Eur. J. Biochem. *196*, 115–122 (1991)

0014295691001 20V

Primary structure of O-linked carbohydrate chains in the cellulosome of different *Clostridium thermocellum* strains

Gerrit J. GERWIG[1], Johannis P. KAMERLING[1], Johannes F. G. VLIEGENTHART[1], Ely MORAG (MORGENSTERN)[2], Raphael LAMED[2] and Edward A. BAYER[3]

[1] Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, Utrecht, The Netherlands
[2] Department of Biotechnology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel
[3] Department of Biophysics, The Weizmann Institute of Science, Rehovot, Israel

(Received September 11, 1990) – EJB 90 1086

The cell-free forms of the multiple cellulase-containing protein complex (cellulosome), isolated from the cellulolytic bacterium *Clostridium thermocellum* strains YS, ATCC 27405 and LQRI, have a total carbohydrate content of 5–7% (by mass), consisting of O-linked oligosaccharide chains. The carbohydrate chains were liberated by alkaline-borohydride treatment and fractionated as oligosaccharide alditols via gel-permeation chromatography and HPLC. The fractions were investigated by 500-MHz ^{1}H-NMR spectroscopy in combination with monosaccharide and methylation analysis and with fast-atom-bombardment mass spectrometry (FAB-MS). In addition to the previously described major oligosaccharide,

D-Gal*p*-α(1 → 3)
\
D-Gal*f*-α(1 → 2)-D-GalOH
/
3-*O*Me-D-Glc*p*NAc-α(1 → 2)

[Gerwig, G. J., de Waard, P., Kamerling, J. P., Vliegenthart, J. F. G., Morgenstern, E., Lamed, R. & Bayer, E. A. (1989) *J. Biol. Chem. 264*, 1027–1035], the following partial structures of this compound could be established:

D-Gal*p*-α(1 → 3)
\
D-Gal*f*-α(1 → 2)-D-GalOH,
/
D-Glc*p*NAc-α(1 → 2)

3-*O*Me-D-Glc*p*NAc-α(1 → 2)-D-Gal*f*-α(1 → 2)-D-GalOH,

D-Glc*p*NAc-α(1 → 2)-D-Gal*f*-α(1 → 2)-D-GalOH,

D-Gal*f*-α(1 → 2)-D-GalOH and GalOH.

Cell-free and cell-associated forms of the cellulosome of *C. thermocellum*, as determined for strain YS, have the same oligosaccharide pattern. Based on the oligosaccharide structures, a biosynthetic pathway is suggested.

C. thermocellum, an anaerobic thermophilic bacterium, is capable of saccharifying crystalline cellulose. To this end the cell produces a high-molecular mass, cellulose-binding, multicellulase-containing protein complex, termed the cellulosome, which displays *endo*-glucanase and glucohydrolase activity [1, 2]. The cellulosome occurs in both extracellular and cell-surface forms and mediates strong adherence of the bacterium to the insoluble substrate [3]. The respective polypeptide compositions of the cell-free and cell-bound forms are quite similar [2].

In order to understand the pathway that leads to the degradation of cellulose by *C. thermocellum* a more extensive knowledge of the properties and character of its cellulase system (cellulosome) is essential. The glycoprotein nature of some of the cellulosomal subunits [2] prompted an investigation of the structures of the carbohydrate chains. The largest protein component (210 kDa) of the cellulosome, a highly glycosylated (40%, by mass) S1 subunit, exhibits no measurable cellulolytic activity, but is highly antigenic and is probably involved in the structural organisation of the complex [4].

To have the structural data of the cellulase complex available is beneficial for the study of the basic mechanism of cellulase action in *C. thermocellum*. In previous work [5], we have shown the occurrence, in the cellulosome, of two major

Correspondence to J. F. G. Vliegenthart, Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, P.O. Box 80.075, NL-3508 TB Utrecht, The Netherlands

Abbreviations. DQF^{1}H-^{1}H COSY, double-quantum-filtered ^{1}H-^{1}H correlation spectroscopy; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; 3-*O*Me-Glc*p*NAc, 2-acetamido-2-deoxy-3-*O*-methyl-D-glucopyranose; GalOH, galactitol; FAB-MS; fast-atom-bombardment mass spectrometry.

Enzyme. Peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F (EC 3.5.1.52).

O-linked oligosaccharide chains of novel type, identified as their alditols, namely, D-Gal*p*-β(1 → 4)-D-GalOH and

D-Gal*p*-α(1 → 3)
\
D-Gal*f*-α(1 → 2)-D-GalOH .
/
3-*O*Me-D-Glc*p*NAc-α(1 → 2)

SDS/PAGE combined with blotting techniques indicated that the tetrasaccharide is associated with the S1 subunit [5].

