Molekulare Nanotechnologie, Universität für Bodenkultur, A-1180 Wien, Zentrum für Ultrastrukturforschung und Ludwig Boltzmann-Institut für A-1180 Wien, Austria
Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia and
January 23, 1999
Michael Puchberger2, Paul Kosma2, Andrew A.Gooley1 and Paul Messner3

Zentrum für Ultrastrukturforschung und Ludwig Boltzmann-Institut für Molekulaire Nanotechnologie, Universität für Bodenkultur, A-1180 Wien, Austria. 
1Macquarie University Centre for Analytical Biotechnology, School of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia and 
2Institut für Chemie, Universität für Bodenkultur, A-1190 Wien, Austria
Received on November 23, 1998; revised on January 15, 1999; accepted on 
January 23, 1999

Structural heterogeneity in the core oligosaccharide of the S-layer glycoprotein from Aneurinibacillus thermoautophilus DSM 10155

Thomas Wugeditsch, Natasha E.Zachara1, Michael Puchberger2, Paul Kosma2, Andrew A.Gooley1 and Paul Messner3

The regularly arranged arrays of macromolecules are usually composed of a single proteinaceous subunit and cover the cell surface completely. For both archaea and bacteria, glycoproteins have been established as constituents of their S-layer (Sumper and Wieland, 1995; Messner, 1997; Messner and Schäffer, in press). Although the number of S-layer glycoproteins identified is small, structural analyses have revealed a great variety of glycan structures, constituents and even new protein–carbohydrate linkages, many of which are not known to occur in eukaryotes (Messner, 1996, 1997). Most of the investigated bacterial S-layer glycoproteins possess a tripartite structure similar to lipopolysaccharides (LPS) of Gram-negative bacteria (Raetz, 1996); they consist of strain-specific glycan chains composed of identical repeating units and different degrees of polymerization, which are linked by short core-oligosaccharides to the S-layer polypeptide via different protein–carbohydrate linkages (Messner et al., 1992, 1995; Schäffer et al., 1999).

Characterization of the S-layer glycoprotein glycan chain of the Gram-positive, thermophilic, aerobic bacterium, Aneurinibacillus (formerly Bacillus) thermoautophilus DSM 10155 (Heyndrickx et al., 1997) confirmed this architectural principle. In our systematic survey of glycosylated S-layer proteins A.thermoautophilus DSM 10155 initially was used as a reference strain for the taxonomic characterization of isolates from the extraction plant of an Austrian beet sugar factory (Meier-Stauffer et al., 1996). This strain was chosen for further analysis because of the unusual properties and composition of its S-layer glycoprotein which have been published recently (Kosma et al., 1995b). d-Glycero-d-manno-heptose, a typical component of LPS of Gram-negative bacteria (Raetz, 1996), had been found in this organism for the first time as a constituent of cell surface carbohydrates from a Gram-positive bacterium.

In this study we report on the structural heterogeneity of the linkage region of the S-layer glycoprotein glycans of A.thermoautophilus DSM 10155; this includes both the structure of the core oligosaccharides of the S-layer glycans and their carbohydrate–protein linkages. Using modified Edman-degradation protocols (Zachara and Gooley, in press), alkali-labile O-glycosidic linkages of N-acetyl-galactosamine (GalNAc) to threonine (Thr) and serine (Ser) have been identified. Unlike the linkage of GalNAc in eukaryotic glycoproteins (an α-O-linkage; for review, see Vliegenthart and Montreuil, 1995), GalNAc appears to be linked to the S-layer protein backbone through a β-O-linkage. The modification of Ser/Thr with GalNAc residues (either α or β anomer) has not been observed before in S-layer glycoproteins and also not in any other prokaryotic glycoprotein (Messner, 1997). NMR analyses of Smith-degradation products of fluorescently labeled glycopeptides and liquid chromatography mass-spectrometry (LC-ESI-MS) analyses of the cleaved glycosylated amino acids after Edman degradation led to the elucidation of the structure of the different core oligosaccharides of the S-layer glycoprotein glycans of A.thermoautophilus DSM 10155.

Introduction

Crystalline bacterial cell surface-layers (S-layers) are one of the most common cell surface structures of archaea and bacteria (Messner and Sleytr, 1992; Beveridge, 1994; Sleytr et al., 1996). The regularly arranged arrays of macromolecules are usually composed of a single proteinaceous subunit and cover the cell surface completely. For both archaea and bacteria, glycoproteins have been established as constituents of their S-layer (Sumper and Wieland, 1995; Messner, 1997; Messner and Schäffer, in press). Although the number of S-layer glycoproteins identified is small, structural analyses have revealed a great variety of glycan structures, constituents and even new protein–carbohydrate linkages, many of which are not known to occur in eukaryotes (Messner, 1996, 1997). Most of the investigated bacterial S-layer glycoproteins possess a tripartite structure similar to lipopolysaccharides (LPS) of Gram-negative bacteria (Raetz, 1996); they consist of strain-specific glycan chains composed of identical repeating units and different degrees of polymerization, which are linked by short core-oligosaccharides to the S-layer polypeptide via different protein–carbohydrate linkages (Messner et al., 1992, 1995; Schäffer et al., 1999).

