ELSEVIER

Carbohydrate Research 311 (1998) 61–69

CARBOHYDRATE RESEARCH

Heterogeneity of glycans at each *N*-glycosylation site of horseradish peroxidase

James S.S. Gray, Byung Yun Yang, Rex Montgomery *

Department of Biochemistry, College of Medicine, University of Iowa, Iowa City, IA 52242, USA

Received 15 June 1998; accepted in revised form 22 July 1998

Abstract

The tryptic glycopeptides of horseradish peroxidase isozyme c (HRPc) were studied by methylation linkage analysis, exoglycosidase degradation, and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDITOFMS). Over 90% of the predicted tryptic peptides and glycopeptides of HRPc could be identified in the unfractionated digest. Four glycans, namely (Xyl)Man_3(Fuc)$GlcNAc_2$ (major species), (Xyl)Man_2(Fuc)$GlcNAc_2$, (Xyl)$Man_3GlcNAc_2$, and Man_3(Fuc)$GlcNAc_2$ (minor species), were observed at all of the *N*-glycosylation sites and account for greater than 95% of the carbohydrate. Other members of this glycan family, namely $(Xyl)_xMan_m(Fuc)_f$ $GlcNAc_2$ ($x=0$ or 1, $f=0$ or 1, $m=$ 4, 5, 6, or 7), account for the rest of the glycans. Only traces of high mannose-type glycans were detected in HRPc. Two sites, namely those at Asn-57 and Asn-267, were found to be more heterogeneous than the sites at Asn-13, Asn-158, Asn-186, 198 (doubly glycosylated peptide), Asn-214, and Asn-255. Two of the glycopeptides were observed as part of disulfide-linked species. MALDITOFMS confirmed the *N*–glycosylation sites previously reported [K.G. Welinder, *Eur. J. Biochem.*, 96 (1979) 483–502] and was used to determine the heterogeneity of the glycan pool at each site.

Keywords: Horseradish peroxidase; *N*-glycosylation sites; MALDITOFMS

1. Introduction

Horseradish peroxidase (HRP), a glycoprotein, occurs as several isozymes. The most intensely studied of these is isozyme c (HRPc), which has been sequenced [1] and cloned and expressed in *Escherichia coli* [2]. The crystal structure of the recombinant enzyme has been determined [3]. The structure of the major glycan is known [4] and the heterogeneity of the glycans at the *N*-glycosylation sites have been demonstrated [5]. There are nine *N*-glycosylation consensus sequences (NXT/S, X≠P) in HRPc, but only eight of these are occupied [1].

It has been observed in our studies and those of others [6] that different samples of HRP from commercial sources usually differ in the composition of the glycans. A previous study [5] confirmed

* Corresponding author. Fax: +1-319-335-9570; e-mail: rex-montgomery@uiowa.edu

Abbreviations: DHB, 2,5-dihydroxybenzoic acid; GP, glycopeptide-containing fractions; 4HCCA, α-cyano-4-hydroxycinnamic acid; HPLC, high performance liquid chromatography; HRPc, horseradish peroxidase isozyme c; MALDITOFMS, matrix assisted laser desorption/ionization time-of-flight mass spectrometry; N_2M_3FX, (Xyl)Man_3(Fuc)$GlcNAc_2$ etc.; RP, reverse phase; SBP, soybean seed-coat peroxidase.

0008-6215/98/$—see front matter
PII S0008-6215(98)00209-2

that the major glycan present at all sites is (Xyl)-Man_3(Fuc)$GlcNAc_2$, accounting for more than 70% of the total carbohydrate present in this sample. Other members of this glycan family, namely $(Xyl)_xMan_m(Fuc)_fGlcNAc_2$ ($x=0$ or 1, $f=0$ or 1, $m=2$, 4, 5, or 6), account for a further 20–25% of the glycans. Only traces of high mannose-type glycans ($Man_xGlcNAc_2$, $x=4$, 5, 6, or 7) were detected in HRPc, in contrast to soybean seed-coat peroxidase where a significant part (~16%) of the total glycan pool is composed of high-mannose species [7,8]. Moreover, the high-mannose glycans of soybean seed coat peroxidase are essentially confined to a single glycosylation site [8].

The most heterogeneous glycan pools from HRPc were detected in three carbohydrate-containing fractions [5]. From these studies and others on the glycan pools released from apo-HRP [9] or from unfractionated digests of HRP [6], nothing is known about their points of origin in the protein. In this paper, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDITOFMS) is used to confirm the previously reported *N*-glycosylation sites [1] and to analyze the heterogeneity of the glycan pool at each site.

2. Results and discussion

HRPc consists of 308 amino acid residues (Fig. 1), including 21 arginine and 6 lysine residues. Consequently, complete digestion by trypsin results in 28 tryptic peptides, seven of which (T1, Z[1–19]R; T5, L[39–62]R; T16, N[154–159]R; T20, L[184–206]R; T21, T[207–224]R; T24, G[242–264]R; T25, S[265–283]R) are glycopeptides. Tryptic peptide T20 contains two glycosylated sites. The major glycan of HRPc was previously determined to be (Xyl)Man_3(Fuc)$GlcNAc_2$ [4,5] and is present on each of the glycosylation sites of HRPc. The predicted masses of the major glycopeptides are presented in Table 1.