In this paper, we describe the structural determination of additional O-linked carbohydrate chains in the cellulosome, which form partial elements of the tetrasaccharide. Furthermore, different strains of *C. thermocellum* are compared with regard to the structures of protein-linked carbohydrates.

MATERIALS AND METHODS

Isolation of the cellulosome

The *C. thermocellum* strains used in this study were YS, ATCC 27405 (NCIB 10682) and LQRI. Cell-associated and cell-free forms of the cellulosome were isolated from whole-cell sonic extracts or from the corresponding cell-free supernatant fluid of cellulose-grown cells, respectively, as described earlier [1, 4]. In brief, the work up involved the batch-wise adsorption of cellulosomes onto cellulose followed by elution with 1% (by vol.) aqueous triethylamine. The material was further fractionated by gel filtration. Before structural analysis, cellulosomal samples were dialyzed against distilled water and then lyophilized [5].

Monosaccharide analysis

Samples (25 μg) were subjected to methanolysis (1.0 M methanolic HCl, 24 h, 85 °C) followed by GLC of the trimethylsilylated (*N*-re-acetylated) methyl glycosides on a capillary WCOT SE-30 fused-silica column (25 m × 0.32 mm, Pierce) [6]. The absolute configuration of the monosaccharides was determined by GLC of the trimethylsilylated (*N*-re-acetylated) (−)-2-butyl glycosides [7, 8].

Methylation analysis

Methylation analysis on 50-μg samples was performed with methylsulfinylcarbanion and methyl iodide in dimethylsulfoxide as described [5].

Alkaline-borohydride treatment

Cellulosome samples (10 mg) were treated with 0.1 M NaOH containing 1 M $NaBH_4$ (48 h, 37 °C) and the oligosaccharide alditols formed were fractionated on a column (65 × 1.8 cm) of Bio-Gel P-2 as reported [5]. A separation of pooled regions from the first P-2 gel filtration was performed on a second column (95 × 1 cm) of Bio-Gel P-2 (200–400 mesh, Bio-Rad) using water as eluent (6 ml/h). For two additional cell-free cellulosome samples of strain YS, different β-elimination conditions were tested, (A) 0.2 M NaOH containing 1 M $NaBH_4$ (16 h, 90 °C) and (B) 0.1 M NaOH without $NaBH_4$ (24 h, 37 °C).

High-performance liquid chromatography

HPLC on a Lichrosorb-10-NH_2 column (250 × 4.6 mm, Chrompack) was carried out as described [5] using a mixture of acetonitrile/water, 80:20 (by vol.) as eluent and a flow rate of 1.0 ml/min.

Fast-atom-bombardment mass spectrometry

Positive-ion mass spectra were recorded as reported previously [5]. The carbohydrate samples (50 μg) were dispersed in glycerol and bombarded with xenon atoms.

500-MHz ^{1}H-NMR spectroscopy

Carbohydrate samples (25–100 μg) were repeatedly exchanged in ^2H_2O (99.96 atom % ^{2}H, Aldrich) with intermediate lyophilization. Resolution-enhanced ^{1}H-NMR spectra were recorded on a Bruker AM-500 spectrometer (Bijvoet Center, Department of NMR Spectroscopy, Utrecht University) operating at 500 MHz with a probe temperature of 300 K. Chemical shifts (δ) are expressed in ppm relative to sodium 4,4-dimethyl-4-silapentane-1-sulfonate, but were actually measured by reference to internal acetone in ^2H_2O ($\delta = 2.225$). Two dimensional ^{1}H-NMR experiments, double-quantum-filtered ^{1}H-^{1}H correlation spectroscopy (DQF ^{1}H-^{1}H COSY) and homonuclear Hartmann-Hahn spectroscopy (HOHAHA) were performed as described [5].

RESULTS

Alkaline-borohydride treatment of cell-free cellulosomes of strain YS and subsequent fractionation on Bio-Gel P-2 gave rise to two major and several minor carbohydrate-containing fractions (Fig. 1 A). As reported [5], the major low-molecular-