Characterization of the S-layer glycoprotein glycan chain of the Gram-positive, thermophilic, aerobic bacterium, Aneurinibacillus (formerly Bacillus) thermoautophilus DSM 10155 (Heyndrickx et al., 1997) confirmed this architectural principle. In our systematic survey of glycosylated S-layer proteins A.thermoautophilus DSM 10155 initially was used as a reference strain for the taxonomic characterization of isolates from the extraction plant of an Austrian beet sugar factory (Meier-Stauffer et al., 1996). This strain was chosen for further analysis because of the unusual properties and composition of its S-layer glycoprotein which have been published recently (Kosma et al., 1995b). d-Glycero-d-manno-heptose, a typical component of LPS of Gram-negative bacteria (Raetz, 1996), had been found in this organism for the first time as a constituent of cell surface carbohydrates from a Gram-positive bacterium.

In this study we report on the structural heterogeneity of the linkage region of the S-layer glycoprotein glycans of A.thermoautophilus DSM 10155; this includes both the structure of the core oligosaccharides of the S-layer glycans and their carbohydrate–protein linkages. Using modified Edman-degradation protocols (Zachara and Gooley, in press), alkali-labile O-glycosidic linkages of N-acetyl-galactosamine (GalNAc) to threonine (Thr) and serine (Ser) have been identified. Unlike the linkage of GalNAc in eukaryotic glycoproteins (an α-O-linkage; for review, see Vliegenthart and Montreuil, 1995), GalNAc appears to be linked to the S-layer protein backbone through a β-O-linkage. The modification of Ser/Thr with GalNAc residues (either α or β anomer) has not been observed before in S-layer glycoproteins and also not in any other prokaryotic glycoprotein (Messner, 1997). NMR analyses of Smith-degradation products of fluorescently labeled glycopeptides and liquid chromatography mass-spectrometry (LC-ESI-MS) analyses of the cleaved glycosylated amino acids after Edman degradation led to the elucidation of the structure of the different core oligosaccharides of the S-layer glycoprotein glycans of A.thermoautophilus DSM 10155.

© 1999 Oxford University Press

Glycobiology vol. 9 no. 8 pp. 787–795, 1999

Key words: core oligosaccharide/eubacteria/glycoprotein/heterogeneity/surface-layer (S-layer)

Introduction

Crystalline bacterial cell surface-layers (S-layers) are one of the most common cell surface structures of archaea and bacteria (Messner and Sleytr, 1992; Beveridge, 1994; Sleytr et al., 1996). The regularly arranged arrays of macromolecules are usually composed of a single proteinaceous subunit and cover the cell
Results

Characterization of S-layer glycopeptides GP A and GP B

As reported previously (Kosma et al., 1995b), two different types of glycopeptides (GP A and B) were obtained after proteolytic digestion of isolated S-layer glycoproteins by Pronase E, followed by isolation and purification of the resulting glycopeptides by gel permeation chromatography, cation-exchange chromatography, chromatofocusing, and RP-HPLC. These glycopeptides differed in their amino acid compositions but carried identical glycans. Based on the results of composition analysis, chemical degradation experiments, NMR spectroscopy, and comparison with a synthesized model substance, the disaccharide \((1\rightarrow4)\)-\(\alpha\)-L-Rhap-\((1\rightarrow3)\)-\(\beta\)-D-glycero-D-manno-Hepp\) was identified as the repeating unit structure (Kosma et al., 1995b). Combined evidence of total carbohydrate content of the S-layer glycoprotein (~15%), the molecular mass of the polypeptide backbone (75 kDa), and the average mass of the glycan chain composed of disaccharide repeating units determined by SDS–PAGE and MALDI-TOF-MS revealed that both glycosylation sites occur only once on the S-layer polypeptide.

<table>
<thead>
<tr>
<th>Component</th>
<th>GP Aa</th>
<th>GP B</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)-L-Rhap</td>
<td>18–24</td>
<td>18–25</td>
</tr>
<tr>
<td>(\beta)-D-glycero-D-manno-Hepp</td>
<td>n.q.(^b)</td>
<td>n.q.(^b)</td>
</tr>
<tr>
<td>GalNAc</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Thr</td>
<td>2.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Ser</td>
<td>—c</td>
<td>—</td>
</tr>
<tr>
<td>Asn</td>
<td>1.1</td>
<td>—</td>
</tr>
<tr>
<td>Glu</td>
<td>—</td>
<td>1.0</td>
</tr>
<tr>
<td>Gly</td>
<td>—</td>
<td>1.0</td>
</tr>
<tr>
<td>Lys</td>
<td>1.0</td>
<td>—</td>
</tr>
</tbody>
</table>

a Values are expressed as molar ratios, with GalNAc arbitrarily set to 1.0.

b Not quantified.

c —, Not found.

Due to the mode of glycopeptide preparation in this study, results of amino acid analyses of the glycopeptides varied slightly but indicated two different sets of glycopeptides (Table I). In both cases glycan chains were released by base-catalyzed \(\beta\)-elimination. Cleavage products were separated by cation-exchange chromatography and subjected to composition analysis. Although under the conditions applied in \(\beta\)-elimination no quantitative reduction was attained, the observed loss of Thr and Ser together with the identification of galactosaminitol suggested a GalNAc-Thr linkage in the case of GP A and a GalNAc-Ser linkage for GP B. These results were supported by the liberation of the S-layer glycans by aqueous hydrazinolysis or \(\beta\)-elimination under nonreductive conditions and subsequently fluorescent labeling of the reducing end of released glycan chains. In addition to 2-aminobenzamide (2-AB)-labeled GalNAc, labeled Rha was also detected by thin-layer chromatography after hydrolysis of isolated glycan chains (data not shown). Under these conditions cleavage of N-acetylgalactosaminylhydroxyamino acid linkages results in the formation of GalNAc as the reducing sugar. This product is not stable in alkaline environment and may undergo another \(\beta\)-elimination reaction forming a new reducing end. This stepwise degradation (“peeling reaction”) (Whistler and Miller, 1958; Lloyd et al., 1968) led to some information about the monosaccharide sequence of the glycan chain. The suggested types of linkages of GP A and GP B were finally confirmed by modified Edman-degradation which not only revealed the peptide sequence but also allowed the site identification of the glycosylated amino acid (Figure 1). The amino acid sequences of GP A and GP B are shown in Table II.

