Evidence for Tyrosine-Linked Glycosaminoglycan in a Bacterial Surface Protein

Jürgen PETERS[a], Sabine RUDOLF[a], Hartmut OSCHKINAT[a], Rainer MENGELE[b], Manfred SUMPER[b], Josef KELLERMANN[a], Friedrich LOTTSPEICH[a] and Wolfgang BAUMEISTER[a]

[a] Max-Planck-Institut für Biochemie, Martinsried bei München
[b] Lehrstuhl Biochemie I der Universität Regensburg

Summary: The S-layer protein of *Acetogenium kivui* was subjected to proteolysis with different proteases and several high molecular mass glycosaminoglycan peptides containing glucose, galactosamine and an unidentified sugar-related component were separated by molecular sieve chromatography and reversed-phase HPLC and subjected to N-terminal sequence analysis. By methylation analysis glucose was found to be uniformly 1,6-linked, whereas galactosamine was exclusively 1,4-linked. Hydrazinolysis and subsequent amino-acid analysis as well as two-dimensional NMR spectroscopy were used to demonstrate that in these peptides carbohydrate was covalently linked to tyrosine. As all of the four Tyr-glycosylation sites were found to be preceded by valine, a new recognition sequence for glycosylation is suggested.

Beweise für Tyrosin-gebundenes Glycosaminoglycan in einem bakteriellen Oberflächenprotein

Zusammenfassung: Das S-Layer Protein von *Acetogenium kivui* wurde mit verschiedenen Proteasen verdaut und mehrere hochmolekulare Glycosaminoglycanpeptide, die Glucose, Galactosamin und eine noch nicht identifizierte Zuckerkomponente enthielten, wurden durch Molekularsiebchromatographie und Reversed-phase-Chromatographie getrennt und N-terminal sequenziert. Die Methylierungsanalyse ergab eine einheitliche 1,6-Verknüpfung der Glucose und eine 1,4-Verknüpfung des Galactosamins. Durch Hydrazinolyse in Verbindung mit Aminosäureanalyse sowie durch 2D-NMR-Spektroskopie wurde eine kovalente Verknüpfung von Tyrosin und Zucker in diesen Peptiden gezeigt. Da allen vier Glycosylierungsstellen die Aminosäure Valin vorangeht, wird eine neue Erkennungssequenz für die Glycosylierung vorgeschlagen.

Key terms: Glycosaminoglycan; tyrosine-bound sugar; NMR-spectroscopy; *Acetogenium kivui*

Glycosylation has been recognized as an important feature of many eukaryotic proteins. The notion that glycosylation is also widespread in prokaryotes is relatively recent (see ref.[1] for a review). In prokaryotic proteins glycosylation sites have only been localized in the flagellins and the cell surface glycoprotein of *Halobacterium* which belongs to the archaebacteriae[2]; evidence for the existence of true glycoproteins among the bacteria is not beyond doubt[1]. We have recently reported the nucleotide sequence-derived amino-acid sequence of this S-layer protein which appeared to be extensively and heterogeneously glycosylated[3]. We have used protein chemical as well as NMR methods to show that certain tyrosine residues of the surface (S)-layer protein of the thermophilic bacterium *Acetogenium kivui* are glycosylated.

Materials and Methods

Isolation of the S-layer, cleavage and chromatographic procedures

The S-layer of *A. kivui* was isolated as described[3]. For cleavage with proteinase K (EC 3.4.21.14; Boehringer, W-6800 Mannheim, Germany) under native conditions 300 mg of protein suspended in 13.5 ml 40mM Tris/Cl, pH 7.5, were treated with 3 mg proteinase K in 3 aliquots at 50 °C for 10 h. The digestion was terminated by the addition of 100 μl of 100mM phenylmethane sulfonylfluoride in 2-propanol. After centrifugation at 13600 × g for 5 min in an Eppendorf centrifuge sedimented residual cell wall material was discarded and the supernatant applied to a 600 × 21.5 mm TSK 3000 column equilibrated with 200mM ammonium acetate, pH 5.9, at a flow rate of 3 ml/min. Polypeptides were precipitated with 75% saturated ammonium sulfate and the sediment dissolved in 3 ml of water.

