ELSEVIER

Investigation of cationic peanut peroxidase glycans by electrospray ionization mass spectrometry

Cunjie Zhang [a], Amanda Doherty-Kirby [a], Robert van Huystee [b,*], Gilles Lajoie [a,*]

[a] *Department of Biochemistry, The University of Western Ontario, London, Ont., Canada N6A 5C1*
[b] *Department of Biology, The University of Western Ontario, London, Ont., Canada N6A 5B7*

Abstract

Cationic peanut peroxidase (CP) was isolated from peanut (*Arachis hypogaea*) cell suspension culture medium. CP is a glycoprotein with three N-linked glycan sites at Asn60, Asn144, and Asn185. ESI-MS of the intact purified protein reveals the microheterogeneity of the glycans. Tryptic digestion of CP gave a near complete sequence coverage by ESI-MS. The glycopeptides from the tryptic digestion were separated by RP HPLC identified by ESI-MS and the structure of the glycan chains determined by ESI-MS/MS. The glycans are large structures of up to 16 sugars, but most of their non-reducing ends have been modified giving a mixture of shorter chains at each site. Good agreement was found with the one glycan previously analyzed by ^{1}H NMR. This work is the basis for the future studies on the role of the glycans on stability and folding of CP and is another example of a detailed structural characterization of complex glycoproteins by mass spectrometry.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: *Arachis hypogaea*; Cationic peanut peroxidase; Electrospray ionization mass spectrometry; Tandem mass spectrometry; LC MS/MS; Proteomics; Glycobiology; Glycomics; Glycans; Glycopeptides

1. Introduction

The concept of proteomics implies that proteins relate to genomic expression in cells (Zivy and de Vienne, 2000; Kvasnička, 2003). However, the genome is not accurately reflected in the proteome for two reasons: (1) mRNA levels and protein levels are not well correlated and (2) many proteins are post-translationally modified (i.e., phosphorylated, glycosylated, removal of signal peptides, oxidation, etc.). In mammals, *N*-glycosylation has been shown to be involved in many physiological functions. These include, but are not limited to, protein folding, biological activity, protein stability, and immunogenicity (Zeleny et al., 1999). It is likely that in plants, *N*-glycosylation will have similar biological functions.

Cationic isozyme peanut peroxidase (CP) (EC 1.11.1.7) is *N*-glycosylated at three sites, but the glycan chains have never been fully characterized. The active enzyme can be obtained in abundant (mg) quantities from peanut (*Arachis hypogaea*) cell suspension culture medium, and it has been studied in detail with respect to its function and its structure (van Huystee et al., 2002). Overall, CP has been shown to be stable (van Huystee and Zheng, 1995; Lige et al., 2001) requiring heme (Chibbar et al., 1984) and Ca^{2+} for optimal activity (Rodriguez Maranon et al., 1993). Enzymatic deglycosylation of CP with PNGase F results in lost activity and increased susceptibility to trypsinolysis presumably due to loss of some sugars from the glycan chain (Hu and van Huystee, 1989; van Huystee and Wan, 1994). Treatment of cell culture with tunicamycin, a glycosylation inhibitor, results in decreased secretion of CP into the medium likely due to degradation of the nonglycosylated protein within the cell (van Huystee et al., 2002; Ravi et al., 1986). CP is *N*-glycosylated at Asn60 (GPb), Asn144 (GPc), and Asn185 (GPa). Single

mutagenesis of each of the relevant Asn residues to Gln showed decreased activity in the N60Q and N144Q mutants with N185Q CP showing no loss of activity. Loss of the glycan at Asn185 decreases thermostability when compared to the other mutants and the wild-type enzyme. The loss of a single glycan did not affect the degradation of the protein as studied by trypsinolyis, but did show varying effects on the folding of the enzyme (Lige et al., 2001). Thus, the evidence indicates that the glycans of CP are functionally important. A formal investigation of the glycans of CP is required for several reasons: (1) to study the effect of sugar loss on the stability of the enzyme, (2) to compare glycan structure at different stages of development, and (3) to investigate the glycosidases secreted with this enzyme.

Previous structural studies on CP could not fully characterize the glycans of this enzyme. An X-ray diffraction analysis confirmed the sites of glycosylation, but due to poor density and high B factors was unable to solve the structure of the glycans (Schuller et al., 1996). The glycan released from Asn144 has been studied by ^{1}H NMR spectroscopy (Shaw et al., 2000). This study was handicapped by the microheterogeneity of CP that was presumably due to the presence of glycosidase in the cell wall and the medium supporting the cell culture (Wan et al., 1994). This present study uses high performance liquid chromatography (HPLC) and electrospray ionization mass spectrometry (ESI-MS) to investigate the structure and microheterogeneity at all three sites.

2. Results and discussion

2.1. Protein purification and mass analysis

Isolation of CP is quite simple due to the relatively few proteins found in the spent medium of the peanut cell suspension culture (van Huystee and Tam, 1988; Sesto and van Huystee, 1988). However, HPLC analysis showed minor components corresponding to other proteins and this could complicate the analysis of CP at the peptide level. Therefore, to ensure as homogeneous sample as possible, 500 μg of CP was further purified by reversed phase HPLC. A single peak eluting at 27–29 min was collected. This purification step also removed the salts as well as the heme from the sample and thus enabled ESI-MS analysis of the intact protein.

ESI-MS analysis of the intact CP under denaturing conditions showed numerous components present due to the heterogeneity of the glycans on the protein (Fig. 1). Deconvolution of the raw data using maximum entropy techniques (MaxEnt, Ferrige et al., 1992) gave masses ranging from 33.0 to 37.8 kDa for intact CP. These results are consistent with earlier SDS-PAGE results of 40 kDa for CP (O'Donnell et al., 1992). The total sugar composition of the three glycans of CP has

been determined by TFA hydrolysis, chemical derivatization and HPLC (Sun et al., 1997). The structure of the Asn144-linked glycan (Shaw et al., 2000) has been determined by NMR. From the composition analysis (Sun et al., 1997), five different sugars were identified for all three glycans: N-acetylglucosamine (N, monoisotopic mass 203.079 Da), the deoxyhexose, fucose (F, 146.058 Da), the pentose, xylose (X, 132.042 Da), and the two hexoses, mannose and galactose (H, 162.053 Da) which are indistinguishable using the MS techniques described in this study. The heterogeneity of the intact protein molecular masses can be accounted for by variations in one or more of these sugar moieties (Fig. 1(b)).

2.2. Isolation of glycopeptides from the trypsin digest of CP

Trypsin digestion of CP generates fragments within an acceptable mass range for the type of analysis presented here and the glycopeptides were separated from the non-glycosylated peptides (Fig. 2) using reverse phase HPLC on a C18 analytical column. The extent of the tryptic digestion of CP was first verified by LC-MS using a microbore precolumn to remove salts and partially separate the peptide mixture. Based on the MS survey spectrum (Fig. 3(a)) and MS/MS data acquired from this and from analysis of the fractions from analytical purification, near complete coverage of CP was observed (Fig. 3(b)). The N-terminal tryptic peptide contained a Gln residue at the amino terminus which cyclized to pyroglutamic acid corresponding to a loss of 17 Da. This is a well documented modification that can occur from sample handling (Khandke et al., 1989). Only a few peptides were not detected in the mass spectra. These were very short peptides (GFTTK, TR, NKK, TNCRK, TN) resulting from tryptic cleavage at proximal sites in the sequence.