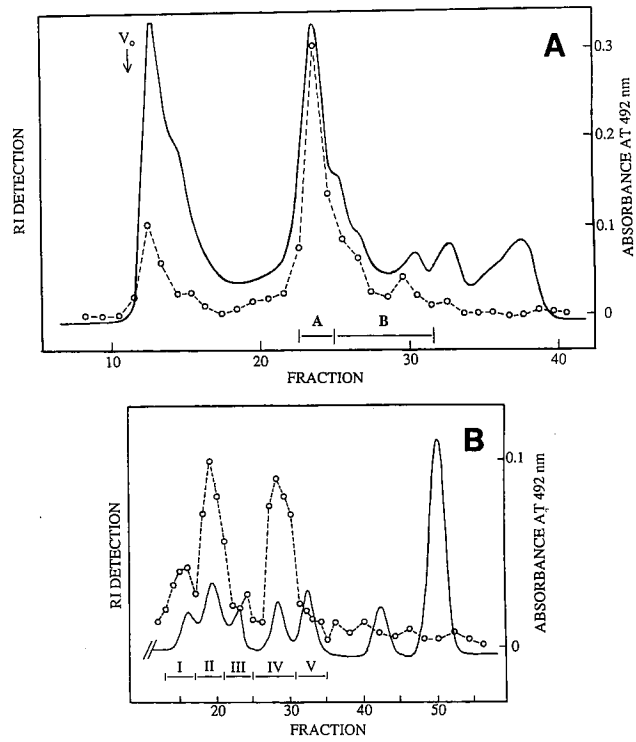


Fig. 1. *Elution profile on Bio-Gel P-2 of alkaline-borohydride-treated (0.1 M NaOH, 1 M $NaBH_4$, 48 h, 37 °C) cell-free cellulosome of* C. thermocellum *YS.* (A) Trace from [5]. (B) The P-2 column (95 × 1 cm) was eluted with bi-distilled water. Fractions of 1 ml were collected at a flow rate of 6 ml/h. The eluate was monitored by refractive index detection (——) and by hexose determination with phenol/sulfuric acid (○-----○). The fractions indicated were pooled. V_o, void volume

Table 1. *Molar carbohydrate composition of derived fractions of alkaline-borohydride-treated cell-free cellulosome of* C. thermocellum *YS*
All sugars possess the D-configuration. For the molar ratio GalOH was taken as 1.0 + mean <0.1. For quantification of 3-*O*Me-GlcNAc, the molar adjustment factor of GlcNAc was used

Monosaccharide	Molar ratio of fraction										
	I	II	III	IV	V	AI	AII	AIII	AI.1	AI.2	AI.3
Gal	1.9	2.0	1.2	0.9	–	1.8	1.5	–	1.0	0.9	2.1
GalOH	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
3-*O*Me-GlcNAc	0.8	–	0.4	–	–	0.7	–	–	0.6	1.0	0.9
GlcNAc	–	0.8	0.4	+	–	0.2	0.1	–	–	–	0.2

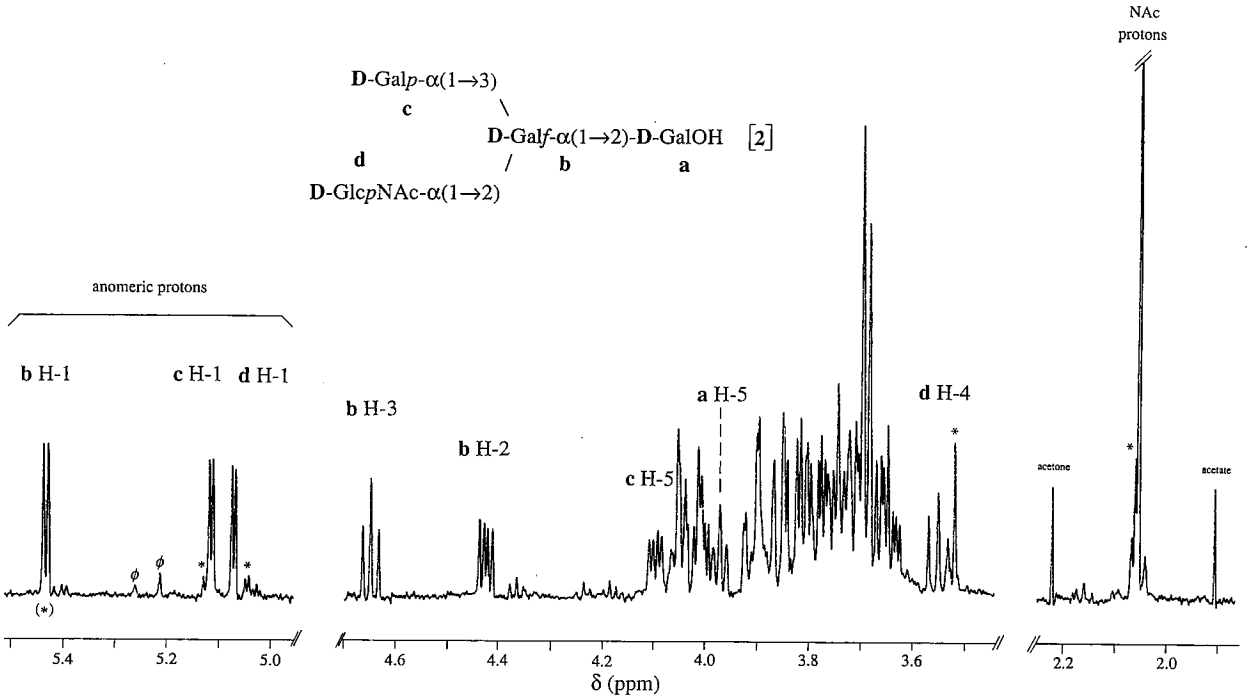


Fig. 2. *500-MHz ^{1}H-NMR spectrum of fraction II derived from* C. thermocellum *YS cell-free cellulosome recorded in ^2H_2O at 300 K.* The assignment of the protons refer to the corresponding residues in the structure. The HO^{2}H signal (4.70–4.90 ppm) has been omitted. (*) signals stemming from **1**; (Φ) non-carbohydrate contamination