After the addition of 9 ml of concentrated formic acid, 2 mg pepsin (Sigma, W-8024 Deisenhofen, Germany) dissolved in 120 ml 10mM HCl were added with stirring and the sample incubated at room temperature for 14 h. The material was lyophylized, dissolved in 200mM ammonium acetate, pH 5.9 and chromatographed in the same solvent on a preparative TSK 3000 column at 4 ml/min.

Reversed-phase chromatography (RPC) was performed on a Li-Chrospher 100 RP 18 column, 5 μm, 250 × 4 mm (Merck, W-6100 Darmstadt, Germany). Peptides were eluted with a gradient of 0–30% acetonitrile, 0.1–0.094% trifluoroacetic acid.

Cleavage with endoproteinase AspN (Boehringer, Mannheim) was carried out in 50mM Tris/Cl, pH 7.5. The enzyme was added to the glycopeptides at 0.5% by weight and samples were incubated at room temperature for 48 h. For subsequent chromatography concentrated formic acid was added to 30% by volume.

Cleavage with pronase (Sigma) was performed in 50mM Tris/Cl, pH 7.5, 1mM CaCl$_2$ at 37 °C for 24 h and 50 °C for 6 h with 2% pronase by weight. The proteinase was added in five aliquots throughout the period of incubation.

Hydrazinolysis was performed as described[4].

Amino-acid and amino-acid sequence analysis

These procedures were performed as described previously[5]. The *o*-phthaldialdehyde method was used for amino-acid analysis.

Carbohydrate analysis procedures

Neutral sugars were routinely detected by the phenol/sulfuric acid method[6]. Glucose was identified as alditol acetate[7] as described in ref.[5] after hydrolysis with 4M trifluoroacetic acid for 4 h at 100°C. Total sugar analysis was performed according to ref.[8]. Amino sugars were determined with ninhydrin on a Beckman model 6300 automated amino-acid analyzer after hydrolysis with 4M HCl for 4 h at 100 °C. Under these conditions deacylation was possible.

For linkage analysis the saccharides were permethylated by the procedure of Hakamori[9], as modified by Waeghe et al.[10]. The permethylated saccharides were purified with Sep-pak C$_{18}$ cartridges (Waters Inc.) by elution with 40% acetonitrile. After hydrolysis in 4M HCl for 4 h at 100 °C the methylated sugars were reduced with NaBH$_4$, peracetylated and analysed by GC-mass spectrometry as described recently[11]. Reference spectra for methylated alditol acetates were taken from ref.[12] for neutral sugars and from ref.[13] for amino sugars.

NMR measurements

About 1 mg of the Asp-N-generated glycosaminoglycanpeptide fraction (see above) was dissolved in 600 μl D$_2$O. The measurements were performed on Bruker Spectrometers (AM 500 and AMX 600) at 300 °K. A suitable Lorentz-to-Gauss conversion of the line shapes was applied prior to Fourier transformation to all

spectra. All chemical shifts given are relative to HDO = 4.76 ppm. The 1D-^{1}H-NMR-spectrum shown in Fig. 3A was recorded at 600 MHz with 96 scans, a spectral width of 10414 Hz, and sampling 4 K data points. The FID was extended to 16 K data points by zero filling. The baseline of the spectrum shown was corrected interactively. A COSY spectrum[14] with double quantum filter using a standard pulse sequence was recorded at 600 MHz with 64 scans, 512 experiments in t_1, 2 K data points int t_2, and a spectral width of 10414 Hz in both dimensions. The relaxation delay was 1.8 s. A NOESY spectrum[15] employing a mixing time of 150 ms and a relaxation delay of 1.5 s was recorded at 500 MHz. Another NOESY spectrum with the same parameters and a mixing time of 70 ms was recorded at 600 MHz. The spectral width of this spectrum was 5000 Hz, 700 experiments were performed in t_1, 2 K data points were taken in t_2, 128 scans were applied, as well as 4 dummy scans at the beginning of each experiment.
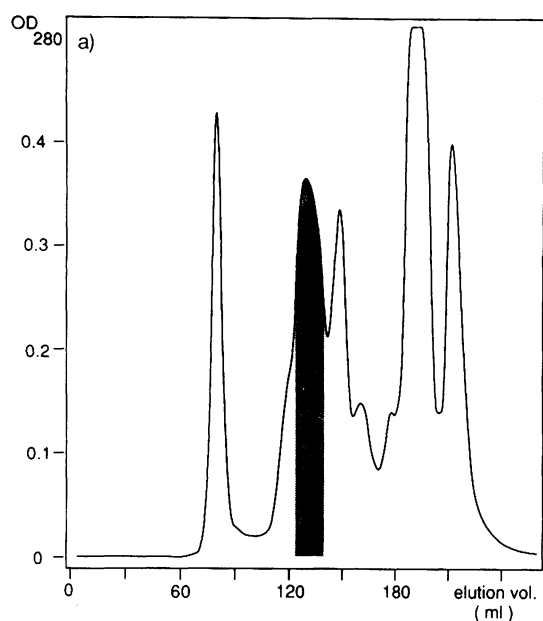
Results and Discussion