The glycopeptide-containing fractions were identified on a Q-TOF (Q-TOF Micro, Micromass, UK) operating in precursor ion discovery (PID) mode (Ritchie et al., 2003). In the PID experiment the voltage on the collision cell is rapidly switched between low and high voltage to obtain spectra of intact and fragmented peptides, respectively. Due to the very labile nature of glycosidic bonds, the loss of carbohydrate specific ions such as (Hex)$^{1+}$, m/z 163.060, (HexNAc)$^{1+}$, m/z 204.084, (HexHexNAc)$^{1+}$, m/z 366.139 is characteristic of a glycopeptide. Using the PID protocol, five fractions containing glycopeptides were found and corresponded to the three glycopeptide groups. GPa (Asp185 linked) eluted in fraction 19 (18–19 min), GPb (Asn65 linked) was collected in fractions 31 and 32 (30–32 min), and GPc (Asn144 linked) eluted in fractions 45 and 46 (44–46 min) (Fig. 2). Each of the eluting glycopeptide fractions were relatively broad reflecting the heterogeneity
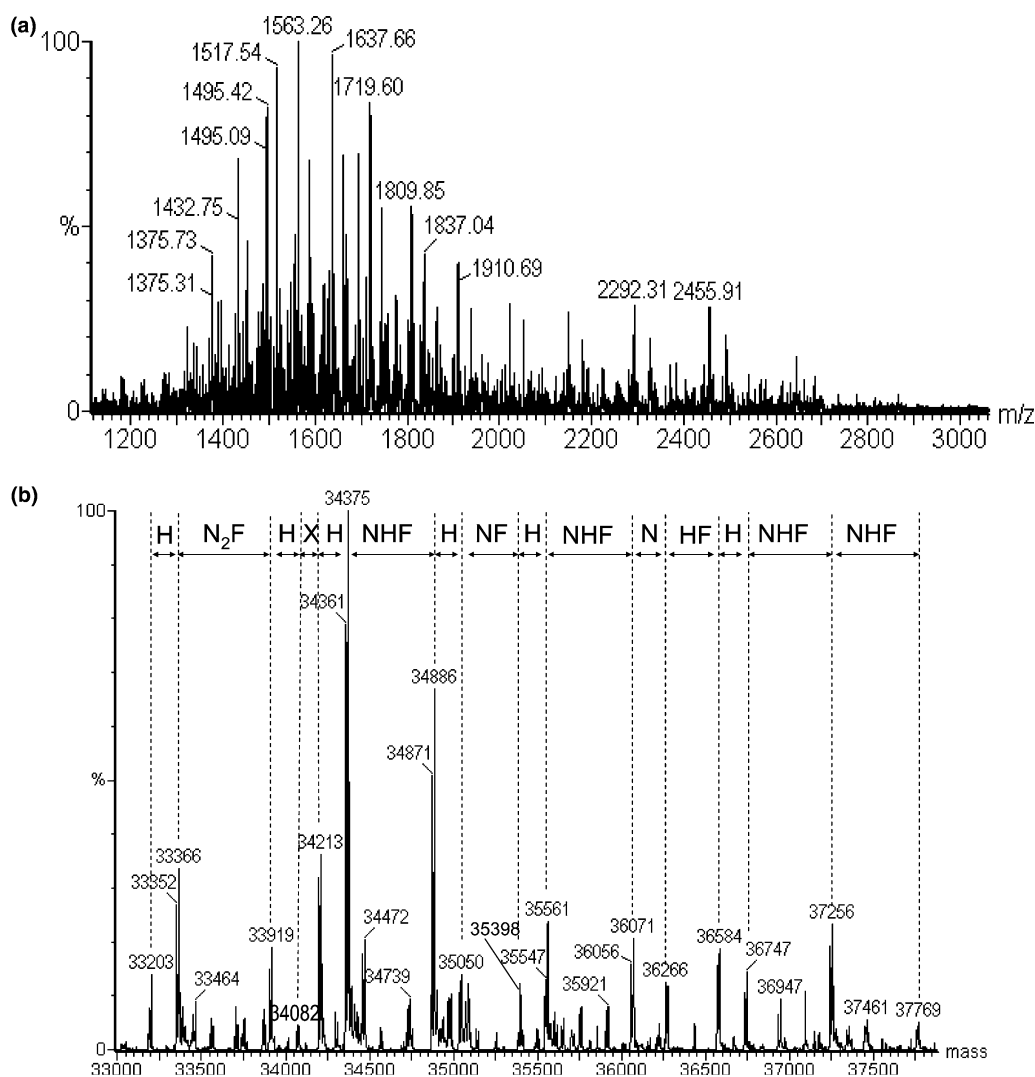
Fig. 1. (a) Base-line subtracted and (b) MaxEnt1 processed ESI-MS data for the HPLC purified CP. Observed masses ranged from 32.0 to 37.8 kDa. The masses between individual species correspond to various sugars (F = fucose, H = hexose, N = N-acetylglucosamine, X = xylose).
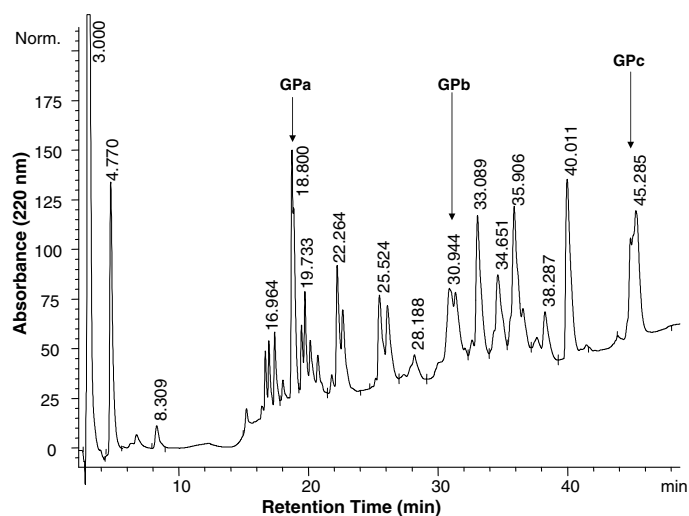