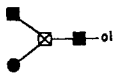
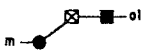
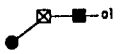
mass fraction (A in Fig. 1A) revealed a novel O-linked oligosaccharide chain, isolated as its alditol:

D-Gal*p*-α(1→3)
\
D-Gal*f*-α(1→2)-D-GalOH . (**1**)
/
3-*O*Me-D-Glc*p*NAc-α(1→2)

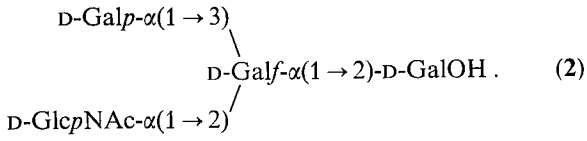
Region B (Fig. 1A) was pooled from three identical samples of alkaline-borohydride-treated cellulosomes and separated on a second Bio-Gel P-2 column (95 × 1 cm). Five carbohydrate-containing fractions, denoted I–V were isolated (Fig. 1B) and their monosaccharide analyses data are given in Table 1. All fractions yielded trace amounts of glucose, which was derived from a contaminant in each case. Fraction I contained the previously reported compound **1**, stemming from tailing of the major Bio-Gel P-2 peak in the original elution profile. Fraction V contained only free galactitol (compound **6**).

Monosaccharide analysis of fraction II gave GalOH, Gal and GlcNAc in a molar ratio of 1:2:1 (Table 1), suggesting a tetrasaccharide alditol. Positive FAB-MS of the underivatized fraction II (spectrum not depicted) showed the presence of a protonated molecular ion $(M+H)^+$ at *m/z* 710 and a cationized molecular ion $(M+Na)^+$ at *m/z* 732, indicating a molecular mass of 709 Da, in accordance with the assumed tetrasaccharide-alditol structure. The ^{1}H-NMR spectrum of fraction II is presented in Fig. 2. Three anomeric signals of equal intensity point to the presence of a single compound. Comparison of the ^{1}H-NMR data with those of **1** (see Table 2) showed that this compound is a non-methylated variant of **1**. The H-1 signals at $\delta = 5.437$ ($J_{1,2} = 4.6$ Hz), $\delta = 5.117$ ($J_{1,2} = 3.6$ Hz) and $\delta = 5.074$ ($J_{1,2} = 3.5$ Hz) belong to Gal*f*, Gal*p* and Glc*p*NAc, respectively, all having the α configuration. Owing to the absence of the OCH_3 group, compared to **1**, shift effects were observed for the NAc ($\delta = 2.057$; $\Delta\delta = -0.006$ ppm) and H-1 ($\delta = 5.074$; $\Delta\delta = +0.026$ ppm) signals of Glc*p*NAc. In addition, a large downfield shift was observed for H-3 ($\delta = 3.738$; $\Delta\delta = +0.221$ ppm), as compared to H-3 of the methoxylated carbon in **1** ($\delta = 3.517$ [5]). Two dimensional DQF ^{1}H-^{1}H COSY and HOHAHA (not depicted) were used to assign relevant

Table 2. Relevant ^{1}*H-NMR parameters of structural reporter groups of constituent monosaccharides for oligosaccharide alditols from* C. thermocellum *cellulosome*
In the table heading, the structures are represented by: (■) Gal*p*; (⊠) Gal*f*; (●) Glc*p*NAc; (■-ol) Galactitol; (m) 3-*O*Me; n.d., not determined. Some coupling constants, *J*, are given in Hz in brackets