For the isolation of glycosaminoglycanpeptides from the S-layer protein of *Acetogenium kivui* the native protein was initially cleaved with proteinase K at 50 °C and subjected to molecular sieve chromatography (Fig. 1a). The shaded fraction in Fig. 1a contained glucose as the only neutral sugar and galactosamine as the only amino sugar. This material was cleaved with pepsin and subjected to molecular sieving. A peak containing the glycosaminoglycanpeptides was found to be well-separated from small peptic peptides by virtue of its relatively high app. M_r (Fig. 1b). When chromatographed on a TSK 2000 molecular sieving column under denaturing conditions (data not shown) the app. M_r of this fraction was found to be approx. 5500 for the main peak and 7500 for the shoulder. When the app. M_r 5500-glycosaminoglycan peptide fraction was separated by RPC (Fig. 1c), 7 peaks containing three different peptides or fragments derived from them were obtained as judged by complete N-terminal amino-acid sequence analysis (Fig. 2) which were congruent with positions 268–285,

$\triangleright$

Fig. 1. a) Molecular sieve chromatography of *Acetogenium kivui* S-layer protein digested with proteinase K under native conditions and separated on a preparative TSK 3000 column (see Materials and Methods). The void volume was 80 ml, the column volume 195 ml. The fraction containing glycosaminoglycan is shaded.

b) Separation of peptic peptides of the shaded fraction in a) under the same conditions as in a). The glycosaminoglycan fraction is again shaded.

c) Reversed-phase chromatography of the peptic glycosaminoglycanpeptides corresponding to the shaded fraction in b) excluding the shoulder. % B, % acetonitrile. Fractions 1 through 3, peptide TSTVYA and derived fragments (see Fig. 2). Fractions 4 through 6, peptide TTVYKNDDNK-TAIKVDDNA and derived fragments. Fraction 7, peptide AVYSDVYDKVNL.

| Glycopeptides	Cleavage
AVYSDVYDKVNL	**Pepsin**
AVYS	Protease Asp-N
DVY	
VY VY	Hydrazinolysis
TTVYKNDDNKTAIKVDDNA	**Pepsin**
TTVYKN	Protease Asp-N
VY	Hydrazinolysis
TSTVYA	**Pepsin**
TSTVYA	Protease Asp-N
VY	Hydrazinolysis

Fig. 2. Glycosaminoglycan-containing peptides of the S-layer.