Fig. 2. HPLC chromatogram of the tryptic digest of CP. The glycopeptides are labeled GPa, GPb, and GPc according to the time of elution. GPa was collected in fraction 19, GPb in fractions 31 and 32, and GPc in fractions 45 and 46.
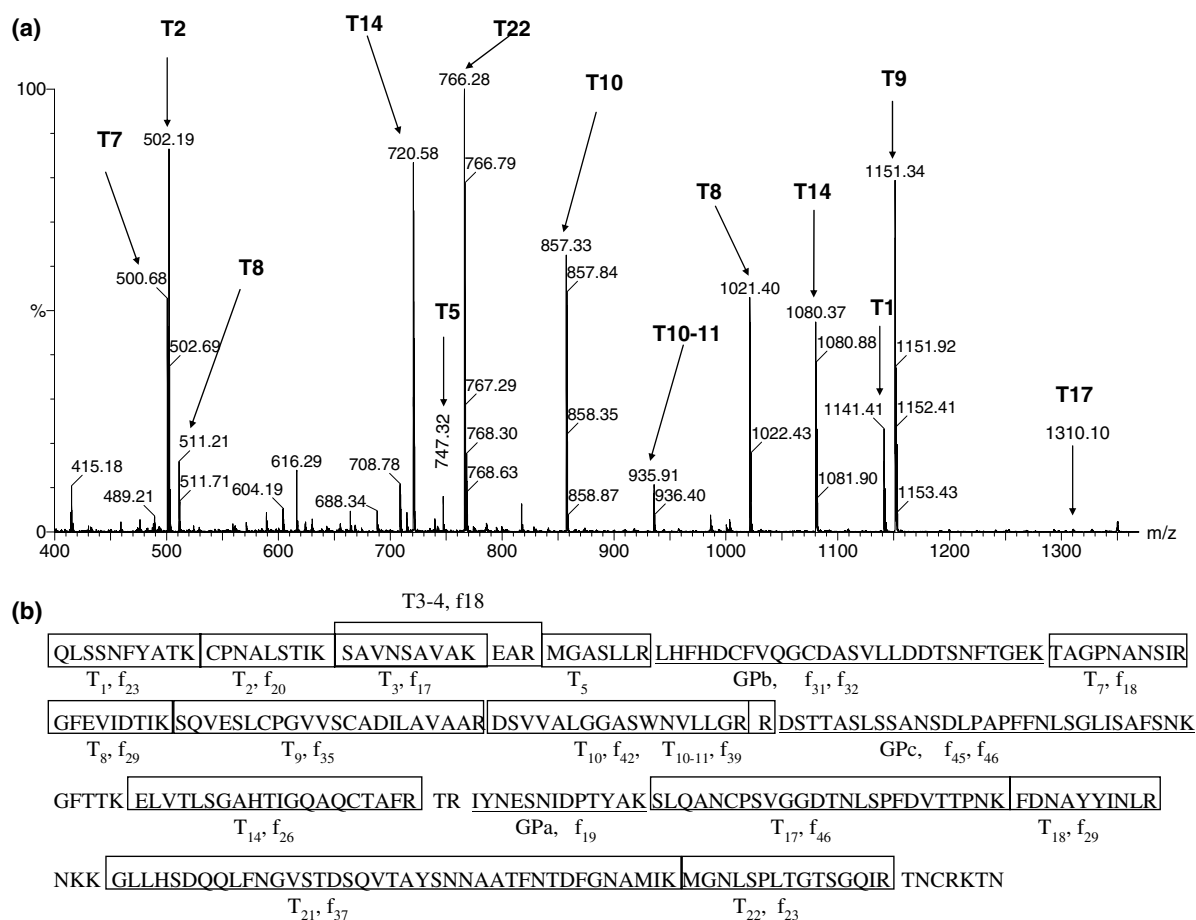
Fig. 3. (a) Survey MS data for the tryptic digest of CP. Peptides labelled correspond to those expected from the theoretical digest of CP (e.g., T2 is the second peptide from the N-terminus that is expected from this digest). Peptides that are not shown here, but were shown in the various analytical HPLC fractions are indicated in (b). (b) Coverage map for CP taking into account the survey data from (a) and the data obtained from HPLC fractions. T corresponds to the tryptic fragment from a theoretical digest while f indicates the HPLC fraction number. T1 was observed as the pyroglutamic acid derivative. In total, the peptides observed corresponded to 92% of the expected protein sequence.

of the glycan chain at each position. This was confirmed later by MS/MS analysis. Taken together these results indicate the complete removal of the leader peptide sequence as well as confirmation of the protein sequence. This also allowed correlation of the tryptic peptides and glycopeptides to the intact protein results as discussed later.

2.3. ESI-MS analysis of GPa, GPb, and GPc

Each of the five glycopeptide fractions was analysed by nano-ESI-MS with a hybrid quadrupole time-of-flight mass spectrometer (Q-TOF2, Micromass, UK) using borosilicate tips. The nano-source allowed for extended spraying of small amounts of the sample so that instrumental parameters could be optimized. From the experimentally determined masses, the possible oligosaccharide compositions were determined using GlycoMod (Cooper et al., 2001; http://www.expasy.ch/tools/glycomod/) with the known monosaccharide

components from the hydrolysis studies. Fig. 4 shows the non-deconvoluted and MaxEnt3 processed MS spectrum for GPb (fraction 32) from CP. The spectrum reveals several glycopeptides. The most abundant glycopeptides for GPb were labelled and range in mass from 3826 to 5304 Da. The theoretical monoisotopic mass for the non-glycosylated, doubly carbamidomethylated tryptic peptide corresponding to GPb is 3111 Da. Based on the added mass combination of the sugar units and the theoretical peptide mass, the output from GlycoMod gave monosaccharide composition in terms of the number of hexose, fucose, xylose, and N-acetylglucosamine. However, GlycoMod does not give any information on the actual position or type (α or β) linkage of the sugar units within the glycan chain. The most abundant glycopeptides for GPa and GPc range in mass from 2697 to 3490 Da and from 4168 to 5866 Da, respectively (data not shown). The theoretical monoisotopic masses for the non-glycosylated peptide of GPa and GPc are 1527 and 3159 Da.
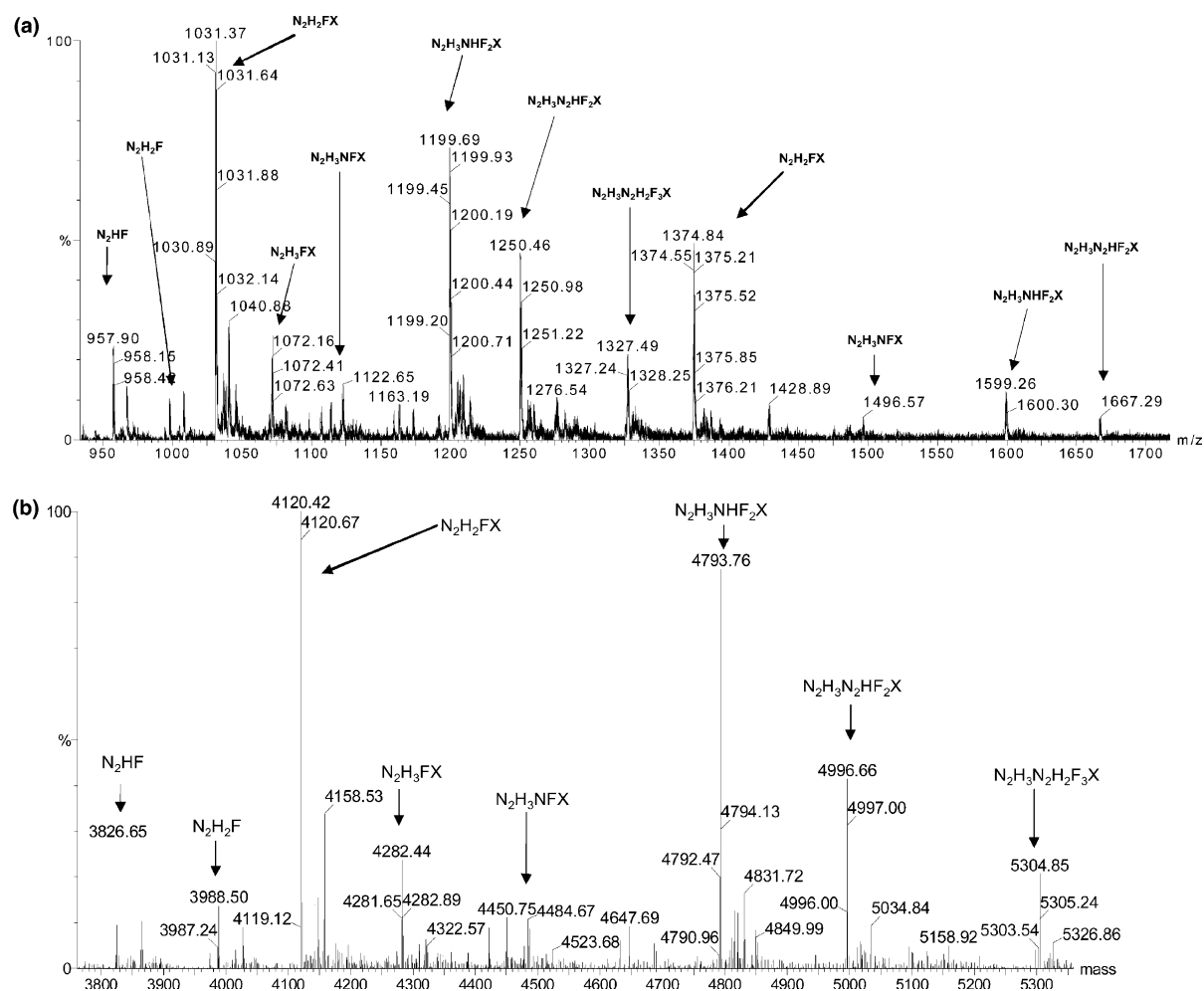
Fig. 4. ESI-MS spectra of GPb (fraction 32) with glycans labelled. Species corresponding to monovalent cation (i.e., NH_4^+, Na^+, and K^+) adducts are not labelled. (a) Baseline subtracted m/z data showing multiply charged species. (b) MaxEnt3 processed data showing monoisotopic MH^+ species. The mass of GPb ranges from 3826 to 5304 Da.