Residue	Reporter group	Chemical shift				
		1	**2**	**4**	**5**	**3**
						
		ppm				
Gal*f*	H-1 ($J_{1,2}$)	5.438 (4.6)	5.437 (4.6)	5.401 (4.8)	5.400 (4.8)	5.265 (4.8)
	H-2 ($J_{2,3}$)	4.429 (8.2)	4.426 (8.1)	4.218 (8.7)	4.220 (8.8)	4.136 (8.6)
	H-3 ($J_{3,4}$)	4.666 (7.2)	4.651 (7.7)	4.363 (8.1)	4.350 (8.5)	4.199 (8.0)
	H-4	4.048	4.045	n.d.	n.d.	3.776
Gal*p*	H-1 ($J_{1,2}$)	5.128 (3.6)	5.117 (3.6)	–	–	–
	H-2	3.837	3.830	–	–	–
	H-3	3.803	3.796	–	–	–
	H-4	4.024	4.018	–	–	–
Glc*p*NAc	H-1 ($J_{1,2}$)	–	5.074 (3.5)	–	5.026 (3.6)	–
	H-2	–	4.007	–	n.d.	–
	H-3	–	3.738	–	n.d.	–
	H-4	–	3.550	–	3.547	–
	NAc	–	2.057	–	2.069	–
3-*O*Me-Glc*p*NAc	H-1 ($J_{1,2}$)	5.048 (3.5)	–	5.001 (3.6)	–	–
	NAc	2.063	–	2.073	–	–
	OCH_3	3.520	–	3.528	–	–
GalOH	H-1 a,b	~3.68	~3.68	~3.79	~3.79	~3.79
	H-2	4.060	4.051	4.054	4.051	4.076
	H-5	3.976	3.973	3.980	3.975	3.973

protons of the monosaccharide residues (see Table 2). In summary, the structure can be given as follows:



(2)

Monosaccharide analysis of fraction IV yielded Gal and GalOH in a molar ratio of 1:1 (Table 1), indicating a disaccharide alditol. Gas-liquid chromatography/mass spectrometry of permethylated fraction IV, proved it to be Gal*f*-(1 → 2)-GalOH (Fig. 3). The furanose form of Gal could be deduced from the ratio of the intensities of the peaks at *m/z* 88 ($MeO\dot{C}H-CH=\overset{+}{O}Me$) and *m/z* 101 ($MeOCH=CH-\dot{C}H=\overset{+}{O}Me$), being ≪1 [6, 9]. The ^{1}H-NMR spectrum (not depicted) showed a doublet for the anomeric proton of Gal*f* at $\delta = 5.265$ with a coupling constant $J_{1,2} = 4.8$ Hz, indicating an α-furanose configuration. The chemical shifts of relevant protons were determined by two-dimensional DQF ^{1}H-^{1}H COSY and are listed in Table 2. The downfield shift of the GalOH H-2 signal ($\delta = 4.076$) in comparison to the H-2 position of free GalOH ($\delta = 3.97$) is in agreement with a substitution at C-2. Combining the data, the structure of the disaccharide alditol is demonstrated to be

D-Gal*f*-α(1 → 2)-D-GalOH . (3)

It should be noted that D-GalOH, which is derived from D-Gal [5], is a symmetrical molecule having in principle the possibility of substitution at C-5 instead of at C-2 for the compounds **1**, **2** and **3**. To clarify this ambiguity a cellulosome sample was subjected to a β-elimination procedure without the addition of sodium borohydride. The reducing forms of compounds **1**, **2** and **3** could be isolated and ^{1}H-NMR spectroscopy demonstrated the presence of Gal H-1 signals at $\delta =$ 4.395 for **1** and **2** and at $\delta = 4.680$ for **3**, all having coupling constants of $J_{1,2} = 7.7-7.9$ Hz, indicating the occurrence of a reducing β-D-Gal*p* unit, thereby excluding the 1 → 5 linkage.

Fraction III contained too little carbohydrate material for further detailed structural analysis, but ^{1}H-NMR spectroscopy suggested the presence of some related oligosaccharide alditols. For this reason another cellulosome sample was treated with alkaline borohydride using more drastic conditions. The mixture of oligosaccharide alditols was separated on Bio-Gel P-2 yielding fractions AI–AIII (Fig. 4). Fraction AIII contained free galactitol and fractions AI and AII turned out to be complex mixtures. HPLC on Lichrosorb-NH_2 of fraction AI (Fig. 5) yielded three carbohydrate-containing subfractions denoted AI.1–AI.3; their monosaccharide analyses are given in Table 1. Fraction AI.1 was identified as a 'peeling product' of **1** [10]. Fraction AI.3 contained compounds **1** and **2**.