MALDITOFMS analysis of an unfractionated tryptic digest of HRPc.—The masses of, and symbols used for the glycans on HRPc are summarized in Table 2.

The unfractionated tryptic digest of apo-HRPc was analyzed by MALDITOFMS with 4HCCA (Fig. 2A, upper panel) or DHB/Fuc as matrix (Fig. 2A, lower panel). All of the glycopeptides with the major glycan ((Xyl)Man_3(Fuc)$GlcNAc_2$) are observed and identified with DHB/Fuc as matrix as protonated species, $[M+H]^+$.

Six of the seven predicted tryptic glycopeptides give strong signals using 4HCCA (Fig. 2A, upper panel) with masses similar to those observed using DHB/Fuc (Fig. 2A, lower panel). T16 is not easily observed using 4HCCA until the scale is greatly expanded. A major advantage of using 4HCCA, despite the poor signal obtained from glycosylated T16, is that many of the higher-mass peptides and glycopeptides give rise to more intense signals.

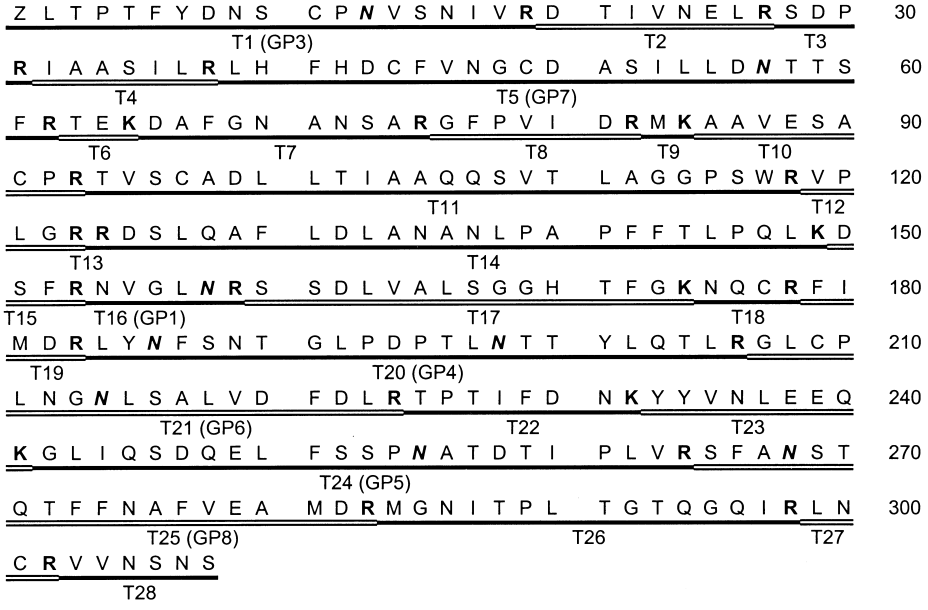

Fig. 1. Sequence of HRPc [1]. The trypsin cleavage sites are indicated by bold letters. The tryptic peptides are numbered from the *N*-terminus and are indicated below the peptides. The sites of carbohydrate attachment are indicated by bold italics.

Table 1
The identification of the peptides associated with the carbohydrate-containing fractions, GP1 to GP8 previously purified by reverse-phase HPLC on a Spherisorb S5 column [5]

Carbohydrate fraction	Retention time (min)	Major tryptic peptide	Predicted mass of tryptic glycopeptide (av) (Da)	Observed mass of tryptic glycopeptide (Da)
GP1	20.8	T16	1843.8	1844.0
GP2	34.2–35.0	T1-SH	3322.4	3323.9
GP3	42.5	T1-T10	4223.4	4224.8
GP4	43.6	T20	4986.1	4987.3
GP5	44.6	T24	3673.9	3674.6
GP6	48.4	T21	3606.8	3607.4
GP7	48.8	T5	3896.1	3896.6
GP8	54.7	T25	3355.5	3355.6

Table 2
Molecular weights, abbreviations, and symbols used for the oligosaccharides in this paper

		Molecular weight (av.)		
Oligosaccharide	Abbreviation	M (calc.)	$M-H_2O$	Symbol used in tables
(Xyl)Man_2(Fuc)$GlcNAc_2$	N_2M_2FX	1027.0	1008.9	**a**
(Xyl)$Man_3GlcNAc_2$	N_2M_3X	1043.0	1024.9	**b**
Man_3(Fuc)$GlcNAc_2$	N_2M_3F	1057.0	1037.0	**c**
(Xyl)Man_2(Fuc)$GlcNAc_2$	N_2M_3FX	1189.1	1171.1	**d**
(Xyl)$Man_4GlcNAc_2$	N_2M_4X	1205.1	1187.1	**e**
Man_4(Fuc)$GlcNAc_2$	N_2M_4F	1219.1	1201.1	**f**
(Xyl)Man_4(Fuc)$GlcNAc_2$	N_2M_4FX	1351.2	1333.2	**g**
$Man_3GlcNAc_2$	N_2M_3	910.8	892.8	**h**
$Man_4GlcNAc_2$	N_2M_4	1073.0	1055.0	**i**
$Man_5GlcNAc_2$	N_2M_5	1235.1	1217.1	**j**
$Man_6GlcNAc_2$	N_2M_6	1397.3	1379.2	**k**
$Man_7GlcNAc_2$	N_2M_7	1559.4	1514.4	**l**