Only peptides containing one single glycan chain are shown. Note that the fraction of peptic peptide AVYSDVYDKVNL that was glycosylated at both Tyr residues was not cleaved by AspN but by pronase (see text).
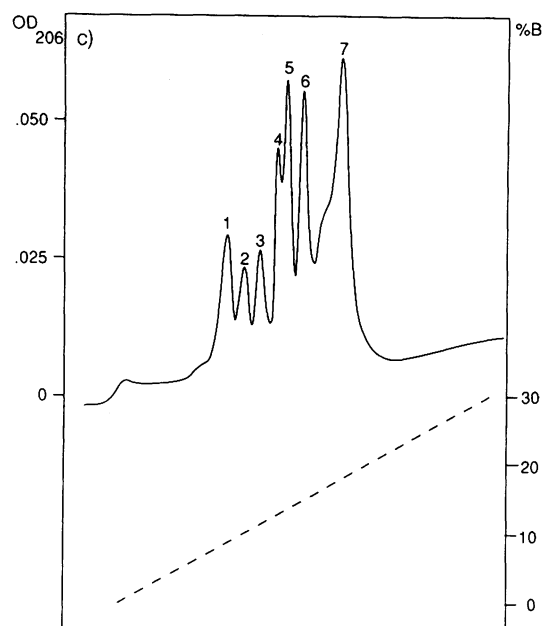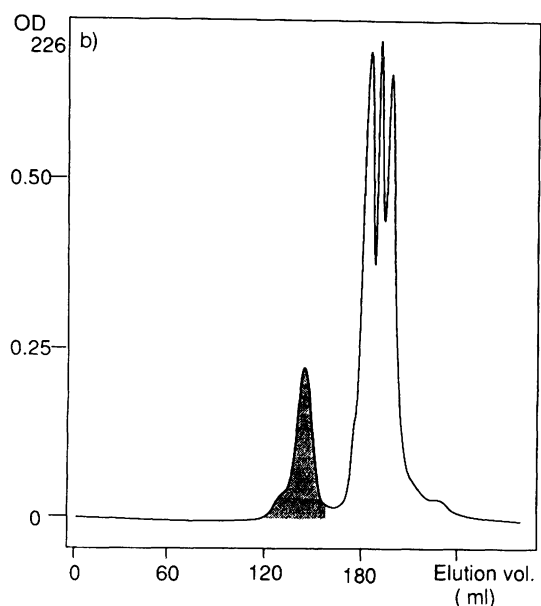
488–499, and 602–607 of the nucleotide sequence-derived amino-acid sequence (see ref.[3]). However, there were no phenylthiohydantoin derivatives missing in the Edman degradation cycle indicating glycosylated Ser/Thr or Asn residues. We concluded that the glycan portions were removed under the conditions of automated Edman degradation. In fact, exposure to anhydrous trifluoroacetic acid for 30 min at 50 °C resulted in the complete removal of the carbohydrate from the peptide portions. In contrast to this finding the peptidyl-glycosyl bond turned out to be resistant to treatment with 1M NaOH at 80 °C for 4 h.

Unfortunately, the success of the separation of peptic peptides was critically dependent on the amount of peptide material applied to the reversed-phase column: Only about 500 pmol of each peptide per run could be obtained in pure form, as judged by N-terminal sequence analysis.

Subcleavage of the total peptic peptide fraction with endoproteinase AspN and the isolation of a glycosaminoglycanpeptide fraction by molecular sieving on a TSK 2000 column resulted in a shift of app. M_r to 3800 for the main peak and a smaller peak at 7500 corresponding to the shoulder in Fig. 1b (not shown).

The app. M_r-7500 peak was shown to contain the peptide AVYSDVYDKVNL which was apparently not cleaved by endoproteinase AspN. However, RPC of the main peak (not shown) and sequence analysis revealed that both AVYS and DVY (see Fig. 2) were present in the glycanpeptide fraction of app. M_r 3800,

indicating that the parent peptide was only cleaved when glycosylated at one single site. The app. M_r value of 7500 of the peak containing the peptide AV-YSDVYDKVNL is also in agreement with the assumption that a fraction of this peptide is double-glycosylated with glycosaminoglycan chains, whereas the value of 3800 corresponds to single glycosylation of either AVYS or DVY. The app. M_r-7500 glycanpeptide was cleaved by pronase and the peptide AVYSD was isolated by RPC (not shown).

Hydrazinolysis was used to truncate the peptide moieties as far as possible to localize the glycan-linking amino acid. Molecular-sieve chromatography and amino-acid analysis of the glycosaminoglycan peptides revealed that about equimolar amounts of valine and tyrosine were essentially the only amino acids that remained bound to the carbohydrate portion both with the peptide AVYSD and the complete AspN-generated peptide fraction used as starting materials (not shown). From the sequencing data it follows that the corresponding dipeptide is uniformly Val-Tyr. It is not clear why the Val-Tyr bond was resistant to hydrazinolysis.