Monosaccharide analysis of fraction AI.2 (Table 1) suggested a trisaccharide alditol, consisting of GalOH, Gal and 3-*O*Me-GlcNAc residues. Positive FAB-MS of fraction AI.2 (spectrum not depicted) showed a protonated molecular ion $(M+H)^+$ at *m/z* 562 and a cationized molecular ion $(M+Na)^+$ at *m/z* 584, indicating a molecular mass of 561 Da,

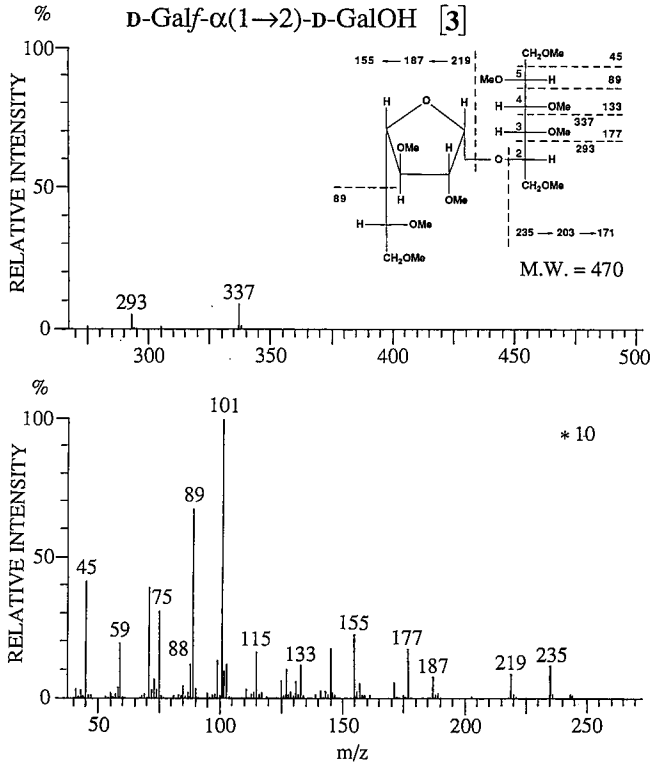


Fig. 3. *70-eV electron-impact mass spectrum of permethylated fraction IV derived from* C. thermocellum *YS cell-free cellulosome*

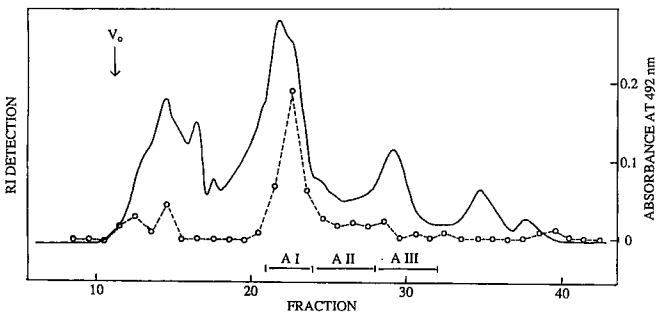


Fig. 4. *Elution profile on Bio-Gel P-2 of alkaline-borohydride-treated (0.2 M NaOH, 1 M $NaBH_4$, 16 h, 90°C) cell-free cellulosome of* C. thermocellum *YS*. The column (65 × 1.8 cm) was eluted with bi-distilled water at a flow rate of 10 ml/h. The eluate was monitored by refractive index detection (——) and by hexose determination with phenol/sulfuric acid (○-----○). The fractions indicated were pooled. V_o, void volume

in accordance with an assumed mono-*O*-methylated trisaccharide-alditol structure. Methylation analysis, including reduction with NaB^2H_4 showed non-labeled 1,3,4,5,6-penta-*O*-methyl-2-mono-*O*-acetylgalactitol representing the original GalOH residue substituted at C-2; 3,5,6-tri-*O*-methyl-1,2,4-tri-*O*-acetylgalactitol corresponding to a Gal*f* residue substituted at C-2; 2-*N*-methylacetamido-2-deoxy-3,4,6-tri-*O*-methyl-1,5-di-*O*-acetylglucitol demonstrating a terminal (3-*O*Me-)Glc*p*NAc residue.

The ^{1}H-NMR spectrum of fraction AI.2 (Fig. 6) showed two anomeric signals of equal intensity at $\delta = 5.401$ ($J_{1,2} = 4.8$ Hz) and $\delta = 5.001$ ($J_{1,2} = 3.6$ Hz), which points to the presence of a single compound. As compared to the ^{1}H-NMR data of **1**, these doublets can be assigned to α-Gal*f* H-1 and

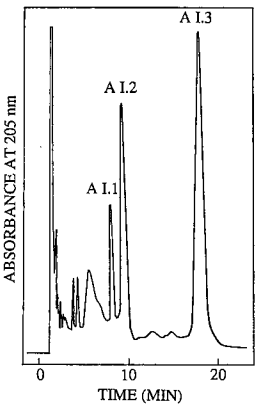


Fig. 5. *HPLC elution profile on Lichrosorb-NH_2 of fraction A I derived from* C. thermocellum *YS cell-free cellulosome*. The column (250 × 4.6 mm) was run isocratically with a mixture of acetonitrile/water, 80:20 (by vol.) at a flow rate of 1.0 ml/min. The eluate was monitored at 205 nm. The peaks indicated were isolated