Upon α-mannosidase digestion, each peak assigned to a major glycopeptide [that is, the peptide substituted with (Xyl)Man_3(Fuc)$GlcNAc_2$] other than the diglycosylated species, T20, shifted to a mass 324 Da lower, corresponding to the loss of two Man residues; the mass of diglycosylated T20, shifted by 648 Da corresponding to the predicted loss of four Man residues, two from each glycan. All of the major glycopeptide peaks disappeared from the mass spectrum after digestion of the tryptic digest with glycopeptidase A (data not shown). These data unambiguously identify the glycopeptides. In addition, the α-mannosidase digestion provides evidence that the Man residues of the major glycans are terminal residues.

Closer examination of the mass spectrum in Fig. 2A reveals a great deal of information about the glycan heterogeneity of the individual glycosylation sites (Fig. 2B and C). Expansion of the region about T5 (Fig. 2B), as an example, indicates that T5 is substituted with small amounts of high-mannose species of glycan in addition to Xyl- and Fuc-containing glycans. The peak at m/z 4058 has been assigned to both T5 substituted with (Xyl)-Man_4(Fuc)$GlcNAc_2$ and to T1-S-S-T10 substituted with (Xyl)Man_2(Fuc)$GlcNAc_2$ for a number of reasons. Firstly, this peak does not disappear upon reduction of the tryptic digest with either 1,4-dithiothreitol or tris(2-carboxyethyl)phosphine hydrochloride under conditions that lead to the complete disappearance of the other disulfide-linked species. Secondly, this peak co-purifies by C_8 RP HPLC with GP7, a peak not containing any T1-S-S-T10, but containing T5; it also co-purifies with T5 upon C_{18} RP HPLC (data not shown). Finally, it is converted to a species of lower molecular mass by α-mannosidase digestion of the individual glycopeptide fractions. The presence of (Xyl)Man_4(Fuc)$GlcNAc_2$ is unusual in that this glycan has not been described in the literature for HRPc, although (Xyl)$Man_4GlcNAc_2$ has been observed on the glycans of ricin [10].

The peak at m/z 4058 also does not completely disappear upon α-mannosidase digestion of an

