The N-glycosylation sites of soybean seed coat peroxidase

James S.S. Gray and Rex Montgomery
Department of Biochemistry, College of Medicine, University of Iowa, Iowa City, IA 52242, USA

To whom correspondence should be addressed

Tryptic digestion of apo-soybean peroxidase (apo-SBP), with and without acetamidation, chromatographic separation of the tryptic fragments and MALDI-TOF analysis of the major components, both before and after digestion with glycopeptidase A, demonstrated the presence of six carbohydrate groups on five peptides. Five of the glycopeptides can be mapped with confidence to the peptides containing Asn16, Asn90, Asn104, Asn169, and Asn174. The sixth N-glycosylation site is not known and does not appear to be Asn145. It may be present on the N-terminus of SBP, which has not been sequenced.

Key words: soybean seed coat peroxidase/glycopeptides/glycosylation sites/MALDI-TOF MS

Introduction

Soybean seed coat peroxidase (SBP) (E.C. 1.11.1.7) is an anionic glycoprotein (pI 4.1) containing 18.2% carbohydrate (Gray et al., 1996) and is the most abundant protein in mature soybean seed coat (Gillikin and Graham, 1991). It possesses exceptional heat stability (McEldoon and Dordick, 1996) and is active and stable at low pH (McEldoon et al., 1995). These properties, together with its activity in organic solvents (Blinkovsky et al., 1994), make the enzyme very attractive to industry.

The glycans of SBP have recently been studied (Gray et al., 1996). Six carbohydrate-containing fractions were purified from a tryptic digest of SBP on a C8 reverse phase column. The carbohydrates were released from the glycopeptides by digestion with glycopeptidase A and characterized by high pH anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), methylation analysis and matrix-assisted laser desorption/ionization mass spectrometry (MALDI-TOF MS; Gray et al., 1996).

Two of the fractions contained substantial amounts of high mannose-type glycans, which together make up about 20% of the total glycan pool (Gray et al., 1996). The other tryptic glycopeptide fractions were substituted mainly with the (Xyl)Man3(Fuc)GlcNAc2 family of glycans (Gray et al., 1996). This is in contrast to horseradish peroxidase isozyme c (HRPc) where the (Xyl)Man2(Fuc)GlcNAc2 family of glycans make up greater than 95% of the glycan pool (Yang et al., 1996).

The recent publication of a partial sequence for SBP (Huangpu et al., 1996) has enabled us to map five of the glycans to their N-glycosylation sites on the polypeptide.

Results and discussion

Sequence of SBP

The partial sequence of SBP (GenBank Accession Number U41657), reported by Huangpu et al. (1996), is presented in Figure 1. This sequence is missing the amino-terminal end and starts from the conserved histidine residue involved in acid-base catalysis (Huangpu et al., 1996). The sequence in Figure 1 is preceded by Arg-Leu-His (not shown) which was determined by sequencing a CNBr fragment of apo-SBP (Huangpu et al., 1996).

There are 21 sites susceptible to cleavage by trypsin in the sequence of apo-SBP depicted in Figure 1 (indicated by underlined, boldface characters) giving raise to 22 tryptic peptides. These are numbered in Figure 1 beginning from the amino-terminus. Five of the peptides, T1, T7, T8, T12, and T14, contain consensus sequences (NXT/S, X ≠ P) for N-glycosylation, T14 containing two consensus sequences. The masses of the peptides and the peptides substituted with either (Xyl)Man2(Fuc)GlcNAc2 or Man3GlcNAc2, previously determined to be the principal glycans present on SBP (Gray et al., 1996), are presented in Table I. The mass of T1, as depicted in Figure 1, is 3864.2 Da; this increases to 4114.0 Da if the preceding Leu-His residues are included. Upon acetamidation of the apo-SBP with iodoacetamide, the masses of the glycopeptides are increased by 57 Da for each cysteine residue present.

Rechromatography of C8 reverse phase purified polypeptides on a Vydac C18 reverse phase column

In a previous study (Gray et al., 1996), the tryptic glycopeptides were purified on a Spherosorb C8 5-S reverse phase column. Six carbohydrate-containing fractions were purified and their glycans investigated (Gray et al., 1996). Carbohydrate fractions GP4, GP5, and GP6 were poorly separated on this column although the fractions were adequately pure for the work reported there (Gray et al., 1996).