3-*O*Me-α-Glc*p*NAc H-1, respectively. Singlets of equal intensity at $\delta = 3.528$ and $\delta = 2.073$ originate from the OCH_3 and *N*-acetyl groups of 3-*O*Me-Glc*p*NAc, respectively. Owing to a substituent at C-2 (see methylation analysis), Gal*f* H-2 ($\delta = 4.218$, $J_{2,3} = 8.7$ Hz) and Gal*f* H-3 ($\delta = 4.363$, $J_{3,4} = 8.1$ Hz) have shifted downfield, in comparison to **3**. The GalOH H-2 signal has shifted to a more downfield position ($\delta = 4.054$) than that of free GalOH ($\delta = 3.97$), in agreement with a substituent at C-2. In summary, it can be concluded that the structure of the compound is

3-*O*Me-D-Glc*p*NAc-α(1 → 2)-D-Gal*f*-α(1 → 2)-D-GalOH . **(4)**

Fraction AII could be separated on Bio-Gel P-2 into two fractions AII.1 and AII.2. Fraction AII.2 contained compound **3**, whereas fraction AII.1 was still a mixture. Owing to the very small amount of material, further subfractionation of AII.1 failed, but on the basis of its ^{1}H-NMR spectrum three compounds in almost equal amounts could be traced. No OCH_3 signals were observed. Two of the components were identified to be compound **2** and a 'peeling product' of **2** [10]. In view of essentially identical chemical-shift differences ($\Delta\delta$) of the anomeric and NAc protons observed in going from **1** to **2** and **4** to **5**, the third component was deduced to be

D-Glc*p*NAc-α(1 → 2)-D-Gal*f*-α(1 → 2)-D-GalOH . **(5)**

The cell-free cellulosomes of two other *C. thermocellum* strains, namely, ATCC 27405 and LQRI, and the cell-associated form of strain YS were investigated in a similar way as described for the cell-free form of strain YS. In Table 3 the carbohydrate content and monosaccharide composition of the different complete cellulosomes are summarized. For all samples, comparable amounts of Gal, GlcNAc and 3-*O*Me-GlcNAc were detected, with Gal as the major component. The carbohydrate content varied from 4.5 to 6.8% (by mass). Separation on Bio-Gel P-2 of the four alkaline-borohydride-treated cellulosome preparations gave rise to similar elution patterns as depicted in Fig. 1A. In Table 4, the monosaccharide compositions of the total low-molecular-mass fractions are presented, showing essentially no difference between the released mixtures of alditols. For each preparation identical carbohydrate structures, as reported above, could be identified in isolated fractions. The relative amounts of compounds **1** – **6** (Table 5), as calculated from hexose deter-