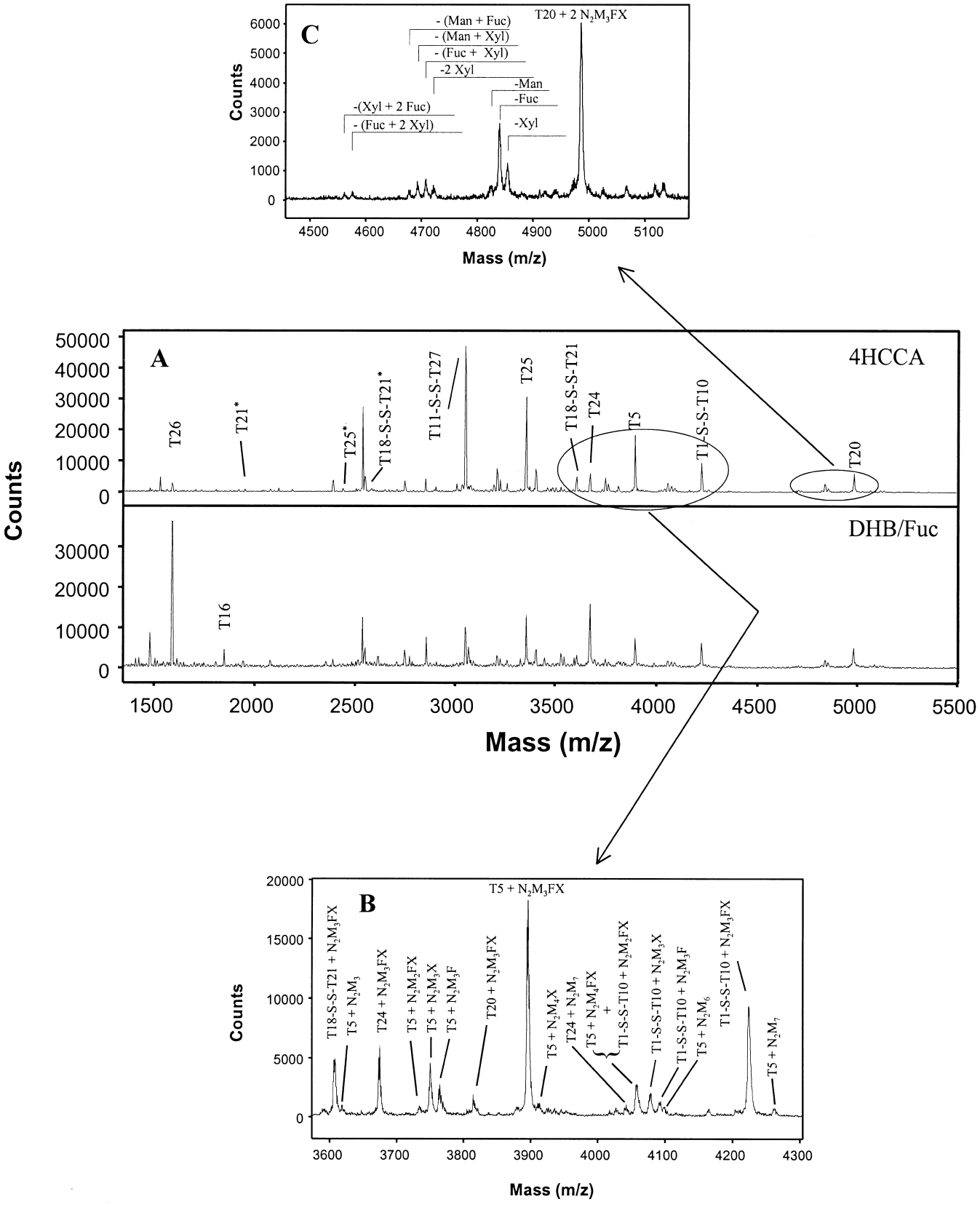

Fig. 2. The MALDITOFMS analysis of an unfractionated tryptic digest of HRPc using either 4HCCA (upper panel) or DHB/fucose (lower panel) as matrix (A); (B) and (C) are expanded regions of the region about tryptic glycopeptides T5 and T20 using 4HCCA as matrix. The peptides are marked with a number that can be identified in Fig. 1. Peptides linked through a disulfide bond are presented as two numbers. Non-glycosylated peptides are marked with an asterisk.

unfractionated trypsin digest of HRPc. This apparent contradiction of the data obtained by α-mannosidase digestion of T5 and T1-S-S-T10 is resolved by the observation that T20 substituted with a combination of (Xyl)Man(Fuc)$GlcNAc_2$ and $ManGlcNAc_2$ (expected to arise after α-mannosidase digestion) gives rise to a peak of this mass, and is indeed observed in the MALDITOFMS spectrum of α-mannosidase-digested GP3. This result indicates that T20 contains some high-mannose species in amounts too small for the individual glycopeptides to be detected by MALDITOFMS.

The diglycosylated glycopeptide, T20 (Fig. 2C) also exhibits considerable heterogeneity, and peaks resulting from the loss of Man, Fuc, and Xyl residues, either singly or in combination, are observed. However, all of the heterogeneity can be explained in terms of combinations of the four basic glycans detected at all of the HRPc *N*-glycosylation sites. In the T5 region, a small peak corresponding to T20 substituted with a single glycan, (Xyl)Man_3(Fuc)$GlcNAc_2$, is observed. Cleavage of diglycosylated T20 with chymotrypsin demonstrated that similar heterogeneity of the glycans was present at each site.

The heterogeneity of the glycans about each *N*-glycosylation site observed in the unfractionated trypsin digest of HRPc is summarized in Table 3. It is clear from these results that the major heterogeneity is found about Asn-57 and Asn-267. The glycans previously reported as being present in GP6 are derived from a mixture of T18-S-S-T21 and T5, thus explaining the apparent heterogeneity of this fraction in the earlier publication [5].

Over 90% of the tryptic peptides can be accounted for in the analysis of an unfractionated tryptic digest with either 4HCCA or DHB/Fuc as matrices. The peptides of lower mass are observed more readily with DHB/Fuc, whereas the larger peptides generate more intense signals with 4HCCA as matrix.

Eight cysteine residues are present in HRPc and the following –S–S– linkages have been established: C11–C91, C44–C49, C97–C301, and C177–C209 [1]. Three of these disulfide bonds are associated with tryptic glycopeptides T1, T5, and T21. The disulfide bond associated with T5 is an intra-residue bond and thus the predicted mass of this glycopeptide is 2 Da less than for the reduced species. Both T1 and T21 form inter-residue disulfide bonds with other tryptic peptides, namely T10 and T18, respectively. The peak at m/z 4224 is due to T1 linked to T10 through a disulfide linkage and corresponds to the disulfide linkage between C11 and C91. It is converted into T1–SH and T10–SH upon reduction, both species being detected by MALDITOFMS analysis. Both T1–SH and T10–SH, arising from the prompt (in-source) fragmentation of the disulfide bond [11], are also observed in a MALDITOFMS analysis of HPLC-purified T1–S–S–T10. Similarly, the peak at m/z 3606 corresponds to T18–S–S–T21 and has analogous behavior upon reduction and MALDITOFMS analysis.

