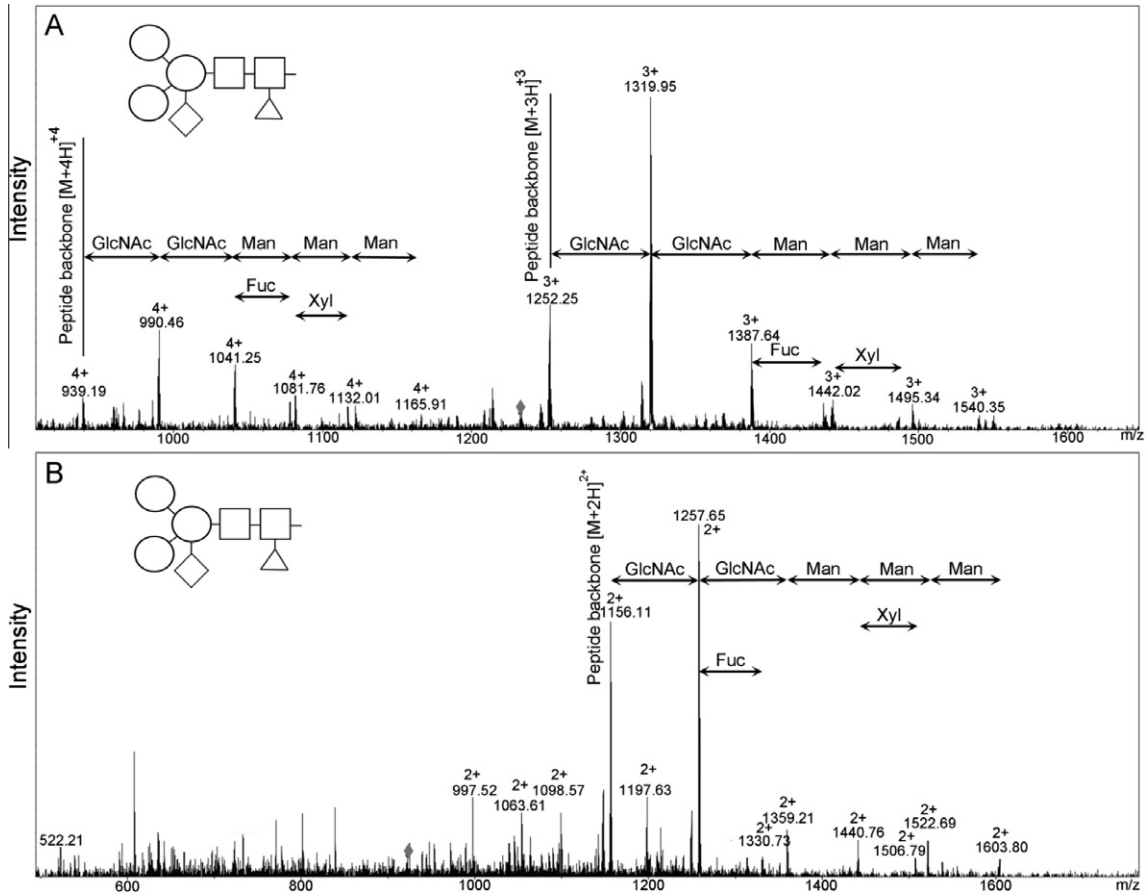

Fig. 5. MS2 sequencing of $XylMan_3FucGlcNAc_2$ glycan at Asn497. (A) Chymotryptic glycopeptide 473–503. The 3+ and 4+ charged peptide-GlcNAc fragments are prominent at m/z 1319.95 and 990.46, respectively and (B) Chymotryptic glycopeptide 484–503. The double-charged peptide-GlcNAc fragment is prominent at m/z 1257.65.

Madison, WI, USA), sequencing grade modified bovine chymotrypsin (Princeton Separation, Inc., Adelphia, NJ, USA), Lys-C and Asp-N (Roche Diagnostic, Mannheim, Germany). In addition, a sample of reduced and S-carboxymethylated TaPAPhy was deglycosylated by 60 mU of glycopeptidase A (PNGase A, Sigma–Aldrich, MO, USA) in 15 μL 0.1 mM ammonium acetate, pH 5 at 37 °C for 18 h, dried by vacuum centrifugation, and digested with chymotrypsin.

Aliquots of the digests were analyzed by a nanoflow capillary liquid chromatography system (Ultimate/Switchos/Farmos, LC-Packings, NL) interfaced directly to an ESI Q-TOF tandem mass spectrometer (MicroQ-TOF, Bruker Daltonics, DE) using a vented column setup (van der Heeft et al., 1998). Peptides were separated on a 10 cm long 50 μm id custom-packed C18 reversed-phase column of Reprosil-Pur C18-AQ, 3 μm particle size (Dr. Maisch GmbH,

DE) using a 35 min gradient of 0–45% acetonitrile (v/v) containing 0.6% (v/v) acetic acid and 0.005% (v/v) heptafluorobutyric acid in MilliQ water, at a flow rate of 175 nL min^{-1}. Mass spectra obtained by automated data-dependent acquisition modes were processed using the DataAnalysis v. 3.4 (Bruker Daltonics, DE).