In this study, the C8 fractionated glycopeptides were further purified by chromatography on a Vydac C18 reverse phase column eluted with a shallower gradient of acetonitrile-TFA, fractions being collected manually (Figure 2A–E). Because of the difficulty of collecting GP5 and GP6 from the C8 column, they were collected together for further purification on the C18 column. This had no effect on the purification of GP5 and GP6 since the GP5 purified on the C8 column is in effect a mixture of GP4 and GP6; hence, GP5 is not a unique glycopeptide.

All of the carbohydrate-containing fractions, particularly the pooled GP5 and GP6 fraction (designated GP5/6) were found to be heterogeneous by MALDI-TOF MS (Figure 2A–E). A comparison of the elution times on the C8 and C18 columns are presented in Table II. Noteworthy is the inverted elution times.
including preceding LH not present in the sequence published in Genbank. The mass of 4114.0 dalton assumes the formation of an intra-residue disulfide bond between C[4-9]C; the mass is 2 Da greater for the reduced species.

Table I. Predicted molecular weights of the tryptic peptides containing N-glycosylation sequences unsubstituted and substituted with carbohydrate

<table>
<thead>
<tr>
<th>Tryptic peptide</th>
<th>Amino acid residues</th>
<th>Sequence</th>
<th>Peptide mass</th>
<th>Glycopeptide mass (+N_3M_2FX^a)</th>
<th>(+N_2M_1^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>1–35</td>
<td>FHDCFVQGCDGSVL[NNTDTIESQDALPN]</td>
<td>4114.0 (^c)</td>
<td>5285.0</td>
<td>5555.0</td>
</tr>
<tr>
<td>T7</td>
<td>85–91</td>
<td>DSLTANR</td>
<td>776.8</td>
<td>1946.8</td>
<td>2341.8</td>
</tr>
<tr>
<td>T8</td>
<td>86–108</td>
<td>TLANQNLPAFNFNLQTK</td>
<td>2030.4</td>
<td>3201.4</td>
<td>3571.4</td>
</tr>
<tr>
<td>T12</td>
<td>143–163</td>
<td>L[YNFSSTGHLHLDTTYEVL]R</td>
<td>2482.8</td>
<td>3653.8</td>
<td>4025.8</td>
</tr>
<tr>
<td>T14</td>
<td>166–189</td>
<td>CP[QNATGDNL]TL[NDLSTPDQDF]RN</td>
<td>2648.8</td>
<td>4990.8</td>
<td>5730.8</td>
</tr>
</tbody>
</table>

\(^a\) \(N_3M_2FX = (Xyl)Man_\(3\)(Fuc)GlcNAC_2\)

\(^b\) \(N_2M_1 = Man_2GlcNAC_2\)

\(^c\) Including preceding LH not present in the sequence published in Genbank. The mass of 4114.0 dalton assumes the formation of an intra-residue disulfide bond between C[4-9]C; the mass is 2 Da greater for the reduced species.

MALDI-TOF MS analyses of the repurified glycopeptides and the peptides released by glycopeptidase A digestion

The predominance of (Xyl)Man_\(3\)(Fuc)GlcNAC_2 in the glycopeptides simplified their analysis by MALDI-TOF MS to identify the corresponding peptide. Selected fractions from the repurified glycopeptides were analyzed with \(\alpha\)-cyano-4-hydroxycinnamic acid as matrix (Figures 2A–E).

A comparison of the chromatographic separation of GP5/6 (Figure 2EJ) and the MALDI-TOF MS analyses (Figure 2EJ) provides clear evidence that the glycan causes considerable broadening of the chromatographic peak and that some discrimination of the various glycoforms occurs.

Four of the five predicted glycopeptides are readily identified in this analysis: GP1 (T7 substituted mainly with (Xyl)Man_\(3\)(Fuc)GlcNAC_2), GP4 (T8+ (Xyl)Man_\(3\)(Fuc)GlcNAC_2), GP3 (T14+ 2×(Xyl)Man_\(3\)(Fuc)GlcNAC_2), and GP5/6 (T1 substituted with high mannose type species, the major peak containing mainly Man_\(2\)GlcNAC_2). The corresponding asparagine residues are: GP1, Asn90; GP3, Asn169 and Asn174; GP4, Asn104; GP6, Asn16.

These data confirm the identities of the major glycan species previously shown to be present on these glycopeptides (Gray et al., 1996), in which GP5 was found to contain mainly (Xyl)Man_\(3\)(Fuc)GlcNAC_2 whereas GP4 was found to contain Man_\(2\)GlcNAC_2 in addition to (Xyl)Man_\(3\)(Fuc)GlcNAC_2. It is apparent from the HPLC fractionation on a C18 reverse phase column that this is due to incomplete fractionation of the glycopeptides. Some preparations of SBP contain T1 (Asn16) substituted with (Xyl)Man_\(3\)(Fuc)GlcNAC_2 in addition to the high mannose-type glycans which contributes to the complexity of the fractions (see below).