MALDITOFMS analyses of the glycopeptides purified by C_8 reverse-phase HPLC from a tryptic digest of HRPc.—MALDITOFMS analysis of the glycopeptides purified by C_8 RP HPLC [5] confirmed the glycosylation site heterogeneity observed in the analysis of the unfractionated tryptic digest. In addition, peaks not observed in the analysis of the unfractionated tryptic digest,

Table 3
Types of glycans observed on each glycosylation site of HRPc. These data were extracted from the MALDITOFMS analysis of the unfractionated tryptic digest of HRPc presented in Fig. 2 and from MALDITOFMS analyses of C_8-purified glycopeptides

HRP Tryptic peptide (C_8 glycopeptide fraction)	Molecular weight (av.), major glycopeptide (Da)	Glycans observed on peptide in unfractionated tryptic digest of HRPc	Glycans observed on peptide in C_8 RP HPLC fractions
T1-SH (DP2)	3322.5	**d**	**d**
T1-S-S-T10 (GP3)	4222.4	**a**, **b**, **c**, **d**, **g**	**a**, **b**, **c**, **d**, **e**,
T5 (GP7)	3896.1	**a**, **b**, **c**, **d**, **e**, **f**, **g**, **h**, **j**, **k**, **l**	**a**, **b**, **c**, **d**, **e**, **f**, **g**, **h**, **j**, **k**, **l**
T16 (GP1)	1843.8	**a**, **b**, **c**, **d**	**a**, **b**, **c**, **d**, **e**, **g**, **k**
T20 (GP4) Monoglycosylated	3815.5	**d**	**d**
T20 (GP4) Diglycosylated	4986.1	**d**+**d**, **a**+**d**, **b**+**d**, **c**+**d**, **a**+**b**, **b**+**c** **d**+**h**, **c**+**c**, **b**+**b**, **b**+**h**, **c**+**h**	**d**+**d**, **a**+**d**, **b**+**d**, **c**+**d**, **a**+**b**, **b**+**c**, **c**+**c**, **d**+**h**
T21-SH (GP6)	3088.3	**d**	**d**
T18-S-S-T21 (GP6)	3604.8	**a**, **b**, **c**, **d**, **i**	**a**, **b**, **c**, **d**
T24 (GP5)	3673.8	**b**, **c**, **d**, **h**, **k**, **l**	**a**, **b**, **c**, **d**, **k**
T25 (GP8)	3355.5	**a**, **b**, **c**, **d**, **e**, **h**, **i**, **j**	**a**, **b**, **c**, **d**, **e**, **k**

either because of signal suppression or due to signal crowding, become visible. Moreover, signals that are poor or highly suppressed in the complex mass spectrum of the unfractionated digest are readily observed in the glycopeptide fractions. Thus, T16, which gives a poor signal in the unfractionated digest, gives a strong signal in the purified glycopeptide fraction, GP1, revealing not only the presence of the major glycans, **a**, **b**, **c**, and **d** but also traces of higher mass glycans, **e**, **g**, and **k**, albeit it in small amounts (Table 3).

Two of the glycosylation sites, Asn-57 (T5) and Asn-267 (T25) are the most heterogeneous sites in terms of the amounts of the higher-mass glycans (Table 3). The heterogeneity of GP6, as previously reported [5], is due to the substantial contamination of this fraction with T5. A MALDITOFMS analysis of the glycopeptides in GP6 and GP7 confirms that T5 is more heterogeneous than T21 and enables the assignment of the different glycans to these two sites. A MALDITOFMS analysis of C_{18} RP HPLC purified T18–S–S–T21 and T5 glycopeptides, minimally contaminated with each other, substantiates this conclusion.

The disulfide-linked species are observed mainly in the oxidized form in the glycopeptide fractions, although a trace of free T1–SH is observed in GP2. These disulfide-linked species, upon reduction with either 1,4-dithiothreitol or tris(2-carboxyethyl)-phosphine hydrochloride, disappeared from the mass spectra with the concurrent appearance of the predicted reduced glycopeptides.

Structure of the glycans.—The masses of the glycans determined by MALDITOMS allows a generic monosaccharide composition to be assigned in terms of the number of hexose, deoxyhexose, pentose, and hexosamine residues present, but does not allow structures to be assigned. The monosaccharide composition is similar for all the glycopeptides [5] and shows an absence of galactosyl residues.

Some conclusions can be drawn from fairly simple experiments, such as α-mannosidase digestion and methylation analyses, and these are reported here.

(i) Digestion of the purified glycopeptides with α-mannosidase indicated that all of the α-D-mannose residues can be removed with the production of $(Xyl)_xMan_1(Fuc)_xGlcNAc_2$, where $x=0$ or 1. Consequently, the α-D-mannose residues are capped only by other α-D-mannose residues.

(ii) All of the methylated alditol acetates detected in the methylation analysis of holo-HRPc are present in the methylation analyses of the tryptic glycopeptides (Table 4). The major methylated alditol acetate peaks in the GLC–FID and GLC–MS traces could be analyzed with some confidence since the contribution of minor non-carbohydrate peaks to the total area of these peaks is small. Quantitation of the minor methylated alditol acetates presented a problem because the contribution of co-eluting non-carbohydrate species to the peak was often substantial. Consequently, it was not possible to use the GLC–FID data to quantitate these peaks, and the use of a GLC–MS was essential. Semi-quantitative data were obtained by integration of extracted selected ion chromatograms, namely those from m/z 118, 129, 130, and 159. Where similar quantitative data were obtained with two or more of these ions, they were averaged.

(iii) Although it is difficult to uniquely define the structure of each glycan from the methylation of the mixture of glycans in each glycopeptide fraction, it is possible to comment on structural characteristics, particularly in the light of the study of Takahashi et al. [6] and the relative proportions of the glycans from an earlier study [5].