Table I. Chemical composition of glycopeptides A and B

Table II. Sequence of glycopeptides A and Ba

<table>
<thead>
<tr>
<th>Component</th>
<th>GP A</th>
<th>GP B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr-Thr</td>
<td>Asn-Xaa-Lys</td>
<td></td>
</tr>
<tr>
<td>Ser-Gly</td>
<td>Thr-Xaa-Xaa</td>
<td></td>
</tr>
</tbody>
</table>

a Glycosylated amino acids are shown in boldface; Xaa represents not identified amino acids.
Structural heterogeneity of S-layer glycoprotein cores

To elucidate the structure of the core oligosaccharide, phenylthiohydantoin (PTH)-amino acids generated by Edman degradation of glycopeptides were subjected to LC-ESI-MS analysis. In addition, dansyl chloride-labeled glycopeptides were subjected to Smith degradation. Degradation products were fractionated by gel permeation chromatography over Bio-Gel P-2, P-4, and P-6 and analyzed by thin-layer chromatography and high-performance anion-exchange chromatography with pulsed electrochemical detection (HPAEC-PED). Fractions of interest were finally purified by RP-HPLC and one of those obtained from dansylated GP B was chosen for NMR analysis.

**LC-ESI-MS analysis of the glycosylated PTH-amino acids of GP A and GP B**

Glycosylated amino acids released by Edman degradation (cycles 1 and 2 for GP A and cycle 1 for GP B) were manually collected prior to injection onto the HPLC (~60 pmol of GP A and ~40 pmol of GP B) and were analyzed by LC-ESI-MS. Analysis of the PTH-amino acids of GP A by LC-ESI-MS (Figure 2) showed that many of the PTH-amino acids contained the diagnostic sugar oxonium ion for HexNAc (204 m/z ion), which is released by in-source collision induced dissociation (CID) and detected using selected ion monitoring (Carr et al., 1993). Examination of this part of the mass spectrum showed many mass species which represent the three following core structures (see Figure 3): PTH-Thr-HexNAc (438 m/z), PTH-Thr-HexNAc-DeoxyHex (584 m/z), and PTH-Thr-HexNAc-DeoxyHex-DeoxyHex (730 m/z).

Each core structure was modified by a varying number of the repeating unit α-D-glycero-β-D-manno-Hep-DeoxyHex (338 m/z; see Table IIIA). Cleavage between the α-D-glycero-β-D-manno-Hep and DeoxyHex was also observed. Variations in the number of repeats was probably due to acid hydrolysis during conversion of the anilinothiazolinone (ATZ)-amino acid to the PTH-amino acid, which is carried out in 33% (v/v) trifluoroacetic acid (TFA), at 70°C for 6 min 40 s. Scans of the major mass species (Figure 4) identified the nature of the peaks in the RP-HPLC chromatogram and allowed the quantitation of the relative amounts of each core structure. Interestingly, there appears to be equal amounts of each structure (see Table IV).

Similar results were found for GP B, except that the glycan was bound to Ser (see Table IIIIB). Table IIIIB presents the mass species detected for GP B, again indicating the presence of three core structures, which are modified by differing numbers of repeating units. These results, taken together with those of the sugar analysis (Kosma et al., 1995b), would suggest that the identity of the DeoxyHex residue is Rha and that of the HexNAc residue is GalNAc (Table I). However, the surprising result was that the glycosylated PTH-amino acids identified in GP A and GP B did not coelute with PTH-Ser or PTH-Thr modified by a reducing terminal α-GalNAc typical of the mucin-like glycopeptides (see Figure 1), which suggests that one possible linkage of the GalNAc to Ser or Thr is through a β- O-linkage. To establish the anomeric and relative configurations of the respective carbohydrate moieties as well as their linkage sites and confirm the MS data, exemplary NMR analysis on a purified peptide-core-oligosaccharide fragment obtained after Smith degradation of the dansylated intact glycopeptide pool was performed.

Fig. 2. LC-ESI-MS analysis of the PTH-amino acids recovered from cycles 1 and 2 of GP A. The absorbance trace at 269 nm is shown in (A), and the acetonitrile gradient from 5–80% (v/v) is indicated. (B) shows the total ion current acquired in the negative ionization mode, while (C) shows the SIM trace for the HexNAc ion 204 m/z. Areas positive for this ion were combined and the spectrum is shown in Figure 3.