The added dimensions of the mass of the glycopeptide and the mass of the released glycans clarifies the identity of the
parent peptide as demonstrated here. Nonetheless, investigation of the released glycans is still essential if the resolution of the present MALDI-TOF MS is insufficient to handle the increased masses of the glycopeptides.

The masses of the glycopeptides present in GP1, GP3, GP4, and GP5/6 are as predicted from the sequence presented in Figure 1 and the nature of the carbohydrate described by Gray et al. (1996) (Figure 4). Asn16 and Asn104 are glycosylated, confirming the data of Huangpu et al. (1996). Moreover, the carbohydrate on Asn16 is mainly of the high mannose type (Figures 1E_m, 4G,H). In addition, Asn90 (Figure 4A), Asn169 and Asn174 (Figure 4C,D) are glycosylated. It is noteworthy that GP3 contains two glycosylation sites, Asn169 and Asn174, both of which are not necessarily glycosylated (Figures 4D,E). The MALDI-TOF MS analyses of all of the glycopeptides containing principally (Xyl)Man$_3$(Fuc)GlcNAc$_2$ contain peaks which can be attributed to isoforms containing (Xyl)Man$_2$(Fuc)GlcNAc$_2$, Man$_3$(Fuc)GlcNAc$_2$ or (Xyl)Man$_3$GlcNAc$_2$ (Figures 2, 4).
Table II. A comparison of the retention times of the SBP tryptic glycopeptides on a Spherisorb C8 and a Vydac C18 reverse phase column

<table>
<thead>
<tr>
<th>Carbohydrate fraction</th>
<th>Tryptic peptide</th>
<th>Retention time (min)</th>
<th>Observed mass of major glycopeptide/peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Spherisorb C8</td>
<td>Vydac C18</td>
</tr>
<tr>
<td>GP1</td>
<td>T7</td>
<td>17.53</td>
<td>21.80</td>
</tr>
<tr>
<td>GP2</td>
<td>Unknown</td>
<td>42.27</td>
<td>47.45</td>
</tr>
<tr>
<td>GP3</td>
<td>T14</td>
<td>43.05</td>
<td>52.35</td>
</tr>
<tr>
<td>GP4</td>
<td>T8</td>
<td>47.79</td>
<td>57.35</td>
</tr>
<tr>
<td>GP5/6</td>
<td>T1</td>
<td>48.11/48.63</td>
<td>56.05</td>
</tr>
</tbody>
</table>

The observed masses of the glycopeptide present in the major peak as well as the peptide released by glycopeptidase A digestion is also presented.

* A monoglycosylated species with a mass of 3878 Da was observed in acetamidated GP3.

The nature of the carbohydrate in GP1, GP3, and GP4 is remarkably similar in all the preparations of SBP investigated to date.

The difference in masses between GP3 and GP5/6 derived from apo-SBP and acetamidated apo-SBP indicate that GP3 has one and GP5/6 two cysteine residues.

Heterogeneity of GP5/6

The MALDI-TOF mass spectra of GP5/6 provides clear evidence for the high mannose nature of its carbohydrate moiety. Signals are present, 162 Da apart, for T1 substituted with \( \text{Man}_5\text{GlcNAc}_2 \), \( \text{Man}_6\text{GlcNAc}_2 \), \( \text{Man}_7\text{GlcNAc}_2 \) (major species), \( \text{Man}_8\text{GlcNAc}_2 \), and \( \text{Man}_9\text{GlcNAc}_2 \) (Figures 2E, 4G,H).

---

![Fig. 3. HPLC analysis of the tryptic peptides of apo-SBP before (A) and after glycopeptidase A digestion (B). Apo-SBP was digested with TPCK-treated trypsin and a portion was chromatographed on a Vydac C18 column as described in the text (A). Another portion was subjected to glycopeptidase A digestion and chromatographed under identical conditions (B). The proposed identity of the glycopeptides and the inclusive amino acid residues are indicated on the figure.](image-url)
Fig. 4. MALDI-TOF MS analyses of the glycopeptides purified from either apo-SBP or acetamidated apo-SBP (top) and the corresponding peptides released by glycopeptidase A digestion (bottom). Purified glycopeptides were digested with glycopeptidase A as described in the text and the products analyzed by MALDI-TOF MS with α-cyano-4-hydroxycinnamic acid as matrix.
However, there is evidence that in some preparations of SBP, GP5/6 is substituted to a limited extent with (Xyl)Man$_3$ (Fuc)GlcNAc$_2$ in addition to the high mannose type glycans (predicted m/z 5286.5 for Tl+(Xyl)Man$_3$ (Fuc)GlcNAc$_2$, observed m/z 5283.8; Figure 5). Whether this difference reflects a species difference or is due to cultivation or extraction conditions is not known.