(iv) From Table 4 it is clear that:

1. Most of the glycans are fucosylated at the 3-hydroxyl of the GlcNAc residue linked to Asn, since 1,3,4,5-tetra-*O*-acetyl-(2-*N*-methylacetamido)-6-*O*-methylglucitol and 1,4,5-tri-*O*-(2-*N*-methylacetamido)-3,6-di-*O*-methylglucitol derived from the core chitobiose in all the glycans are approximately equimolar.
2. The majority of the glycans carry either fucosyl and/or xylosyl residues, since these are present in each GP fraction to a similar extent.
3. The xylosyl residues substitute at the 2-OH group of the β-D-mannosyl residue, which is never substituted by an α-D-mannosyl residue at its 6-*O*-position (because of the total absence of a 1,2,5,6-tetra-*O*-acetyl-3,4-di-*O*-dimethylmannitol in any methylation analysis) as has been reported for stem bromelain [12]. This

Table 4
Methylation analyses of holo-HRPc, an apo-HRPc hydrazinolyzate, and glycopeptides purified from a tryptic digest of apo-HRPc (GP1 to GP8) by RP HPLC. The major methylated alditol acetates were quantitated by GLC with FID detection and by GLC–MS with integration of the total ion chromatogram. The minor peaks were quantitated from the extracted selected-ion chromatograms as described in the legend

Methylated alditol acetate	Relative mol									
	Holo-HRPc	Hydrazinolyzate of apo-HRPc	GP1	GP2	GP3	GP4	GP5	GP6	GP7	GP8
2,3,4-Me_3 Xyl[*]	0.9[a]	0.8[a]	0.8[a]	0.6[a]	0.7[a]	1.0[a]	0.9[a]	1.2[a]	0.9[a]	1.0[a]
2,3,4-Me_3 Fuc	1.0[a]	0.8[a]	0.9[a]	0.6[a]	0.9[a]	1.0[a]	1.1[a]	1.1[a]	1.0[a]	0.9[a]
2,3,4,6-Me_4 Man	2.0[a,b]	2.0[a,b]	2.0[a,b]	2.0[a,b]	2.0[a,b]	2.0[a,b]	2.0[a,b]	2.0[a,b]	2.0[a,b]	2.0[a,b]
3,4,6-Me_3 Man	< 0.1[a,d,e]	< 0.1[a,d,e]	< 0.1[b,d,e]	< 0.1[d,e]	< 0.1[b,d,e]	0.1[b,d,e]	—[f]	0.1[d,e]	0.1[d,e]	< 0.1[b,d,e]
2,4,6-Me_3 Man	0.2[b,c]	< 0.1[b,c]	0.1[b,c,d]	0.2[d]	0.1[b,c,d]	0.1[b,c,d]	0.4[b,c,d]	0.2[c,d]	0.1[b,c,d]	
2,3,4-Me_3 Man	< 0.1[c,d]	< 0.1[c,d]	0.1[c,d]	< 0.1[c,d]	< 0.1[c,d]	0.1[c,d]	< 0.1[c,d]			
4,6-Me_2 Man	< 0.1[b,d]	< 0.1[b,d]	0.1[d]	0.3[d,]	0.2[b,d]	0.3[c]	0.3[b,c]	+[g]	0.5[d]	0.3[b,d]
2,4-Me_2 Man	< 0.1[c,d]	< 0.1[c,d]	—[f]	0.2[b,c]	0.1[b,c]	0.1[c]	0.1[c]	0.1[b,c]	0.1[b,c,d]	0.1[b,c,d]
4-Me Man	0.9[a]	0.8[a]	0.6[a]	0.7[a]	0.6[a]	0.9[a]	0.6[a]	0.8[a]	1.0[a]	1.0[a]
2-*N*-Me-3,6-Me_2 GlcNAc[h]	1.1	1.1	0.9	1.1	1.1	1.0	1.0	1.0	1.0	1.0
2-*N*-Me-6 GlcNAc[h]	0.9	0.9	1.1	0.9	0.9	1.0	1.0	1.0	1.0	1.0

[*] 2,3,4-Me_3 Xyl = 1,5-di-*O*-acetyl-(1-deuterio)-2,3,4-tri*O*-methylxylitol etc.
[a] Quantitation from GC-FID using equivalent carbon response factor derived by Sweet et al. [19].
[b] Quantitation from integrated total ion chromatogram (GC–MS).
[c] Quantitation from extracted and integrated *m*/*z* 118 chromatogram (GC–MS).
[d] Quantitation from extracted and integrated *m*/*z* 129 chromatogram (GC–MS).
[e] Quantitation from extracted and integrated *m*/*z* 130 chromatogram (GC–MS).
[f] None detected.
[g] Present but at levels too low (< 0.01 relative mol) to quantitate.
[h] Quantitation from extracted and integrated *m*/*z* 159 chromatogram (GLC–MS).

also contrasts with the proposal that Man α6(Xylβ2)Manβ4GlcNAcβ4(Fucα3)GlcNAc is present in HRPc [6].