Fig. 3. Different glycoforms of PTH-Thr detected by LC-ESI-MS of GP A. Each core structure produced both singly and doubly charged ions. Each mass species is denoted by two numbers, for example 1–2 represents the core structure 1 with two repeats and depending on the charge state the same structure may be represented twice. Core structure 1, β-D-GalNAc-O-Thr; Core structure 2, α-L-Rhap-(1–3)-β-D-GalNAc-O-Thr; and Core structure 3, α-L-Rhap-(1–3)-α-L-Rhap-(1–3)-β-D-GalNAc-O-Thr. Table IIIA itemizes each structure found.
Table III. Mass of glycosylated PTH-amino acids of glycopeptide A (A) and glycopeptide B (B)

<table>
<thead>
<tr>
<th>Number of repeats</th>
<th>Core 1</th>
<th>Core 2</th>
<th>Core 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z1-</td>
<td>Z2-</td>
<td>Z1-</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>438</td>
<td>584</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>776</td>
<td>922</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1114</td>
<td>1260</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>—</td>
<td>725</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>—</td>
<td>894</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>—</td>
<td>1063</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>—</td>
<td>1232</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>424</td>
<td>570</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>762</td>
<td>908</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1100</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>—</td>
<td>718</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>—</td>
<td>887</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>—</td>
<td>1056</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>—</td>
<td>1225</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Core 1 represents β-D-GalNAc-O-Thr/Ser, Core 2 α-L-Rhap-(1–3)-β-D-GalNAc-O-Thr/Ser and Core 3 α-L-Rhap-(1–3)α-L-Rhap-(1–3)-β-D-GalNAc-O-Thr/Ser. All masses are reported as mass/charge (m/z) ratios, the charge state is indicated. Note that the mass of the repeat unit is 338.

Table IV. Quantitation of different core structures on glycopeptides A and B

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Core 1 (mAU²)</th>
<th>Core 2 (mAU²)</th>
<th>Core 3 (mAU²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP A</td>
<td>12.23</td>
<td>10.67</td>
<td>12.02</td>
</tr>
<tr>
<td>GP B</td>
<td>3.94</td>
<td>3.44</td>
<td>4.46</td>
</tr>
</tbody>
</table>

The peak area under the PTH-amino acids carrying Core 1, Core 2, and Core 3 (identified in Figure 3) were calculated. While less of the Core 2 structure exists in both GP A and GP B, there are similar levels of each core structure.

NMR analysis of the dansylated Smith-degraded glycopeptide B

Assignments of the carbohydrate signals (units A-C). The 300 MHz 1H NMR spectrum (Figure 5) of the dansylated Smith-degraded GP B was recorded at 300 K and displayed, inter alia, low-field shifted signals in the region of 8.58–7.80 p.p.m. (6 H, aromatic protons of the dansyl group), three anemic protons at 5.00 p.p.m. (H-1A, J1,2 1.6 Hz), 4.81 p.p.m. (H-1B, J1,2 1.7 Hz), and 4.42 p.p.m. (H-1C, J1,2 8.8 Hz), and four upfield-shifted signals corresponding to methyl groups at 1.98 p.p.m. (3 H, singulet), 1.28 and 1.25 p.p.m. (6 H, two doublets) and 1.15 p.p.m. (less than 3 H by integration, doublet). Since the signal of the anemic proton of residue B was close to the residual water signal, the spectrum was also recorded at 330 K which gave the signal of H-1B in an intensity similar to that of H-1A. Whereas the small values of the homonuclear coupling constants J1,2 of the anemic protons of unit A and B were indicative of manno-configured systems, the large value observed for the homonuclear coupling constant of H-1C could be attributable to a β-glucopyranosyl or galacto-configured system. Furthermore, the signals at 1.28 and 1.26 p.p.m. showing a coupling constant J of 5.6 Hz were assigned to 6-deoxy protons of two deoxy sugars, whereas the doublet at 1.15 p.p.m. (J 6.2 Hz), which had only 70% of the signal-intensity of the neighboring methyl groups, was assigned to the methyl group of a threonine residue.

Further assignments were achieved by gradient-enhanced inverse homo- and heteronuclear correlation spectroscopy (COSY, TOCSY, HMQC) which allowed the establishment of proton–proton connectivity (Figure 6) and proton–carbon correlations. Analysis of the coupling constants observed for the carbohydrate ring protons indicated that the three carbohydrate moieties occurred in the chair conformation. In addition, residues A and B were assigned to rhamnopyranosyl units, whereas unit C was identified as a 2-acetamido-2-deoxy-galactopyranosyl residue in the β-anomeric configuration (Table V).

The small amount of material did not allow a reliable measurement of the heteronuclear coupling constant JCH of the rhamnopyranosyl residues for the assignment of the anomeric configuration. The chemical shift values of the anomeric carbons of
Fig. 5. 300 MHz proton NMR spectrum of the dansylated, Smith-degraded glycopeptide B recorded at 300 K and part of the anomeric region recorded at 330 K.

Moreover, since the signals of all three anomeric carbons were in the range >100 p.p.m., the galactosamine unit had to occur in a configuration different from the rhamnosyl units, otherwise an upfield-shift for the signal of one anomeric carbon (C-1B) would be expected (Shashkov et al., 1988; Vinogradov et al., 1994). Thus, in the glycopeptide core-oligosaccharide of a related S-layer glycoprotein from *Aneurinibacillus thermoaerophilus* GS4-97 containing the sequence \(\alpha\)-D-Rha\((1\rightarrow3)\)-\(\alpha\)-D-Rha\((1\rightarrow3)\)-\(\beta\)-D-GalpNAc\((1\rightarrowO)\)-Thr, the internal Rhap unit displayed the signal of the anomeric carbon at 97.09 p.p.m. (Schäffer et al., 1999).

In the HMQC spectra two downfield-shifted signals for glycosylated carbons were observed at 79.5 and 80.7 p.p.m., respectively. The signal at 80.7 p.p.m. was unambiguously assigned to carbon 3 of the GalNAc residue. The \(^\text{13}\)C NMR data of the GalNAc unit C substituted at position 3 by an \(\alpha\)-L-Rhap residue are in close agreement with published values for a polysaccharide repeating unit \(\rightarrow3\)-\(\alpha\)-L-Rhap\((1\rightarrow3)\)-\(\beta\)-D-GalpNAc\((1\rightarrowO)\)-from *Serratia marcescens* O9 (Oxley and Wilkinson, 1987).