To prove unequivocally that the binding of tyrosine residues to carbohydrate is indeed covalent we used ^{1}H-NMR spectroscopy. Since the individual glycosaminoglycanpeptides could only be purified in amounts of a few μg per run, the entire Asp-N-peptide fraction (ca. 1 mg) was used. N-terminal amino-acid sequence analysis of this fraction confirmed that only the four Asp-N-generated glycopeptides listed in Fig. 2 were present. Furthermore, the linkage and composition data of the carbohydrate portions presented below and the uniformity of the glycosyl-peptide linkage as demonstrated below justify this approach.

The 1D-^{1}H-NMR spectrum of the glycanpeptide fraction is shown in Fig. 3a. Assignments were made according to chemical shift and COSY data. The signal pattern is that of a polysaccharide with repetitive units in the region between 3.4 and 5.2 ppm and small signals in the region between 0.8 and 3.3 ppm which are due to side chains of the amino acids. There are also some signals at 4.2/4.45 ppm and 7.0/7.2 ppm which are assigned to the α-protons and the aromatic ring protons of the amino acids, respectively. The heterogeneity of the peptide portions lead to a slight variation in the chemical shifts of the peptide signals as can be seen in the aromatic region of the spectrum. The large signals at 5.2 and 4.9 ppm are due to the C_1'-H of two α-anomeric sugars, the signal at 4.6 ppm showing a larger splitting is caused by the C_1'-H of an oligosaccharide in its β-form. There is a number of large signals originating from methyl groups around

1.95/2.05 ppm which may be tentatively assigned to the signals of an acylated sugar unit (probably Gal-NAc) due to cross peaks to the C_1'-proton at 5.2 ppm. The small signals to the left of the C_1'-signals of the α-anomeric sugars are probably due to the terminal repetitive unit, the C_1'-signal of the sugar bound to tyrosine occurs as a shoulder on the right hand side of the signal at 5.2 ppm.

There are in principle two different ways of demonstrating the connectivity of the aromatic ring and the sugar chain. The most definite proof would be the detection of small long-range couplings between carbon and hydrogen atoms in both units. Unfortunately, these couplings across the oxygen link are expected to be small (< 4 Hz), and it is unlikely that they can be resolved in biological samples with a considerable linewidth. On the other hand, such samples show a strong negative NOE which may be utilized to detect a close distance between the C_ϵ-H of the tyrosine ring and the C_1'-H of the linking unit of the sugar chain. In this case, however, it is necessary to rule out the possibility of long-range interactions which may be caused by conformational effects or micelle formation. If either of these two effects would be responsible for the detected NOE, a number of NOEs involving the protons at the tyrosine ring should be found and the observed NOE should be small.