4. No terminal *N*-acetylglucosaminyl residues were detected in the methylation linkage analyses of either apo-HRP or its tryptic peptides, in contrast to the findings of [6], demonstrating again the variability of the glycans in different preparations of HRP.
5. The high-mannose glycans were principally (1→3)-linked with some (1→2)-linked residues.

The methylation data obtained for the various glycopeptides, with the exceptions noted above, in general support the structures of the glycan species by Takahashi et al. [6].

The only site on HRPc with a substantial amount of high mannose-containing glycans is T5 where $(Xyl)_x Man_m (Fuc)_f GlcNAc_2$ ($x,f = 0$ or 1; $m = 4$–7) makes up nearly 15% of the carbohydrate [5]. It is this site on soybean seed-coat peroxidase (SBP) that contains mainly high mannose-type glycans [7,8]. The amino acid sequence of HRPc and a partial sequence of SBP [13] reveals significant homology between the two proteins. This extends to some extent to the *N*-glycosylation sites, three of which are closely conserved in the two peroxidases. There is some evidence that a site on the unsequenced amino-terminus of SBP is also glycosylated and this may correspond to a similar site on HRPc (Asn-13). Consequently, four of the eight *N*-glycosylation sites in HRPc and four of the six sites in SBP appear to be conserved.

Interestingly, not all potential glycosylation sites are fully glycosylated, and traces of non-glycosylated peptides (T21, T18–S–S–T21, T25, T20 glycosylated at one site) were detected in a MALDITOFMS analysis of the unfractionated trypsin digest of HRPc. Moreover, the site at Asn-286, is not glycosylated at all.

These studies have revealed that two of the seven glycosylation sites in HRPc, namely T5 and T25, are more heterogeneous than the other five sites (T1, T16, T20, T21, and T24). It has also been demonstrated that a MALDITOFMS analysis of an unfractionated tryptic digest of HRPc provides much of the information available from a MALDITOFMS analysis of the purified glycopeptides, thus demonstrating the power of MALDITOFMS as a technique for analyzing the heterogeneity of the glycans at *N*-glycosylation sites.

3. Experimental

Materials and methods.—Materials. Sequencing-grade modified porcine trypsin was purchased from Promega (Promega Corporation, Madison, WI). Sequencing-grade chymotrypsin was from Boehringer–Mannheim (Indianapolis, IN). α-Mannosidase from Jack Bean and lacto-*N*-difucohexaose II [β-D-Gal-(1→3)[α-L Fuc-(1→4]β-D-GlcNAc-(1→3)-β-D-Gal-(1→4)[α-L Fuc-(1→3]Glc) were obtained from Oxford Glycosystems (Bedford, MA). Tris(2-carboxyethyl)phosphine was from Pierce Chemical Company (Rockford, IL); 1,4-dithiothreitol was from Sigma Chemical (St. Louis, MO). All other materials have been described previously [5,7,8].

Methods. Trypsin digestion, HPLC and MALDITOFMS the purification of HRPc, removal of the heme group from holo-HRPc, trypsin digestion of apo-HRPc, glycopeptidase A digestion, and HPLC fractionation of the tryptic peptides on a Spherisorb C_8 reverse-phase column were performed as described previously [5,7]. Digestion of apo-HRP with sequencing grade modified porcine trypsin was carried out in 50 mM NH_4HCO_3 for 2 h at 37 °C at a final trypsin:apo-HRP ratio of 1:50. Fractionation of the tryptic peptides on a Vydac (Vydac/The Separations Group, Hesperia, CA 92345) reverse-phase C_{18} peptide and protein column (218TP, 300 Å, 250×4.6 mm) is described by Gray and Montgomery [8]. Chymotrypsin digestions were performed in 100 mM NH_4HCO_3, pH 7.8 at a substrate/protease ratio of 50:1 for 1 h at room temperature (23–25 °C). The reaction was stopped by freezing and the NH_4HCO_3 removed by lyophilizing three times from water.

α-Mannosidase was reconstituted with digestion buffer (100 mM NaOAc buffer, pH 5.0 containing 2 mM Zn^{2+}) to give an enzyme concentration of 100 U/mL. Digestion of the unfractionated apo-HRP or purified glycopeptides was performed in 20 μL of digestion buffer with a final α-mannosidase concentration of 50 U/mL. Digestion, under a toluene atmosphere, was carried out for 16–18 h at 37 °C.

Salts were removed from enzyme digests on a C_{18} microcolumn as described by Kuster et al. [14]. Preliminary experiments showed that none of the glycopeptides were lost after washing the microcolumn with 0.1% CF_3CO_2H and elution of the peptides with 80% MeCN–0.1% CF_3CO_2H.

Reduction of disulfide linkages was performed in 0.1% CF_3CO_2H at 23–25 °C with either tris(2-carboxyethyl)phosphine hydrochloride or 1,4-dithiothreitol (5-fold molar excess over disulfide linkages). Samples were either removed directly for MALDITOFMS analysis using 4HCCA as a matrix or were cleaned on a C_{18} microcolumn as already described; the peptides were dissolved in 30% MeCN–0.1% CF_3CO_2H for analysis.

All peptides and glycopeptides were dissolved in 30% CH_3CN–0.1% CF_3CO_2H for MALDITOFMS analysis, and stored at −20 °C.