2.4. Tandem mass spectrometry (MS/MS) analysis of CP glycopeptides

The glycopeptide precursor ions observed for GPa, GPb, and GPc were each subjected to ESI-MS/MS analysis. Although many studies on glycopeptides are performed on the glycan released from the glycopeptide, the MS study of intact glycopeptides has several advantages. First, sample handling is minimized as enzymatic deglycosylation and subsequent derivatization of the glycans is not needed. While fragmentation of the glycan is more favoured, fragmentation of the peptide backbone also occurs allowing determination of the peptide sequence as well as the site of glycosylation (Dell and Morris, 2001). Analysis of the MS/MS spectra for the various glycopeptides confirmed the sequence of the peptides to be GPa: $IYN_{185}ESNIDPTYAK$, GPb: LHFHDCFVQG $CDASVLLDDTSN_{60}FTGEK$, and GPc: DSTTASLS-$SANSDLPAPFFN_{144}LSGLISAFSNK$.

MS/MS experiments yielded critical information about the structure of the glycan chains. The MS/MS data, together with additional results from the composition and the preliminary NMR studies were used to determine the glycan structure for each of the glycopeptides in GPa, GPb, and GPc. For each of these glycopeptides a number of glycan structures were present. The deduction of these structures was based on the following: (1) the relative composition glycans (Sun et al., 1997); (2) the sugar combinations predicted by GlycoMod for each glycopeptide with the calculated mass of the glycan or the observed mass of the glycopeptide (Cooper et al., 2001); (3) the B and Y fragments ions which were identified manually from the MS/MS spectra; and (4) the core structure of GPc as determined by NMR (Shaw et al., 2000).

Two examples of MS/MS spectra from GPb-derived glycopeptides and their corresponding structures are given in Figs. 5 and 6. The supporting MS/MS data are
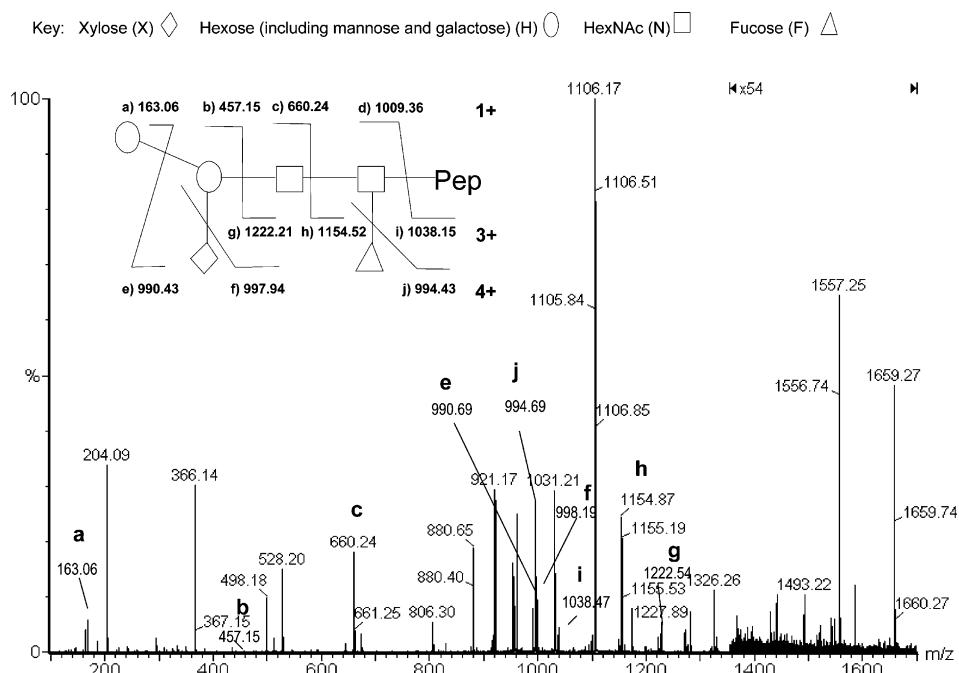
Fig. 5. ESI-MS/MS data for m/z 1030 (4$^+$) species of GPb. The interpreted glycan structure is also shown. Glycan specific fragments confirming the assignment of this structure are labelled in the spectrum. The summary of the MS/MS data is found in Table 1.
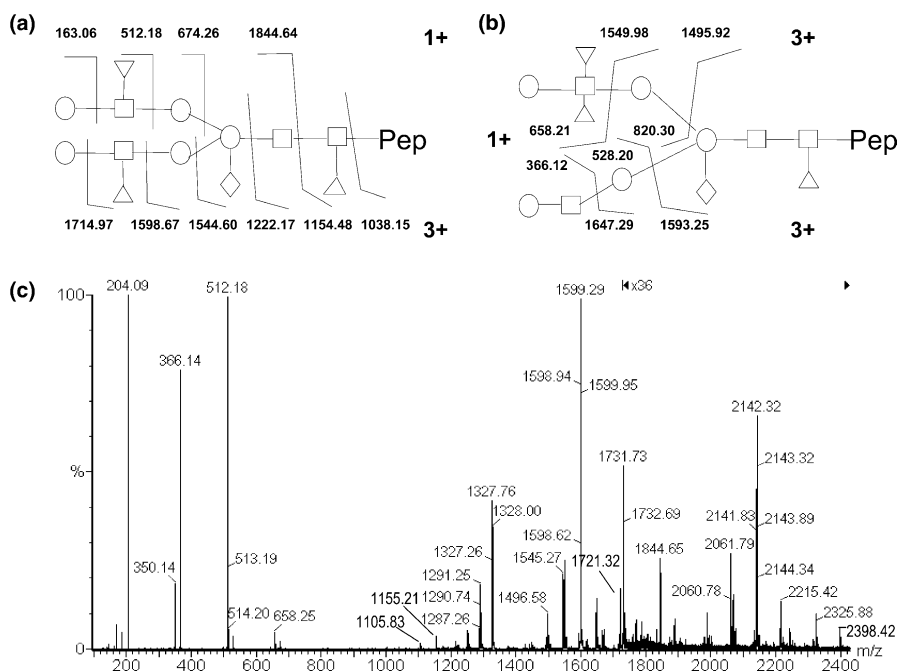
Fig. 6. Major deduced structures for the 1327 (4$^+$) species of GPb (a,b) and (c) the corresponding spectrum. The MS/MS data is summarized in Table 2.