**GP2**

The major glycan released from GP2 upon glycopeptidase A digestion is (Xyl)Man$_3$ (Fuc)GlcNAc$_2$ (Gray et al., 1996). The mass of the glycopeptide detected in the MALDI-TOF MS analysis of GP2 is 4657.5 Da (Figures 1B, 4B) and of the peptide released by glycopeptidase A digestion, 3485.5 Da (Figure 4B). The difference in masses correlates well with the loss of anhydro (Xyl)Man$_3$ (Fuc)GlcNAc$_2$ (calculated mass 1172 Da).

No peptide corresponding to this mass can be found in the partial sequence of SBP (Figure 1). The mass of the peptide containing a potential N-glycosylation site, Asn145, has a calculated mass of 2482.8 Da, far removed from the observed mass of the peptide arising from glycopeptidase A digestion of GP2. If the flanking regions, i.e., T11 and T13 or T13 and T14 are added to T12, the mass of the glycopeptide comes to 4702 Da, 45 Da greater than that determined by MALDI-TOF MS analysis (Figures 1B, 4B). It is greater than the mass of the released peptide by a similar amount. This difference in mass is too great to be accounted for by experimental error.

Thus, unlike the situation with GP1, GP3, GP4, and GP5/6, where the N-glycosylation sites are clearly identified, that associated with GP2 is not, and it is proposed that GP2 arises from the unsequenced amino-terminus of seed-coat SBP.

**Conclusions**

A previous study found that SBP contained 18.2% carbohydrate, sufficient to account for the presence of five or six glycosylation sites. The present study has confirmed and extended these findings and has mapped five of the six glycans onto the peptide chain and has demonstrated that Asn16, Asn90, Asn104, Asn169, and Asn174 are all glycosylated. The site at the N-terminal end, Asn16, has been clearly defined as being substituted mainly with high mannose-type glycans although there is evidence that this site may also be substituted with (Xyl)Man$_3$ (Fuc)GlcNAc$_2$ in different batches of SBP. Asn169 and Asn174 are both glycosylated but there is also evidence that both sites on this single polypeptide are not always substituted. Whether one or other of the sites is preferentially glycosylated was not determined in this study. The sixth site of glycosylation could not determined and is the subject of further investigation. The studies show that Asn145 is not glycosylated and that a peptide, GP2, derived from the unsequenced N-terminus of SBP, contains the sixth glycosylation site. This is the subject of ongoing investigations.

**Materials and methods**

**Purification of SBP and production of apo-enzyme**

Soybean seed coat peroxidase was purchased either from Enzymol (Enzymol International, Inc., Columbus, OH) or from Sigma (St. Louis, MO) with an initial RZ value of 1.3 and purified by a combination of gel filtration on Sephadex G-75, DEAE-cellulose, and FPLC on a Mono-Q column as described previously (Gray et al., 1996). The final RZ value obtained was >2.6.

The heme group was removed from SBP as described by Chibbar and van Huystee (1983) with acidic acetone (20 mM HCl in acetone). The apo-enzyme was lyophilized after extensive dialysis against water.

**Acetamidation of apo-SBP**

Apo-SBP (1.8 mg in 2 ml of 0.36 M Tris-HCl, 3 mM EDTA, pH 8.6) was reduced for 30 min in the dark at 37°C with 10 μl of 1 M dithiothreitol.

---

![Fig. 5. Evidence for the presence of (Xyl)Man$_3$ (Fuc)GlcNAc$_2$ on GP5/6. A glycopeptide purified by C$_8$ reverse phase chromatography from a tryptic digest of apo-SBP obtained from Enzymol was analyzed by MALDI-TOF MS. The signal at m/z 5283.8 correlates with the predicted mass of T1 substituted with (Xyl)Man$_3$ (Fuc)GlcNAc$_2$ (5286.5 Da).](image-url)
Acetamidation with 25 μl of 1 M iodoacetamide was allowed to proceed at room temperature for 30 min in the dark. After dialysis against deionized water (3 x 2 l), the acetamidated apo-SBP was isolated by lyophilization.