MALDITOFMS analyses were performed on a PerSeptive Biosystems (Framingham, Massachusetts) Voyager DE-STR MALDITOFMS with α-cyano-4-hydroxycinnamic acid (4HCCA) as matrix using either the dried drop method or the polycrystalline-layer method as described by Beavis and Chait [15]. The polycrystalline-layer method was further modified as follows.[1] One microliter of a saturated solution of 4HCCA in 50% MeCN–0.3% CF_3CO_2H was allowed to dry on a gold-plated target and then wiped off with a tissue leaving a discolored smear. Sample, diluted in the 4HCCA matrix, was spotted on the smear, whereupon crystallization proceeded rapidly. The spots were rinsed with 5 μL of water (twice) and allowed to dry. Analysis of the unfractionated tryptic digests were also performed with a co-matrix composed of 2,5-dihydroxybenzoate and fucose (DHB/Fuc) as described by Billeci and Stults [16].

Linkage analysis by methylation. Analysis of the methylated alditol acetates from under-methylated methyl α-mannoside, prepared as described by Handa and Montgomery [17], was used to obtain the retention times of the methylated mannitol acetates. This preparation was also used to calculate the relative response factors for the quantitative analysis of the tri- and di-*O*-methylmannitol acetates by GLC–MS. The methylated alditol acetates prepared from asparaginyl carbohydrates A and B [18] were used as standards for identifying the methylated alditol acetates derived from terminal mannose, 1,2-, 1,4-, and 1,3,6-linked mannose and terminal and 1,4-linked GlcNAc. The methylated alditol acetates from the methylation of fetuin provided standards for 1,4-linked Gal whereas the methylation of lacto-*N*-difucohexaose II provided standards for 1-linked Gal, 1,3-linked Gal, 1,3,4-linked Glc, 1-linked Fuc and 1,3,4-

[1] This modification was described to us by Mr. Peter Anagnos, Service Engineer, PerSeptive Biosystems, who obtained it from Dr. Brian T. Chait, Rockefeller University.

linked GlcNAc. The latter compound was used to establish a response factor for 1,5-di-*O*-acetyl-2,3,4-tri-*O*-methyl fucitol on the GLC–MS.

Acknowledgements

This work was supported by the Iowa Biotechnology Consortium (USDA Grant No. 94-34188-0039). We thank Kurt Kulhavy and David Venzke for technical assistance.

References

[1] K.G. Welinder, *Eur. J. Biochem.* 96 (1979) 483–502.
[2] A.T. Smith, N. Santam, S. Dacey, M. Edwards, R.C. Bray, R.N.F. Thorneley, and J.F. Burke, *J. Biol. Chem.* 265 (1990) 13335–13343.
[3] M. Gajhede, D.J. Schuller, A. Henrickson, A.T. Smith, and T.L. Poulos, *Nature Struct. Biol.* 4 (1997) 1032–1038.
[4] A. Kurosaka, A. Yano, N. Itoh, Y. Kuroda, T. Nakagawa, and T. Kawasaki, *J. Biol. Chem.* 266 (1991) 4168–4172.
[5] B.Y. Yang, J.S.S. Gray, and R. Montgomery, *Carbohyd. Res.* 287 (1996) 203–212.
[6] N. Takahashi, K.B. Lee, H. Nakagawa, Y. Tsukamoto, K. Masuda, and Y.C. Lee, *Anal. Biochem.* 255 (1998) 183–187.
[7] J.S.S. Gray, B.Y. Yang, S.R. Hull, D.P. Venzke, and R. Montgomery, *Glycobiology* 6 (1996) 23–32.
[8] J.S.S. Gray and R. Montgomery, *Glycobiology* 7 (1997) 679–685.
[9] J. Clarke and L.M. Shannon, *Biochim. Biophys. Acta* 427 (1976) 428–42.
[10] Y. Kimura, S. Hase, Y. Kobayashi, Y. Kyogoku, T. Ikenaka, and G. Funatsu, *J. Biochem. (Tokyo)* 103 (1988) 944–949.
[11] D.L. Crimmins, M. Saylor, J. Rush, and R.S. Thoma, *Anal. Biochem.* 226 (1995) 355–361.
[12] H. Ishihara, N. Takahashi, S. Oguri, and S. Tejima, *J. Biol. Chem.* 254 (1979) 10715–10719.
[13] J. Huangpu, M.C. Graham, and J.S. Graham, *Plant Physiol.* 110 (1996) 714.
[14] B. Kuster, S.F. Wheeler, A.P. Hunter, R.A. Dwek, and D.J. Harvey, *Anal. Biochem.* 250 (1997) 82–101.
[15] R.C. Beavis and B.T. Chait, *Meth. Enzymol.* 270 (1996) 519–551.
[16] T.M. Billeci and J.T. Stults, *Anal. Chem.* 65 (1993) 1709–1716.
[17] N. Handa and R. Montgomery, *Carbohydr. Res.* 11 (1969) 467–484.
[18] C.C. Huang, J. Mayer, and R. Montgomery, *Carbohydr. Res.* 13 (1970) 127–137.
[19] D.P. Sweet, R.H. Shapiro, and P. Albersheim, *Carbohydr. Res.* 40 (1975) 217–225.