summarized in Tables 1 and 2. Fig. 5 and Table 1 describe the MS/MS spectra for the one of the smallest observed glycopeptides from the GPb fraction (precursor ion, m/z 1030 (4$^+$)) (Table 3, Structure III). The MS/MS data shown in Fig. 6c and summarized in Table 2 correspond to larger, more complex glycans also derived

Table 1A
Tabulated data corresponding to glycan and peptidic species for the MS/MS fragmentation of m/z 1030 (4^+ state) of GPb and assignments

#	Assigned glycan	Expected mass	Observed glycan fragments (m/z)		Observed peptide fragments		
			MH$^+$	Relative intensity (RI)	Assignment		m/z
1	H	162.05	163.06	14	b-ions	y-ions	
2	N	203.08	204.09	112		y3^{1+}	333.19
3	HX	294.09	295.10	9		y4^{1+}	434.23
4	H2	324.10	325.12	2		y5^{1+}	581.3
5	NX	335.12	336.13	1	b9^{2+}		592.77
6	NF	349.14	350.15	4	b10^{2+}		621.28
7	NH	365.13	366.14	100	b11^{2+}		701.27
8	H2X	456.15	457.15	0.8	b12^{2+}		758.8
9	NHX	497.17	498.18	33	b13^{2+}		794.33
10	NHF*	511.19	512.20	9	b14^{2+}		837.87
11	NH2	527.18	528.20	50	b15^{2+}		887.38
12	N2H	568.21	569.19	0.5	b16^{2+}		943.92
13	NHFX*	643.23	644.25	5	b17^{2+}		1000.46
14	NH2X	659.23	660.24	60	b18^{2+}		1057.98
15	NH2F*	673.24	674.25	11	b19^{2+}		1115.5
16	N2HF	714.27	715.30	<0.2	b20^{2+}		1166.03
17	N2H2	730.26	731.31	0.4			
18	NH2FX*	805.28	806.30	18		pep^{2+}	1556.7
19	N2H2X	862.30	863.32	0.6		Pep^{3+}	1038.15
20	N2H2FX	1008.36	1009.36	1			

Data marked with * support a structure with fucose on the GlcNAc two residues from the backbone Asn. The body of the data support the glycan structure given in Fig. 5.

Table 1B
Glycopeptide data observed and the respective assigned species

#	Assigned species	Expected mass	Observed Y-ions of glycopeptide				Non-reducing end
			2$^+$	3$^+$	4$^+$	RI	
1	pep	3111.37	1556.74	1038.15		19	
2	pep-N	3314.45	1658.27	1105.84	829.63	336	
3	pep-NF	3460.51	1731.33	1154.52		81	
4	pep-N2	3517.53		1173.54	880.40	88	
5	pep-N2F	3663.59		1222.21	916.92	19	
6	pep-N2H	3679.58		1227.55	920.90	116	
7	pep-N2HX	3811.62		1271.58	953.92	67	
8	pep-N2HF	3825.64		1276.23	957.43	33	
9	pep-N2H2	3841.63		1281.56	961.43	107	
10	pep-N2HFX	3957.68		1320.27	990.43	29	H
11	pep-N2H2X	3973.68		1325.59	994.43	128	F
12	pep-N2H2F	3987.69			997.94	46	P
13	pep-N2H2FX	4119.73			1030.96	97	

from GPb (m/z 1327 (4^+)) (Table 3, Structures XIa and XIb, Fig. 6a and b). Structures XIa and XIb differ by the position of one fucose residue.

From Tables 2A and 2B, the most intense fragments (512.18 (1^+), 1598.67 (3^+)) as well as other strong fragments (674.26 (1^+), 1544.60 (3^+)) overwhelmingly support the structure shown in Fig. 6(a) while fragments at 1549.94 (3^+), 1647.29 (3^+), and 658.21 (1^+) correspond to the structure of shown in Fig. 6(b). Tables 2A and 2B list the fragment ions which support these structures. It is worth noting that several fragments ions (e.g., 204 (1^+), 366 (1^+), 528 (1^+), 1038 (3^+), 1155 (3^+), 1174 (3^+), 1222 (3^+), 1731 (2^+)) were common to both the smaller (III) and larger (XIa, XIb) glycopeptides

indicating that these species have the same peptide backbone and the same core structure for the N-linked glycan. Using this approach, the glycan structures GPa, GPb, and GPc were deduced and are presented in Table 3.

2.5. Assignment of the glycopeptide on the intact protein

One benefit of this study over several other structure elucidation of glycoproteins is the observation of masses of the intact proteins as well as characterization of most glycoforms at each site. It is therefore straightforward to assign the major masses observed in mass spectrum of the intact protein by adding the deduced glycan (average

Table 2A

Tabulated data corresponding to glycan and peptidic species for the MS/MS fragmentation of m/z 1327 (4^+ state) of GPb and assignments (2A)

#	Assigned glycan	Expected mass	Observed glycan fragments (m/z)		Observed peptide fragments	
			MH $^+$	Relative intensity (RI)	Assignment	m/z
1	H	162.05	163.06	1		
2	N	203.08	204.09	127	y^{3+}	333.19
3	H2	324.10	325.10	<0.2	y^{4+}	434.23
4	NX	335.12	336.13	0.5	y^{5+}	581.3
5	NF	349.14	350.14	24	$b9^{2+}$	592.77
6	NH	365.13	366.14	100	$b10^{2+}$	621.28
7	NF2	495.19	496.20	0.5	$b11^{2+}$	701.27
8	NHX	497.17	498.18	1	$b12^{2+}$	758.8
9	NHF	511.19	512.18	126	$b13^{2+}$	794.33
10	NH2	527.18	528.20	5	$b14^{2+}$	837.87
11	N2H	568.21	569.22	<0.2	$b15^{2+}$	887.38
12	NHFX	643.23	644.25	0.6	$b16^{2+}$	943.92
13	NHF2	657.25	658.21	6	$b17^{2+}$	1000.46
14	NH2X	659.23	660.24	0.8	$b18^{2+}$	1057.98
15	NH2F	673.24	674.26	3	$b19^{2+}$	1115.5
16	NH3	689.23	690.21	0.3	$b20^{2+}$	1166.03
17	N2HF	714.27	715.25	<0.2	pep^{3+}	1038.15
18	N2H2	730.26	731.23	<0.2		
19	NH2FX	805.28	806.29	<0.2		
20	NH2F2	819.30	820.30	0.4		
21	NH3X	821.28	822.26	0.5		
22	NH3F	835.29	836.28	<0.2		
23	N2H2F	876.32	877.33	<0.2		
24	N2H3	892.31	893.32	<0.2		
25	NH3XF	967.33	968.34	0.25		
26	N2H3X	1024.36	1025.37	0.3		
27	NH4FX	1129.39	1130.35	0.22		
28	N2H3FX	1170.41	1171.40	0.4		
29	N2H4X	1186.41	1187.45	0.4		
30	N2H4FX	1332.47	1333.44	1.4		
31	N3H5F2X	1843.65	1844.65	0.9		

mass) at each site to the average mass of the non-modified protein (31,196 Da minus 8H for the four disulfide bonds) (Table 4). For example, the mass at 34,375 corresponds to the addition of glycan IV at N_{185} and glycan III at N_{60} and N_{144}. The mass at 34, 886 corresponds to the addition of glycan VIII at N_{185} and glycan III at N_{60} and N_{144}. The mass at 37,256 corresponds to the addition of glycan XI at N_{185} and N_{144} and glycan VIII at N_{60} or glycan VIII at N_{60} and N_{185} and glycan XIII at N_{144}. The major species present in the MS spectrum of the intact protein can be explained by various combinations of observed glycans (Table 4). The smallest reported glycoprotein can only be rationalized by the absence of a glycan at N_{144}. It has been observed for horseradish peroxidase (HRP) that full glycosylation does not occur at all potential glycosylation sites (Gray et al., 1998).