Due to the close proximity of H-2B and H-3B at 3.87 p.p.m. in the proton domain, the second downfield-shifted carbon could not be unambiguously correlated. Since, however, a 2-substituted Rhap group would have been destroyed during the previous Smith degradation, the signal at 79.5 p.p.m. was assigned to C-3 of residue B. In addition, no \(\beta\)-shift was observed for C-1B which would be observed upon substitution at the 2-position.

ROESY spectra displayed interresidue ROEs of H-1A to H-3B, of H-1B to H-3/4C and H-1C to C-2 of a serine unit. Thus, the sequence \(\alpha\)-L-Rhap\((1\rightarrow3)\)-\(\alpha\)-L-Rhap\((1\rightarrow3)\)-\(\beta\)-D-GalpNAc\((1\rightarrowO)\)-serine is compatible with the NMR data (Figure 7).

Fig. 6. TOCSY spectrum of the dansylated, Smith-degraded glycopeptide B.
Table V. NMR data for the dansylated glycopeptide from *Aneurinibacillus thermoaurphilus* DSM 10155

<table>
<thead>
<tr>
<th>Atom</th>
<th>H/C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>6′</th>
<th>CO</th>
<th>Other signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-Rha</td>
<td>1H</td>
<td>5.00</td>
<td>4.05</td>
<td>3.79</td>
<td>3.45</td>
<td>3.79</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>J</td>
<td>1.65</td>
<td>3.3</td>
<td>9.6</td>
<td>9.7</td>
<td>5.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13C</td>
<td>103.6</td>
<td>70.9</td>
<td>70.7</td>
<td>72.5</td>
<td>70.9</td>
<td>17.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→(3)-α-Rha</td>
<td>1H</td>
<td>4.81</td>
<td>3.87</td>
<td>3.87</td>
<td>3.50</td>
<td>3.84</td>
<td>1.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>J</td>
<td>1.7</td>
<td>−3.0</td>
<td>9.7</td>
<td>9.7</td>
<td>5.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13C</td>
<td>103.6</td>
<td>71.1</td>
<td>79.5</td>
<td>72.1</td>
<td>70.7</td>
<td>17.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→(3)-β-GalNAc</td>
<td>1H</td>
<td>4.42</td>
<td>3.88</td>
<td>3.68</td>
<td>3.95</td>
<td>3.60</td>
<td>3.77</td>
<td>3.72</td>
<td>1.98 (NHAc)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>J</td>
<td>8.8</td>
<td>n.d.</td>
<td>3.1</td>
<td>&lt;1.0</td>
<td>n.d.</td>
<td>n.d.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13C</td>
<td>103.1</td>
<td>52.3</td>
<td>80.7</td>
<td>68.9</td>
<td>76.0</td>
<td>61.8</td>
<td>175.0</td>
<td>23.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>→(O)-Ser-Dansyl</td>
<td>1H</td>
<td>4.12</td>
<td>3.90/3.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>J</td>
<td>6.2</td>
<td>n.d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13C</td>
<td>57.6</td>
<td>69.8</td>
<td>n.d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>1H</td>
<td>3.57 (2H)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>13C</td>
<td>43.4</td>
<td>171.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>1H</td>
<td>4.42</td>
<td>2.13/1.90</td>
<td>2.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>J</td>
<td>8.6</td>
<td>7.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13C</td>
<td>53.8</td>
<td>26.7/27.3</td>
<td>31.7</td>
<td>178.5</td>
<td>n.d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>1H</td>
<td>4.28</td>
<td>4.29</td>
<td>1.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>J</td>
<td>4.9</td>
<td>6.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13C</td>
<td>60.7</td>
<td>69.3</td>
<td>20.1</td>
<td>n.d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dansyl group</td>
<td>1H</td>
<td>—</td>
<td>8.37</td>
<td>7.80</td>
<td>8.49</td>
<td>—</td>
<td>7.83</td>
<td>7.84</td>
<td>8.58</td>
<td>3.27</td>
</tr>
<tr>
<td>13C</td>
<td>n.d.</td>
<td>131.8</td>
<td>127.1</td>
<td>129.1</td>
<td>144.8</td>
<td>119.0</td>
<td>129.8</td>
<td>124.8</td>
<td>47.11</td>
<td></td>
</tr>
</tbody>
</table>

*Spectra were recorded at 300 K, chemical shifts were referenced against internal DDS for 1H and external 1,4-dioxane (δ 67.40 p.p.m.) for 13C, coupling constants are in Hz.*

Assignment of the peptide portion. In addition to one nitrogen-linked carbon connected to C-2 of the GalNAc unit, four additional C-N signals attributable to α-carbons of amino acids were found (Table V). The 1H NMR signals observed for the amino acid residues occurred in two different intensities, indicating that a second species was present in the investigated glycopeptide pool in a proportion of ~30%. The inspection of the connectivities allowed the assignment of an unsubstituted threonine unit (G) in the major species (~70%) of the mixture. Moreover, two different sets of signals (also in a ~2.5:1 ratio) were seen for signals attributable to a Glu or Glu residue (F) and for a Gly residue (E) as well, which were assigned by comparison with literature data (Dill *et al.*, 1985). An explanation for the discrepancy between the sequencing and the NMR results could be that in the course of Smith-degradation glutamic acid residues undergo an amidation reaction, resulting in a conversion to Gln. Finally, a Ser residue (D) was identified which, based on the downfield-shifted signal for the O-linked CH2-group at 69.8 p.p.m. and the observed interresidue ROE from H-1C, had to be glycosylated by the GalNAc residue.