Wheat protein sequences were translated from all available wheat EST sequences, (DFCI Wheat Gene Index, Release 12.0, July 24th, 2008 at http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb = wheat) and organized in a custom made wheat protein database by Dr. Jeppe Emmersen (Emmersen, 2007). The lists of MS1 and MS2 data from each type of digestion were analyzed and searched by Mascot software v2.2 (http://www.matrixsciences.com) and scored against the wheat protein database, or against the translated TaPAPhy sequences, which gave identical results for phytases. Search parameters for trypsin were: enzyme, semi-trypsin, allowing three missed cleavages; complete modification: carboxymethylated; partial modification: oxidized methionine; peptide tolerance: 0.1 Da, peptide ion score at least 15. The tolerance of 0.1 Da was also used for interpreting peptide fragment MS2 data. Settings for chymotrypsin, Lys-C, and Asp-N were similar. Allowing for a search of 'semi-protease' specificity permits identification of the N- and C-terminal peptides of the processed mature phytase. Chymotrypsin has a broader specificity than generally realized, i.e. cleavage after all large hydrophobic residues including Trp, Tyr, Phe, His, Met, and Leu.

Glycopeptide data was manually extracted from the raw MS2 spectra containing the glycan fragment ions 204.1 (GlcNAc) and 366.1 (GlcNAc-Man) using DataAnalysis ver 3.4 (Bruker Daltonics, Bremen, DE). The corresponding MS1 spectra were scanned over the corresponding glycopeptide peak eluting from the on-line LC column into the mass spectrometer. The GPMAW software (www.gpmaw.com) was used to search for candidate glycopeptides. The Mascot searches of all digests were redone allowing Asn-GlcNAc as a variable modification of Asn. The MS2 of this type of glycopeptide usually gave mediocre spectra with a low peptide ion score. Error tolerant searches were performed to identify deglycosylated asparagines in the digest of a glycopeptidase A-treated sample, as deglycosylation will change Asn to Asp.

Acknowledgements

We thank Dr. Jeppe Emmersen for creating the wheat protein database. The work was supported by Grant 274-06-0331 to M.J. from the Danish Research Council for Technology and Production.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2010.12.023.

References

Bendtsen, J.B., Nielsen, H., von Heijne G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol., 340:783–795.
Brinch-Pedersen, H., Olesen, A., Rasmussen, S.K., Holm, P.B., 2000. Generation of transgenic wheat (*Triticum aestivum* L.) for constitutive accumulation of an *Aspergillus phytase*. Mol. Breed 6, 195–206.
Brinch-Pedersen, H., Sorensen, L.D., Holm, P.B., 2002. Engineering crop plants: getting a handle on phosphate. TIPS 7, 118–125.
Dionisio, G., Madsen, C.K., Holm, P.B., Welinder, K.G., Jørgensen, M., Stoger, E., Arcalis, E., Brinch-Pedersen, H., 2011. Cloning and characterization of purple acid phosphatase phytases from wheat, barley, maize, and rice. Plant Physiol. 147, in press.
Dvorakova, J., 1998. Phytase: sources, preparation and exploitation. Folia Microbiol (Praha) 43, 323–338.
Emmersen, J., 2007. Generating unigene collections of expressed sequence tag sequences for use in mass spectrometry identification. Methods Mol. Biol. 367, 77–86.
Gavel, Y., von Heijne, G., 1990. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. Protein Eng. 3, 433–442.
Gray, J.S.S., Yang, B.Y., Montgomery, R., 1998. Heterogeneity of glycans at each N-glycosylation site of horseradish peroxidase. Carbohydr. Res. 311, 61–69.
Kasturi, L., Chen, H., Shakin-Eshleman, S.H., 1997. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem. J. 323, 415–419.
Kornfeld, R., Kornfeld, S., 1985. Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. 54, 631–664.
Lei, X.G., Porres, J.M., Mullaney, E.J., Brinch-Pedersen, H., 2007. Phytase: source, structure, and application. In: Polaina, J., MacCabe, A.P. (Eds.), Industrial Enzymes: Structure Function and Applications. Springer, Dordrecht, pp. 505–530.
Lott, J.N.A., 1984. Accumulation of seed reserves of phosphorus and other minerals. In: Murray, D.R. (Ed.), Seed Physiology. Academic Press, New York, pp. 139–166.
van der Heeft, E., ten Hove, G.J., Herberts, C.A., Meiring, H.D., van Els, C.A., de Jong, A.P., 1998. A microcapillary column switching HPLC-electrospray ionization MS system for the direct identification of peptides presented by major histocompatibility complex class I molecules. Anal. Chem. 70, 3742–3751.
Welinder, K.G., Jørgensen, M., 2009. Covalent structures of potato tuber lipases (patatins) and implications for vacuolar import. J. Biol. Chem. 284, 9764–9769.
Whitley, P., Nilsson, I., von Heijne, G., 1996. A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. J. Biol. Chem. 271, 6241–6244.
Wilson, I.B.H., 2002. Glycosylation of proteins in plants and invertebrates. Curr. Opin. Struc. Biol. 12, 569–577.
Wuhrer, M., Catalina, M.I., Deelder, A.M., Hokke, C.H., 2007. Review glycoproteomics based on tandem mass spectrometry of glycopeptides. J. Chromatogr. B 849, 115–128.