2.6. Correspondence of GPc with the NMR structure

The structure of a glycan isolated from Asn144 (GPc) has been studied by NMR. The structure reported by Shaw et al. (2000) contains only 11 sugar residues and could be derived by the action of glycosidases on the largest GPc structure described in Table 4. Glycosidases are present in the spent medium of peanut (Wan et al., 1994) and microheterogeneity was indeed found in the sample used for NMR (Shaw et al., 2000). Therefore, only the major resonances and NOEs were used to assign the reported structure and minor signals that might correspond to heterogeneous ends were not used (G. Shaw, personal communication). Mass spectrometry is a more selective technique in that all of the components (i.e., not just the major one) of a mixture can be detected if they ionize similarly. It should be noted that the NMR structure did not agree with previous composition results (Sun et al., 1997) which indicated four GlcNAc were present. The longer structures presented in Table 3 accounts for the differences in NMR structure and those from compositional analysis.

2.7. Comparison of the individual sites

It is clear that there is a similar amount of heterogeneity at each of the three sites of glycosylation of CP and that the glycans of GPc are generally larger than those from GPb or GPa. However, the structures for

Table 2B
Glycopeptide data observed and the respective assigned species
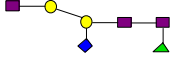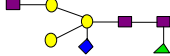
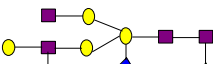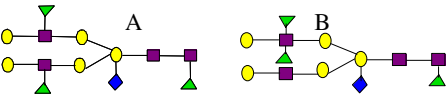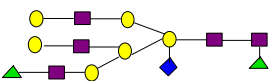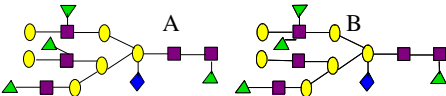| #	Assigned species	Expected mass	Observed Y-ions of glycopeptide				Non-reducing end
			2+	3+	4+	RI	
1	pep	3111.37		1038.1		0.5	
2	pep-N	3314.45	1658.21	1105.80		3.5	
3	pep-NF	3460.51	1731.20	1154.48		6.4	
4	pep-N2	3517.53		1173.46		0.9	
5	pep-N2F	3663.59	1832.80	1222.17		1.7	
6	pep-N2HFX	3957.68	1979.76			0.1	
7	pep-N2H2X	3973.68	1987.76			0.4	
8	pep-N2H2F	3987.69	1994.69			0.1	
9	pep-N2H3	4003.69	2002.77		1002.00	0.8	
10	pep-N2H2FX	4119.73	2060.78	1374.25		2.1	
11	pep-N2H3X	4135.73	2068.83	1379.50		1.2	
12	pep-N2H3F	4149.74	2075.86	1384.26		0.5	
13	pep-N3H2X	4176.76		1393.26		0.5	
14	pep-N3H2F	4190.77		1397.98		0.4	
15	pep-N3H3	4206.77		1403.22		0.5	
16	pep-N2H2F2X	4265.79		1422.89		0.4	
17	pep-N2H3FX	4281.79	2141.92	1428.23	1071.75	4.6	
18	pep-N3H2FX	4322.81		1441.92		1.6	
19	pep-N3H3X	4338.81	2170.28	1447.24		2.5	
20	pep-N3H3F	4352.82		1451.91		1.3	
21	pep-N3H4	4368.82		1457.28		0.5	
22	pep-N2H3F2X	4427.84	2214.93	1476.96		1.0	
23	pep-N3H2F2X	4468.87		1490.63		1.5	
24	pep-N3H3FX	4484.86	2243.32	1495.92		12.7	
25	pep-N3H3F2	4498.88		1500.62		2.4	
26	pep-N3H4X	4500.86		1501.29		2.4	
27	pep-N3H4F	4514.87		1505.93		1.9	
28	pep-N3H3F2X	4630.92	2316.38	1544.60		26.9	
29	pep-N3H4FX	4646.92	2324.33	1549.94		32.2	
30	pep-N3H4F2	4660.93		1554.58		4.3	
31	pep-N4H4X	4703.94		1568.95	1176.70	1.7	
32	pep-N3H3F3X	4776.98		1593.25		5.5	
33	pep-N3H4F2X	4792.97	2397.38	1598.67		125.7	
34	pep-N3H4F3	4806.99		1603.23		7.1	
35	pep-N4H3F2X	4834.00		1612.33	1209.49	1.4	
36	pep-N4H4FX	4850.00		1617.62	1213.48	6.3	
37	pep-N4H4F2	4864.01		1622.27	1217.00	1.1	
38	pep-N4H5X	4865.99		1623.00	1217.47	1.1	
39	pep-N4H5F	4880.01		1627.66		0.4	
40	pep-N3H4F3X	4939.03		1647.29		18.1	
41	pep-N3H5F2X	4955.03		1652.65		4.1	
42	pep-N4H3F3X	4980.06			1246.00	0.5	
43	pep-N4H4F2X	4996.05		1666.31	1249.97	13.2	
44	pep-N4H4F3	5010.07		1671.65	1253.97	12.5	G
45	pep-N4H5F2	5026.06		1676.30	1257.48	2.3	
46	pep-N3H5F3X	5101.08		1701.28		1.5	N
47	pep-N4H4F3X	5142.11		1714.97	1286.51	8.9	
48	pep-N4H5F2X	5158.10		1720.32	1290.50	44.6	F
49	pep-N4H5F3	5172.12			1294.00	3.4	P
50	pep-N4H5F3X	5304.16			1327.05	53.0	
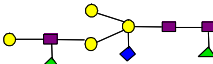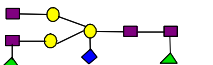
Data support two major structures as shown in Fig. 6.

each set of glycans share several common features such as fucose attached to the GlcNAc residue linked to Asn and the majority contain fucose, xylose, hexose, and GlcNac indicating that they are all complex type (Lerouge et al., 1998). This is in contrast to the glycans reported for HRP where a single site of the eight reported has high mannose containing structures (Gray et al., 1998). Of the eight sites of HRP, two are more heterogeneous than the remaining six. The heterogeneity of each of the CP glycans could arise from either incomplete processing or partial modification by glycosidases and glycosyltransferases in the Golgi apparatus or through degradation by exoglycosidases (Lerouge et al., 1998).