The part of the NOESY spectrum which provides evidence for the tyrosyl-glycosyl linkage comprises the region with the chemical shifts of the Tyr-C_β-H (2.9/3.1 ppm) and the C_1'-H of the sugars in F_1, and that of the aromatic protons of the tyrosine in F_2 (Fig. 3b). The C_δ'-H signals of tyrosine can be identified by the cross peaks involving the C_β-H (upper part of the spectrum in Fig. 3b) whereas the covalent linkage is indicated by a relatively strong cross peak between the C_ϵ-H and the C_1'-H at 5.2 ppm. The relative sizes of the cross peaks give rise to some surprise, because the ones involving C_δ-H and C_β-H of the tyrosine are quite small compared to the one between sugar and tyrosine. There are a number of reasons which affect the amplitudes of the relevant cross peaks. Firstly, the peptide portion is heterogeneous, hence the intensity is smeared out over a certain chemical shift range: The peaks become very broad. Secondly, the mobility increases strongly towards the end of the peptide, so that the effective correlation times, and hence the negative NOE, become small. In fact, nearly no NOE is observed between the side chain protons of different amino acids. On the other hand, the latter gives good evidence for the interpretation of the NOE in the lower part of the spectrum as an intramolecular NOE, because the high mobility excludes the pres-
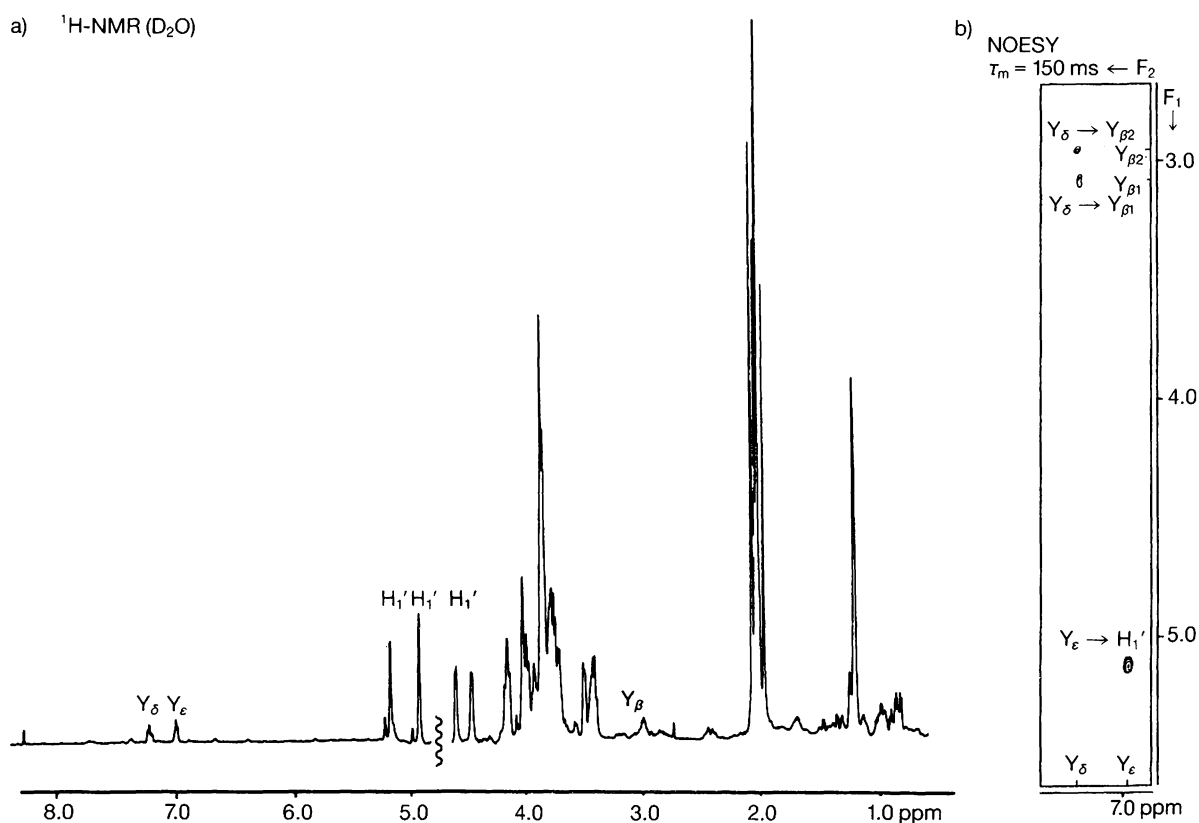
Fig. 3. a) ^{1}H-NMR spectrum of the AspN-generated glycosaminoglycanpeptide fraction.

The resonances of the tyrosine (Y) protons and the anomeric sugar protons are indicated. b) Region of the NOESY spectrum ($\tau =$ 150 ms, $T = 300$ °K) showing cross-peaks to the aromatic resonances of the tyrosine residues. The relatively weak signals involving the β-protons enable the distinction between C_δ and C_ε. The strong signal in the lower part establishes the peptidyl-glycosyl link. No other signals involving protons in the peptide and the glycan were observed.

ence of a rigid tertiary structure or the presence of micelles. To confirm this, a NOESY spectrum with a shorter mixing time was recorded (70 ms), showing only very weak cross peaks within the tyrosine system and still a cross peak of substantial size between tyrosine and sugar. The relative size of the latter peak compared to the ones within the sugar system at 70 ms mixing time suggests a distance of less than 3.0 Å between the two protons. Considering the fact that motion only decreases the amplitudes, an intermolecular NOE is ruled out. Also, there are no cross peaks to other sugar protons, thus suggesting an extended arrangement of sugar and peptide portions. Moreover, no sizeable cross peaks between other aminoacid side chains and sugar units were detected. In conclusion, the NMR data provide strong evidence for the covalent attachment of the carbohydrate moieties to the aromatic ring of tyrosine.