MALDI-TOF-MS analysis of the mixture showed the presence of two molecular ions (M+Na+) at m/z 1145.1 and 1043.9 which corresponds to the calculated mass of the dansylated glycopeptide (1145.2) containing Ser, Gly, Gln (or Lys), and Thr and a second moiety without a terminal Thr residue (1044.1), respectively. Hence the peptide sequence Ser-Gly-Gln-Thr or Ser-Gln-Gly-Thr may be derived from these data. An explanation for the discrepancy between sequencing and the MALDI-TOF results could be that in the course of Smith-degradation glutamic acid residues undergo amidation reaction, resulting in a conversion to Gln.

Discussion

In previous studies we have shown that a remarkable heterogeneity exists in ultrastructural and biochemical properties of glycosylated S-layer proteins (for review, see Messner, 1996). This diversity is
also present in S-layer glycoproteins of the newly described genus *Aneurinibacillus* (Meier-Stauffer et al., 1996; Heyndrickx et al., 1997). The glycan repeating unit structures of different members of the species *A.thermoaerophilus* have been compared and showed considerable differences. Strains L420–91T and GS4-97 possess branched hexaasaccharide repeats (Kosma et al., 1995a; Schäffer et al., in press) whereas the glycans of strain DSM 10155 are assembled of disaccharide repeats with the structure \( \alpha-L-\text{Rha} \rightarrow (1 \rightarrow3) \beta-D-\text{glycero-}D-
\text{manno}-\text{Hepp} \rightarrow (1 \rightarrow \) (Kosma et al., 1995b).

In this study, we have focused our efforts on the elucidation of the structures of core and linkage region of strain *A.thermoaerophilus* DSM 10155. So far, cores of S-layer glycoproteins have only been examined in a few archaea (for review, see Sumper and Wieland, 1995) and in thermophilic strains of the *Bacillaceae* (for review, see Messner, 1996). Of particular interest was the observation of novel linkage constituents of O-linked oligosaccharides such as \( \beta-D-\text{glucose} \rightarrow \text{tyrosine} \) (Messner et al., 1992) and \( \beta-D-\text{galactose} \rightarrow \text{tyrosine} \) (Bock et al., 1994; Messner et al., 1995). However, in strain *A.thermoaerophilus* DSM 10155 identical glycan chains are linked to different glycosylation sites of the S-layer protein backbone by two novel types of O-glycosidic linkages to GalNAc. In the case of GP A the acceptor amino acid was found to be Thr, and for GP B it is Ser. Interestingly, the Ser and Thr linkages in either prokaryotes or eukaryotes, except \( \alpha-L-\text{Rha} \rightarrow \alpha-D-\text{galactose} \rightarrow \text{N-acetylglucosamine} \), as it is known from mucin-like glycosylation (Sharon and Lis, 1997), but \( \beta-D-\text{GalNAc} \) and the \( \beta-D-\text{galactose} \rightarrow \beta-D-\text{glucose} \) that the core structure of the S-layer glycoprotein of the closely related strain *Thermoanaerobacter thermosaccharolyticum* E207–71 was shown to possess also a \( \beta-D-\text{GalNAc} \rightarrow (1 \rightarrow \text{N})-\text{Thr} \) linkage (Schäffer et al., 1999).

In addition to the novel linkage in strain DSM 10155 another interesting feature of its S-layer glycoprotein is the obvious heterogeneity in the core structure. LC-ESI-MS analyses of glycosylated amino acids obtained by a modified Edman-degradation procedure (Zachara and Gooley, in press) showed that the core structure \( \alpha-L-\text{Rha} \rightarrow (1 \rightarrow3) \alpha-L-\text{Rha} \rightarrow (1 \rightarrow3) \beta-D-\text{GalpNAc} \rightarrow (1 \rightarrow\text{O})-\text{Thr}/\text{Ser} \) (Figure 7) is present in equal amounts together with truncated core forms, missing either one or both of the Rha residues. So far, only oligosaccharide chain length variation of S-layer glycans has been examined in detail in *Thermoanaerobacter thermohydrodsulfuricus* L111–69 (Bock et al., 1994). They revealed a variation in the degree of polymerization of the glycan building blocks between approximately 20 and 35 units. There was no indication of a microheterogeneity of the core glycan. Preliminary examination of the S-layer glycopeptide from *Thermoanaerobacter thermosaccharolyticum* E207–71 showed that only one type of core oligosaccharide is present in this organism (C.Schäffer and N.E.Zachara, unpublished observations).

Since similarities between S-layer glycan biosynthesis and O-antigen synthesis of LPS are expected (Messner, 1996; Schäffer et al., 1996), comparable mechanisms for the S-layer glycan synthesis seem to be possible. The consequence for the formation of a complete S-layer glycan chain in *A.thermoaerophilus* would be that to either type of existing core lipid-bound repeating units, as they are known from LPS biosynthesis (Raetz, 1996), should be attached. Despite the heterogeneity this transfer could be facilitated by a relatively simple system such as a single putative transferase because for both the \( \beta-D-\text{GalpNAc} \) and the \( \alpha-L-\text{Rha} \) residue the spatial arrangement for an attachment of ligands to position 3 is very similar (Shashkov et al., 1988).

Further biochemical and genetic analyses will be required to unravel the pathways for S-layer glycan synthesis in Gram-positive bacteria.