Table 3
Summary of proposed glycan structures for CP

	Proposed glycan structure	Abbreviation	Glycan mass monoisotopic (average)	Main species observed at m/z (charge states)		
				GPa	GPb	GPc
				m/z (CS) RI	m/z (CS) RI	m/z (CS) RI
I		N_2HF	714.27 (714.67)		958 (4+)**	
II		N_2H_2F	876.32 (876.82)		998 (4+)*	
III		N_2H_2FX	1008.36 (1008.93)		1031 (4+)**** 1374 (3+)	1390 (3+)**** 1397 (3+,Na)
IV		N_2H_3FX	1170.41 (1171.08)	1350 (2+)***	1070 (4+)**	
V		N_2H_2NFX	1211.44 (1212.13)			1457 (3+) 1465 (3+, Na)****
VI		N_2H_3NFX	1373.49 (1374.27)	1451 (2+)* 1462 (2+,Na)	1122 (4+)* 1496 (3+)	1519 (3+, Na)*
VII		$N_2H_3NF_2X$	1519.54 (1520.41)			1568 (3+, Na)**

VIII		$N_2H_3NHF_2X$	1681.60 (1682.56)	1070 (3+)**** 1605 (2+)	1199 (4+)**** 1598 (3+)	
IX		$N_2H_3N_2 F_2X$	1722.62 (1723.61)			1636 (3+, Na)**
X		$N_2H_3 N_2HF_2X$	1884.68 (1885.75)	1139 (3+)**** 1707 (2+)	1251 (4+)*** 1666 (3+)	1689 (3+, Na)**
XI	A B	$N_2H_3 N_2H_2F_3X$	2192.79 (2194.04)	1241 (3+)****	1327 (4+)**	1339 (4+)**** 1349 (4+, K) 1084 (5+, Na, K)
XII		$N_2H_3 N_3H_3 F_2X$	2411.86 (2413.23)	1326 (3+,K)*		
XIII	A B	$N_2H_3N_3 H_3F_4X$	2703.97 (2705.52)			1472 (4+)** 1186 (5+, 3Na) 1963 (3+, Na)

The relative intensity (RI) is a qualitative description based on the intensity of the observe m/z in the mass spectrum. The most abundant species are labeled with **** and the least abundant species by *.

Table 4
Molecular masses of the intact protein forms from experimental MS data (see Fig. 1) and calculated from various combinations of proposed glycans and the unmodified protein

Glycoprotein mass (Da)		GPa#	GPb#	GPc#
Observed protein mass (Fig. 1) RI	Calculated using proposed glycans (Table 3)			
333,66**	33,368	IV	III	
34,082*	34,083	IV	I	III
34,213**	34,215	III	III	III
34,375****	34,377	IV	III	III
34,886***	34,888	VIII	III	III
35,050*	35,051	VIII	IV	III
35,398*	35,400	VIII	III	VII
35,561**	35,562	VIII	VIII	III
36,071**	36,073	VIII(XI)	XI(VIII)	III
36,276*	36,277	VIII(X)(XI)	XI(XI)(X)	V(III)(III)
36,584**	36,585	VIII	XI	VII
36,747**	36,747	VIII	VIII	XI
37,256**	37,258	XI(VIII)	VIII(VIII)	XI(XIII)
37,769*	37,770	XI	XI	XI

The calculated masses take into account the presence of four disulfide bridges and uses the average masses for the glycans and the protein. The relative intensity (RI) is qualitative description based of the intensity of the corresponding mass in the deconvoluted MS spectrum. The most abundant species are labelled **** and the less abundant species specified by *.

Several attempts have been made to elucidate the structure of the glycans of cationic peanut peroxidase with this study being the most complete to date. With these data, it is now possible to examine the role of glycans in CP stability and activity by treating the protein with exo or endoglycosidases and comparing glycan structure with stability, and function of the enzyme. This study once more demonstrates the power of mass spectrometry to resolve complex mixtures of post-translationally modified proteins. The characterization of these modifications will be critical for the detailed understanding of protein function in plants and other organisms.

3. Experimental

3.1. Materials

Acetonitrile, HPLC grade water (for MS experiments), and formic acid were from Fisher Scientific. Trifluoroacetic acid (TFA), dithiothreitol (DTT), and iodoacetamide (IAA) were purchased from Sigma–Aldrich, sequencing grade trypsin was obtained from Promega, and borosilicate tips for nano-MS experiments were bought from Proxeon.

3.2. CP production and purification

Peanut (*A. hypogaea*) cells derived from cotyledonary tissue (Verma and van Huystee, 1970) were grown in suspension medium with the recommended components. The spent medium was collected by filtration every two weeks and the proteins were acetone precipitated. Proteins were separated by carboxymethyl (CM) chromatography followed by ConA affinity chromatography

(Sesto and van Huystee, 1988). The amount of CP was calculated by measuring the heme absorption at $A_{405 \ nm}$ with the molar absorptivity coefficient of $111 \ mM^{-1} \ cm^{-1}$ for heme assuming a 1:1 ratio of heme:CP (Chibbar et al., 1984).

3.3. HPLC purification

To remove minor contaminants from the CP preparation, analytical reversed-phase HPLC was carried out on an Agilent 1100 system equipped using a C18 column (Zorbax SB-C18, 4.6×250 mm). CP (500 µg) was redissolved in NH_4HCO_3 (100 mM, 100µl) and injected onto the column. Elution was performed at room temperature at a flow rate of 1 ml/min using a gradient of 0–40 min, 10–65% B; 40–45 min, 65–95% B; 45–50 min, 95–10% B (A: water + 0.1% TFA, B: acetonitrile + 0.1% TFA) with detection of the protein at both 220 and 280 nm. The major protein peak from 27 to 29 min was collected and pre-analyzed by flow injection analysis on a triple quadrupole mass spectrometer (Quattro Micro, Micromass, UK) prior to drying at room temperature in a SpeedVac (Savant).

3.4. Protein molecular mass analysis

ESI-MS analysis of intact CP was carried out on a Q-TOF mass spectrometer (Q-TOF Micro, Micromass, UK). LC-MS was carried out on a 1×150 mm microbore column using the same gradient as for purification with the exceptions that flow rate was set to 30 µl/min and formic acid was used to replace TFA. MS experiments were performed at a cone voltage of 40 V, capillary voltage of 3.2 kV, collision energy of 10 V, source temperature of 80 °C, and desolvation temperature of 200 °C. Data

were calibrated prior to acquisition using the charge state envelope of trypsinogen (MW 23,981 Da). Calibration and CP data were acquired over a m/z range of 1100–3900 Da. Data were acquired using Masslynx 3.5 (Micromass). Data were processed using MaxEnt1 of Masslynx 4.0 with a final output mass range 30–40 kDa.

3.5. Protein digestion

HPLC-purified CP was dissolved to 4 μg/μl in 100 mM NH_4HCO_3, reduced with 50 mM DTT at 56 °C for 45 min, and treated with 200 mM IAA at room temperature in the dark for 30 min to alkylate reduced Cys residues. The sample was subjected to centrifugal ultrafiltration with a Nanosep 3K OMEGA (PALL Corporation) and washed with 2×200 μl of 50 mM NH_4HCO_3 to remove free IAA and DTT. CP was resuspended to 2 μg/μl in 50 mM NH_4HCO_3, 2.5 mM $CaCl_2$, pH 7.4 and digested with trypsin (1:40 w/w) for 2 h at 37 °C. The digestion was evaluated using a Q-TOF mass spectrometer (Q-TOF Micro, Micromass, UK). Sample was desalted on a 0.5×5 mm C18 precolumn (LC Packings) using a gradient of A: water + 0.1% formic acid (FA) and B: acetonitrile + 0.1% FA. MS and MS/MS data were acquired in a data-dependent fashion where switching from MS survey to MS/MS was set to occur for multiply charged species. Mass spectrometer parameters were set as follows: cone voltage, 35 V; capillary voltage, 3.2 kV; collision energy for MS data, 10 V; collision energy for MS/MS data depended on the charge state and m/z of the precursor ions; MS data acquisition range, 400–1495; MS/MS data acquisition range, m/z 100–1800. Data were acquired using Masslynx 3.5. Data were processed using the peptideauto.exe and MaxEnt3 algorithms of Masslynx.