Only one glycosaminoglycanpeptide, AVYSD (see above) could be purified by RPC in an amount sufficient for linkage analysis of permethylated sac-charides. The entire AspN-generated peptide fraction as used for the NMR studies was also subjected to linkage analysis. In both cases glucose was found to be uniformly 1,6-linked whereas galactosamine was exclusively 1,4-linked. Derivatisation of methanolysed saccharides with pentafluoropropionic acid anhydride and subsequent GC/mass spectrometric analysis established that glucose, galactosamine and an as yet unidentified sugar-related component were present in a ratio of approximately 1:1.1:1 in both cases. These data are in good agreement with the results of the NMR studies which show the presence of two α-anomeric and one β-anomeric sugars.

The data presented show clearly that glycosaminoglycan is linked to tyrosine in the S-layer protein of A. kivui. Our data provide solid evidence for the existence of glycoproteins in the bacterial kingdom. The linkage of sugar to tyrosine has to our knowledge previously only been demonstrated in glycogenin from rabbit skeletal muscle[16].

Out of 45 tyrosine residues in the amino-acid sequence of the S-layer protein, 8 residues are preceded by valine. The statistical probability of four glycosylated Tyr residues being preceded by valine is therefore 0.00047, as calculated according to ref.[17]. It is thus likely that Val-Tyr is a sequon specifying glycosylation of tyrosine in *A. kivui*.

References

1   Lechner, J. & Wieland, F. (1989) *Annu. Rev. Biochem.* **58**, 173–194.
2   Sumper, M. (1987) *Biochim. Biophys. Acta* **906**, 69–79.
3   Peters, J., Peters, M., Lottspeich, F. & Baumeister, W. (1989) *J. Bacteriol.* **171**, 6307–6315.
4   Fincher, G.B., Sawyer, W.H. & Stone, B.A. (1974) *Biochem. J.* **139**, 535–545.
5   Peters, J., Peters, M., Lottspeich, F., Schäfer, W. & Baumeister, W. (1987) *J. Bacteriol.* **169**, 5216–5223.
6   Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A. & Smith, F. (1956) *Anal. Chem.* **28**, 350–356.
7   Henry, R.J., Blakeney, A.B., Harris, P.J. & Stone, B.A. (1983) *J. Chromatogr.* **256**, 419–427.
8   Lechner, J., Wieland, F. & Sumper, M. (1985) *J. Biol. Chem.* **260**, 860–866.
9   Hakamori, S. (1964) *J. Biochem. (Tokyo)* **55**, 205–207.
10   Waeghe, T.J., Darvill, A.G., McNeil, M. & Albersheim, P. (1983) *Carbohydr. Res.* **123**, 281–304.
11   Sumper, M., Berg, E., Mengele, R. & Strobel, I. (1990) *J. Bacteriol.* **172**, 7111–7118.
12   Jannson, E., Kenne, M., Lindgren, M., Lindberg, B. & Loenngren, J. (1976) *Chem. Commun.* **8**, 1–74.
13   Stellner, K., Saito, H. & Hakamori, S.-I. (1973) *Arch. Biochem. Biophys.* **155**, 464–472.
14   Aue, W.P., Bartholdi, E. & Ernst, R.R. (1976) *J. Chem. Phys.* **64**, 2246–2299.
15   Jeener, J., Meier, B.H., Bachmann, P. & Ernst, R.R. (1979) *J. Chem. Phys.* **71**, 4546–4553.
16   Smythe, C., Caudwell, F.B., Ferguson, M. & Cohen, P. (1988) *EMBO J.* **7**, 2681–2686.
17   Weber, E. (1972) in *Grundriß der biologischen Statistik*, pp. 516–519, Gustav-Fischer-Verlag, Stuttgart.

J. Peters*, H. Oschkinat, J. Kellermann, F. Lottspeich and W. Baumeister, Max-Planck-Institut für Biochemie, Am Klopferspitz, W-8033 Martinsried bei München, Germany;

R. Mengele and M. Sumper, Lehrstuhl Biochemie I der Universität Regensburg, Universitätsstr. 31, W-8400 Regensburg, Germany.

* To whom correspondence should be addressed.