Trypsin digestion

Acetamidated apo-SBP (1.8 mg) or apo-SBP (1.5 mg) was dissolved in 200 μl of 100 mM NH₄HCO₃ buffer, pH 8.2, and denatured by heating for 5 min at 100°C. The samples were cooled and TPCK-treated trypsin (Worthington Biochemical Corporation, Freehold, NJ 07728) in 15 μl 0.1% TFA, 2 mM CaCl₂, was added to the acetamidated apo-SBP and apo-SBP to a final trypsin/protein concentration of 1% (w/w). Digestion proceeded for 4 h at 37°C and an equal aliquot of trypsin was added to each sample together with 10 μl of toluene. The incubation continued at room temperature for a further 19 h. The toluene was evaporated under a gentle stream of N₂, and the buffer was removed by lyophilization. The trypsin digests were dissolved in water (500 μl), and the lyophilization step was repeated. A trypsin blank for each digest was prepared in parallel.

Purification of glycopeptides

The trypsin digest of acetamidated SBP was fractionated by HPLC on a Vydac (Vydate/Sepration Group, Hesperia, CA) reverse-phase C₁₈ peptide and protein column (218TP, 300 Å, 250 x 4.6 mm) at a flow rate of 0.5 ml min⁻¹. Eluent A is 0.1% TFA and eluent B is 0.085% TFA in 90% acetonitrile. The column was equilibrated in 5% B in A and eluted as follows: 0-5 min, 5% B in A, 5-80 min, 5-59% B in A. The concentration of B in A was increased to 80% to wash the column before reequilibrating under initial conditions. Fractions were collected manually, dried either under a stream of nitrogen or under vacuum (Savant SpeedVac), dissolved in 200 μl water, and lyophilized.

Glycopeptidase A digestion

The carbohydrate was removed from the glycopeptides by digestion with glycopeptidase A (Seikagaku America, Inc., Rockville, MD). The enzyme was reconstituted in 100 mM citrate-phosphate buffer, pH 5.0, as directed by the company, aliquoted, and stored at -20°C. Before use, the buffer was exchanged for 100 mM NH₄HCO₃ buffer, pH 5.0, by centrifugation through a 10K MWCO MicroCon microcentrator ultrafilter (Amicon, Inc., Beverly, MA). The reaction mixture contained: glycopeptide A (0.2 μl) and glycopeptide (5–10 μg) in a final volume of 20 μl. After incubation at 37°C for 16–18 h under an atmosphere of toluene, the reaction was stopped by heating at 100°C for 5 min and the toluene evaporated under a stream of nitrogen. The buffer was removed by lyophilization, repeated once from water (50 μl), and the released peptides were analyzed by MALDI-TOF MS.

For HPLC analysis of the released peptides from unfractionated trypsin digests, the reaction mixture contained enzyme (0.5 μl) and trypsin digest (100–150 μg). The glycopeptidase A digestion and sample work-up was as described above. The sample was analyzed by HPLC on the Vydac C₁₈ reverse phase column as described above.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)

MALDI-TOF MS analyses were performed on a Perceptive Biosystems Voyager RP workstation (Framingham, MA). The analyses were conducted in the linear mode, unless otherwise stated, with an accelerating voltage of 30 kV. The vacuum was better than 6 x 10⁻⁷ torr. Glycopeptides and peptides were analyzed with α-cyano-4-hydroxycinnamic acid (10 mg/ml in 50% (v/v) acetonitrile in 0.1% TFA) or sinapinic acid (10 mg/ml in 33% (v/v) acetonitrile in 0.1% TFA) as a matrix. All samples were prepared by mixing the matrix and the sample, in that order, directly on the sample plate.

Acknowledgments

We particularly thank John S. Graham for sharing with us the partial protein sequence of SBP before publication. We also thank the Biotechnology By-products Consortium (USDA Grant No. 91–34188–5943), the Carbohydrate Structure Facility for the use of its equipment and David Venzke for technical assistance.

Abbreviations

FPLC, fast protein liquid chromatography; GP, glycopeptide-containing fractions; HPAEC-PAD, high pH anion exchange chromatography with pulsed amperometric detection; HPLC, high performance liquid chromatography; HRPc, horseradish peroxidase isozyme c; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; SBP, soybean holl peroxidase; TFA, trifluoroacetic acid; TPCK, L-(t-oxyamido-2-phenyl) ethyl chloromethyl ketone; RZ, ratio of the absorbance at 403 nm and 280 nm.

References


Received on November 19, 1996; revised on February 17, 1997; accepted on February 18, 1997.