3.6. Separation of glycopeptides by reverse phase HPLC

One hundered microliters of each tryptic digest was separated on a C18 analytical column ((Zorbax SB-C18, 4.6×250 mm) using an Agilent 1100 system. Elution was carried out at room temperature at a flow rate of 1 ml/min using a gradient of 0–5 min, 0–5% B; 5–10 min, 5–20% B; 10–35 min, 20–25% B; 35–50 min, 25–45% B; 50–60 min, 45–100% B; 60–70 min, 100–0% B (A = MilliQ water + 0.1 % TFA, B = HPLC grade actonitrile + 0.1% TFA) according to Lige et al. (2001). Absorbance was monitored at 220 and 280 nm. Fractions were collected automatically using a rate of one fraction per minute followed by evaporation a Speedvac (Savant).

3.7. Identification glycopeptides by LC-MS/MS with PID

The glycopeptide-containing fractions were identified using a Q-TOF Micro mass spectrometer operating in precursor ion discovery (PID) mode (Ritchie et al., 2003). The collision energy was switched alternately between high (28 V) and low (7 V) every second. MS/MS data acquisition was triggered by the detection of product fragment ions at m/z 163.060, 204.084, and 366.139 found in the high CE function of this experiment. Mass spectrometer parameters were set as follows: cone voltage, 35 V; capillary voltage, 3.2 kV; collision energy for MS/MS data depended on the charge state and m/z range; MS data acquisition range, 400–2500Δ; MS/MS data acquisition range 100–2500. Calibration was achieved using an MS/MS spectrum of Glu-fibrinopeptide-b. Data were acquired using Masslynx 3.5.

3.8. Determination of glycopeptide structure using ESI-MS and ESI-MS/MS

Each identified glycopeptide group (GPa, GPb and GPc) was further analyzed by using positive ion nano ESI-MS and ESI-MS/MS using borosilicate tips. MS measurements were performed at a cone voltage of 45 V and a collision energy of 5 V to get the parent ion information. For MS/MS acquisition, a collision energy ramp of 15–50 V was manually applied to obtain a diverse range of glycopeptide fragment ions in order to provide as much structural and peptide sequence information as possible. Capillary voltage was set to optimize spray from the borosilicate tip. Data were acquired from m/z 900 to 2500 for MS mode and 100 to 2500 or 3000 (depending on the glycan size) for MS/MS mode. Data acquisition and basic analysis (i.e., MaxEnt3) was performed using Masslynx 4.0.

References

Chibbar, R.N., Cella, R., van Huystee, R.B., 1984. The heme moiety in peanut peroxidase. Can. J. Biochem. Cell Biol. 62, 1046–1050.

Cooper, C.A., Gasteiger, E., Parker, N.H., 2001. GlycoMod – a software tool for determining glycosylation compositions from mass spectrometric data. Proteomics 1, 340–349 http://us.expasy.org/tools/glycomod).

Dell, A., Morris, H.R., 2001. Glycoprotein structure determination by mass spectrometry. Science 291, 2351–2356.

Ferrige, A.G., Seddon, M.J., Green, B.N., Jarvis, S.A., Skilling, J., 1992. Disentagling electrospray mass spectra with maximum entropy. Rap. Commun. Mass Spectrom. 6, 707–711.

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.

Hu, C.F., van Huystee, R.B., 1989. Role of carbohydrate moieties in peanut (Arachis hypogaea) peroxidases. Biochem. J. 263, 129–135.

Khandke, K.M., Fairwell, T., Chait, B.T., Manjula, B.N., 1989. Influence of ions on cyclization of the amino terminal glutamine residues of tryptic peptides of streptococcal PepM49 protein. Resolution of cyclized peptides by HPLC and characterization by mass spectrometry. Int. J. Pept. Prot. Res 34, 118–123.

Kvasnička, F., 2003. Proteomics: general strategies and application to nutritionally relevant proteins. J. Chromatogr. B 787, 77–89.

Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Lainé, Gomord, V., Faye, L., 1998. N-glycoprotein biosynthesis in plants: recent developments and future trends. Plant Mol. Biol. 38, 31–48.

Lige, B., Ma, S., van Huystee, R.B., 2001. The effects of the site-directed removal of N-glycosylation from cationic peanut peroxidase on its function. Arch. Biochem. Biophys. 386, 17–24.

O'Donnell, J.P., Wan, L., van Huystee, R.B., 1992. Characterization of the two forms of cationic peanut peroxidase from cultured peanut cells. Biochem. Cell Biol. 70, 166–169.

Ravi, K., Hu, C., Reddi, P.S., van Huystee, R.B., 1986. Effect of tunicamycin on peroxidase release by cultured peanut cells. J. Exp. Bot., 37:1708–1715.

Ritchie, M.A., Gill, A.C., Iqbal, M., Pande, A., MacCauley, J., McKenna, T., Langridge, J.I. 2003. In: Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics, Montreal, Quebec, Canada, June 2003.

Rodriguez Maranon, M.J., Stillman, M.J., van Huystee, R.B., 1993. Co-dependency of calcium and porphyrin for the integrated molecular structure of peanut peroxidase: a circular dichroism analysis. Biochem. Biophys. Res. Commun. 194, 326–332.

Schuller, D.J., Ban, N., van Huystee, R.B., McPherson, A., Poulos, T.L., 1996. The crystal structure of peanut peroxidase. Structure 4, 311–321.

Sesto, P.A., van Huystee, R.B., 1988. Purification and yield of a cationic peroxidase from a peanut suspension cell culture. Plant Sci. 61, 163–168.

Shaw, G.S., Sun, Y., Barber, K.R., van Huystee, R.B., 2000. Sequence specific analysis of the heterogeneous glycan chain from peanut perxidase by ^{1}H NMR spectroscopsy. Phytochemistry 53, 135–144.

Sun, Y., Lige, B., van Huystee, R.B., 1997. HPLC determination of the sugar compositions of the glycans on the cationic peanut peroxidase. J. Agric. Food Chem. 45, 4196–4200.

van Huystee, R.B., Tam, A.S.K., 1988. Peptides released by cultured peanut cells during growth. J. Plant Physiol. 133, 645–647.

van Huystee, R.B., Wan, L., 1994. Carbohydrate moiety of peanut peroxidase necessary for enzyme activity. C.R. Acad. Sci. Paris 317, 789–794.

van Huystee, R.B., Zheng, X., 1995. Half-life of peroxidase in actively growing and arrested peanut cells. C.R. Acad. Sci. Paris 318, 655–658.

van Huystee, R.B., Sun, Y., Lige, B., 2002. A retrospective look at the cationic peanut peroxidase structure. Crit. Rev. Biotechnol. 22, 335–354.

Verma, D.P.S., van Huystee, R.B., 1970. Cellular differentiation and peroxidase isozymes in cell culture of peanut cotyledons. Can. J. Bot. 48, 429–431.

Wan, L., Gijzen, M., van Huystee, R.B., 1994. Heterogeneous glycosylation of cationic peanut peroxidase. Biochem. Cell. Biol. 72, 411–417.

Zeleny, R., Altmann, F., Praznik, W., 1999. Structural characterization of the N-linked oligosaccharides from tomato fruit. Phytochemistry 51, 199–210.

Zivy, M., de Vienne, D., 2000. Proteomics: a link between genomics, genetics, and physiology. Plant Mol. Biol. 44, 575–580.