### Materials and methods

#### Materials

Arylamine disks, Sequelon AA, were from PerSeptive Biosystems Division of Perkin Elmer (Foster City, CA). Protein sequencing (Buffers and Chemicals) were from Hewlett Packard (Palo Alto, CA). All other chemicals were of analytical grade.

#### Isolation of S-layer glycoprotein and S-layer glycopeptides

S-Layer glycoprotein and S-layer glycopeptides were isolated as described previously (Kosma et al., 1995b) with the exception of RP-HPLC chromatography of the glycopeptides, which was performed as described by Schäffer et al. (1999).

#### Glycan release

**β-Elimination under reducing conditions.** Alkaline-catalyzed \( \beta \)-elimination was performed by incubation in 0.1 M NaOH and 1 M NaBH4, pH 11, for 24 h at 45°C. The reaction was neutralized with 2 M acetic acid and passed over a Dowex 50W-X8 (H\(^+\)) column. The released glycan chains containing effluent and water wash from the column were combined and lyophilized. Peptides and uncleaved glycopeptides were eluted with a step gradient of ammonia (2 M and 5 M NH\(_3\)), dried under vacuum, and separated by gel permeation chromatography over Sephadex G-10.

**β-Elimination under nonreducing conditions.** S-layer glycopeptides were solubilized in 2 M Na\(_2\)CO\(_3\), pH 12, at a concentration of 3 mg/ml and incubated for 24 h at 45°C. Neutralization and subsequent isolation of the cleavage products were performed as described for \( \beta \)-elimination under reducing conditions.

#### Aqueous hydrazinolysis

Release of O-linked glycans in unreduced form was also achieved by aqueous hydrazinolysis using the procedure described by Cooper et al. (1994). The obtained glycan hydrazones were deblocked using the aqueous acetone method according to Cooper et al. (1994).

#### Fluorescent labeling

**2-Aminobenzamide labeling of reducing sugars.** S-layer glycans released by \( \beta \)-elimination under nonreducing conditions or aqueous hydrazinolysis were fluorescent-labeled with 2-AB according to Townsend et al. (1996). After neutralization of the reaction products with 1 M NaOH labeled glycans were purified by gel permeation over Bio-Gel P-2.

**Dansylation of S-layer glycopeptides.** Intact S-layer glycopeptides were solubilized in 40 mM Li\(_2\)CO\(_3\), pH 9.0, at a concentration of 5 mg/ml; 1/15 of the total volume of 10% (w/v) dansylchloride (5-dimethylaminonaphthalene-1-sulfonyl chloride) in acetone was added and the mixture was incubated for 30 to 45 min at 37°C in the dark. Free dansylchloride was removed by addition of 1/15 volume of ethanolamine. After incubation at room temperature for 20 min the solvents were evaporated under vacuum, and S-layer glycopeptides were purified by gel permeation chromatography over Bio-Gel P-2. Non-dansylated glycopeptides were separated by RP-HPLC and subjected to another dansylation reaction. Collected fluorescently labeled glycopeptides were combined.

#### Smith degradation

Chemical degradation of the glycan chain of fluorescently-labeled S-layer glycopeptides by Smith-degradation was performed.
according to Goldstein et al. (1965). Dansylated glycopeptides were incubated in 0.1 M sodium metaperiodate in 0.1 M sodium acetate, pH 4.5, for 6 days at 4°C in the dark. Excess periodate was destroyed by addition of ethylene glycol. Oxidized glycopeptides were separated by gel permeation chromatography over Bio-Gel P-2 and subsequently reduced with 1 M NaBH₄ at pH 8.0 (adjusted with ammonia) overnight. After neutralization with 2 M acetic acid the reaction product was finally dried under reduced pressure. Boric acid was removed with methanol (Albersheim et al., 1967). Smith-type hydrolysis was performed for 3 h in 2% acetic acid at 110°C followed by 3 days in 4 M TFA at room temperature. After evaporation of the acids, degradation products were fractionated by gel permeation chromatography over Bio-Gel P-2, P-4, and P-6 and analyzed by thin-layer chromatography. Dansylated degradation products were finally purified by RP-HPLC using the conditions applied for unlabeled glycopeptides. The eluant was monitored by adsorption at 220 nm and by fluorescence detection with an excitation wavelength of 350 nm and an emission wavelength of 510 nm, using an in-line fluorimeter (RF-551, Shimadzu).

**Analytical methods**

Amino acid analyses were performed according to Bock et al. (1994). Carbohydrate composition analyses were performed by HPAEC-PED as described recently (Kosma et al., 1995b). Thin-layer chromatography of dansylated Smith-degradation products was carried out on glass backed Silica Gel 60 plates (0.25 mm thickness, Merck) in n-butanol/pyridine/0.1 N HCl, 5:3:2. Thin-layer chromatography of 2-AB labeled monosaccharides was performed after hydrolysis of labeled glycopeptides with 4 M TFA at 110°C for 4 h using n-butanol/acetic acid/water, 3:2:1, and n-butanol/ethanol/water, 4:1:1 as solvent system.

**Edman degradation**

Attachment of GP A and GP B to arylamine support disks. GP A (~ 300 µg) and GP B (~ 200 µg) were attached via their carboxyl groups to arylamine activated PVDF membrane disks (Sequelon-AA), using water soluble N-ethyl-N’-3-dimethylaminopropyl-carbodiimide (EDC). Briefly, samples in 20% (v/v) acetonitrile/1 M NaCl were applied to an arylamine disk and dried at 50 °C. One milligram of EDC was dissolved in 50 µl of coupling buffer, and this was applied to the disk. The sample was incubated at 4°C to increase the coupling yield (Laursen et al., 1991), for 30 min, before being washed (3×) alternately with 100% methanol and Milli Q water.

**Edman sequencing.** Covalently coupled GP A and GP B were sequenced on a Hewlett Packard G1000A protein sequencer (HP G1000A) using the sequencing program, 3.1 PVDF, which was modified to include a methanol extraction from the column to the converter (Zachara and Gooley, in press). The PTH-amino acids were either analyzed “on-line” as part of the normal sequencing routine or “off-line” by LC-ESI-MS. For analysis using LC-ESI-MS, 60% of the glycosylated amino acids (~60 pmol GP A and ~40 pmol GP B, calculated from the initial yield) was removed from the sequencer prior to injection onto the HPLC.

**LC-ESI-MS of the PTH-glycosylated amino acids from GP A and GP B**

PTH-amino acids collected during Edman degradation (~60 pmol GP A and ~40 pmol GP B) were separated on a Pharmacia SMART HPLC equipped with a Shandon HPLC Beckman Hypersil amino acid C₁₈ reversed phase column (3 µm, 2.1 mm × 250 mm) equilibrated in 5 mM triethyl ammonium formate, 5% (v/v) acetonitrile. This buffer system is compatible with LC-ESI-MS (Pisano et al., 1995) although the PTH-amino acids do have a different elution pattern using this solvent system, from those obtained using the HP G1000A PTH-amino acid separation buffers. The concentration of acetonitrile was raised to 27% (v/v) after the salt eluted at 7.5 min. Following this, the concentration of acetonitrile was raised to 80% (v/v) using a nonlinear acetonitrile gradient (indicated in Figure 2) to elute the PTH-amino acids from the column. All flow rates were 100 µl/min, at 45°C, with a 15% post column split to the mass spectrometer.

Mass spectra were acquired on a Quattro II (Micromass), triple quadrupole mass spectrometer, equipped with an electrospray atmospheric pressure ionization source. Samples were introduced from a SMART HPLC to the ionization source via a microbore PEEK capillary (75 µm i.d. × 500 mm). The electrospray capillary was held at a potential of ~2 kV. The sampling cone voltage was 30 V. Full scan spectra were acquired over a mass range of 150–2000 Da, at a rate of 5 scan/s. SIM for the 204 m/z+ ion in the positive ion mode was also acquired using a cone voltage of 100 V.

**NMR spectroscopy**

A solution of the lyophilized, dansylated, Smith-degraded glycopeptide (0.5 mg) in D₂O (0.6 ml; 99.95%; euriso-top) was used for the NMR experiments. Spectra were recorded at 300 K and 330 K non-spinning in a 5 mm tube at 300.13 MHz for 1H and at 75.47 MHz for 13C with a Bruker AVANCE 300 spectrometer equipped with a 5 mm QNP–probehead with z-gradients. 1H spectra were referenced to internal DSS (δ = 0 p.p.m.); 13C spectra were referenced to external 1,4-dioxane (δ = 67.40 p.p.m.). COSY: spectra were acquired using pulsed field gradients. The data were collected as a 256 × 1024 data matrix, which was processed using sinusoidal multiplication followed by zero filling to afford a final data matrix consisting of 512 × 512 points. TOCSY, ROESY, and HMQC spectra were recorded in phase sensitive mode using TPPI (Bax and Davis, 1985). TOCSY: the spin lock time was 80 ms. ROESY: the mixing time was 250 ms. HMQC: the spectrum resulted from a 512 × 2048 data matrix. The data matrix was zero filled and Fourier transformed to 1K × 1K data points using an exponential function (LB1 = 5, LB2 = 10) as a window function. HMBC: the spectrum was obtained from a 256 × 1024 data matrix, with 480 scans per t1 value and a pulse delay of 2.5 s.

**MALDI-TOF-MS**

Analysis was performed on a Thermo Bioanalysis mass spectrometer according to Hillenkamp et al. (1991).

**Acknowledgments**

We thank Sonja Zayni and Andrea Scheberl for excellent technical assistance. MALDI-TOF-MS of the dansylated Smith-degradation product was performed by Dr. Friedrich Altmann, Institut für Chemie, Universität für Bodenkultur Wien. We also thank Dr. Daniel Jardine, School of Chemistry, Macquarie University, Sydney for his help obtaining ESI-MS spectra. The work was supported by the Austrian Science Fund, project S7201-MOB and P12966-MOB (to P.M.) and the Austrian Federal Ministry of Science and Transportation. A.A.G. acknowledges support from
the Australian Proteome Analysis Facility and N.E.Z. is supported by an Australian Postgraduate Award.

**Abbreviations**

2-AB, 2-amino benzamide; ATZ, anilinothiazolinone; CID, collision induced dissociation; GalNAc, N-acetyl-L-galactosamine (2-acetamide-2-deoxy-2-galactose); COSY, correlated spectroscopy; DeoxyHex, deoxyhexoses; DQF-COSY, double-quantum filtered COSY; GP, glycophosphate; EDC, N-ethyl-N'3-dimethyl-aminopropylcarbodiimid; Gin, glutamine; Glu, glutamic acid; Gly, glycine; GP A, glycopeptide A; GP B, glycopeptide B; HexNAc, N-acetyl-hexosamine; HPAEC-PED, high-performance anion-exchange chromatography with pulsed electrochemical detection; HMQC, heteronuclear multiple-quantum correlation; LC-ESI-MS, liquid chromatography with pulsed electrochemical detection; HMQC, heteronuclear multiple-quantum correlation; LEC-TOCSY, total correlation spectroscopy; TPPI, time proportional phase incrementation.

**References**