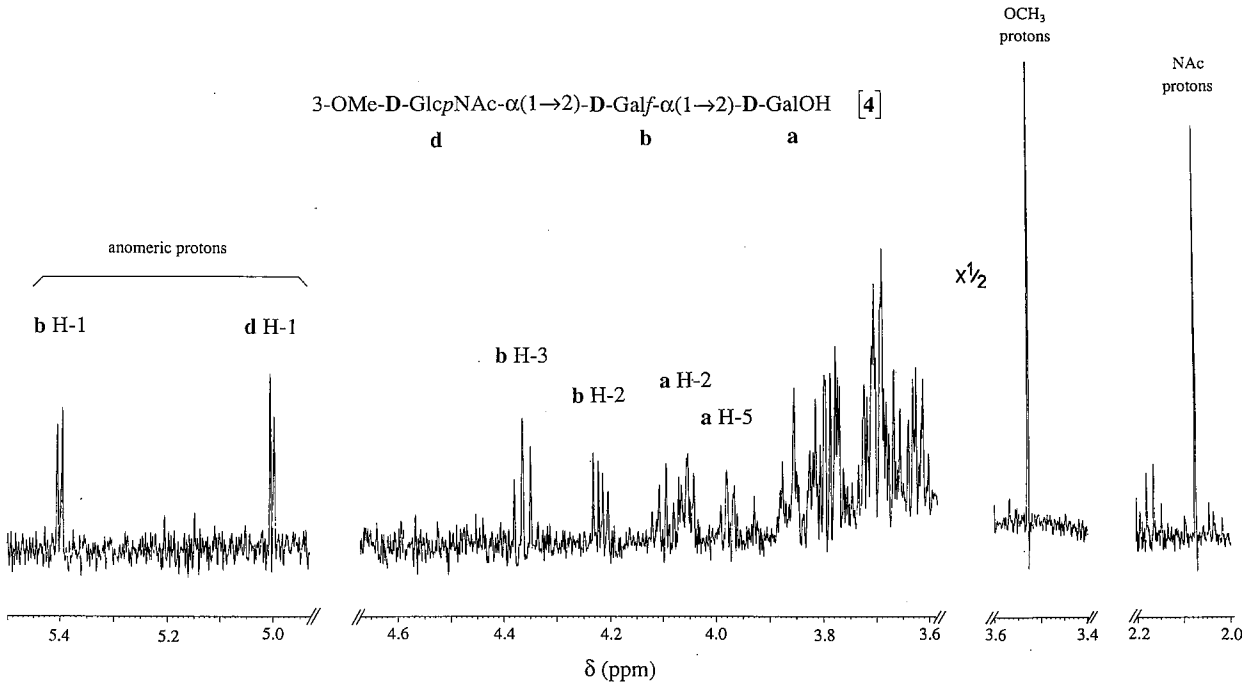


Fig. 6. *500-MHz ^1H-NMR spectrum of fraction A I.2 derived from* C. thermocellum *YS cell-free cellulosome recorded in ^2H_2O at 300 K.* The assignment of the protons refer to the corresponding residues in the structure. The HO^2H signal (4.70–4.90 ppm) has been omitted. The relative intensity scale of the *O*-methyl and *N*-acetyl regions differs from that of the other part of the spectrum as indicated

Table 3. *Carbohydrate content and composition of the cellulosome of different* C. thermocellum *strains*
All sugars possess the D-configuration. For the molar ratio 3-*O*Me-GlcNAc was taken as 1.0. Amount of contaminating Glc can differ with different batches. For quantification of 3-*O*Me-GlcNAc, the molar adjustment factor of GlcNAc was used

Monosaccharide	Molar ratio of strain			
	YS		ATCC 27405 cell-free	LQRI cell-free
	cell-free	cell-bound		
Gal	6.3	6.1	5.9	5.7
Glc	0.5	1.5	0.7	0.7
3-*O*Me-GlcNAc	1.0	1.0	1.0	1.0
GlcNAc	0.3	0.5	0.3	0.4
	%			
content (by mass)	6.8	4.5	5.5	5.1

Table 4. *Molar carbohydrate composition of the low-molecular-mass fraction of alkaline-borohydride-treated cellulosome of different* C. thermocellum *strains*
All sugars possess the D-configuration. For the molar ratio GalOH was taken as 1.0. For quantification of 3-*O*Me-GlcNAc, the molar adjustment factor of GlcNAc was used

Monosaccharide	Molar ratio of strain			
	YS		ATCC 27405 cell-free	LQRI cell-free
	cell-free	cell-bound		
Gal	1.6	1.5	1.7	1.8
GalOH	1.0	1.0	1.0	1.0
3-*O*Me-GlcNAc	0.5	0.4	0.4	0.5
GlcNAc	0.2	0.2	0.1	0.1

mination with phenol/sulfuric acid and from signal intensity ratios of anomeric, OCH_3 and NAc protons in the ^1H-NMR spectra of the alditol mixtures, are comparable for the four cellulosome preparations, demonstrating that there is no apparent type specifity and no significant difference between cell-free and cell-bound cellulosomes concerning O-linked carbohydrate chains. In view of the applied isolation procedure and the absence of free oligosaccharides, it is unlikely that **2**–**6** are formed by partial solvolysis of **1**.

DISCUSSION

The prokaryotic genus *Clostridium* belongs to the Eubacteria. For several years, it was generally assumed that the presence of prokaryotic glycoproteins was restricted to Archaebacteria and represented an important biochemical distinction from Eubacteria in concept of evolution [11, 12]. However, glycosylation of proteins has recently been observed in Eubacterial species, especially with respect to their crystalline surface layers [13–15].

Several reports on the cellulase complex of different *C. thermocellum* strains mention the occurrence of covalently bound or associated carbohydrates. The cellulase complex of strain NCIB 10682 was reported to have a carbohydrate content of 9.9% (by mass) [16], whereas the strain LQRI was found to secrete an *endo*-glucanase containing 11.2% (by mass) carbohydrate [17]. Investigation of the strains JW 20 (ATCC 31449) and YM4 [18–21] showed that their cellulosomes had 6–13% carbohydrate, with the highest amount of carbohydrate (Gal and probably GlcNAc) associated with a 210 kDa subunit. In our work on the cell-free

Table 5. *Relative amount of the oligosaccharide alditols* **1**–**6** *found in alkaline-borohydride-treated cellulosome of different* C. thermocellum *strains*
The numbers in bold refer to the oligosaccharide-alditol structures displayed in Fig. 7

Oligosaccharide alditol	Strain			
	YS		ATCC 27405 cell-free	LQRI cell-free
	cell-free	cell-bound		
	%			
1	47	47	47	47
2	14	19	14	12
3	15	18	23	20
4	5	<5	<5	<5
5	<5	<5	<5	<5
6	15	12	10	15

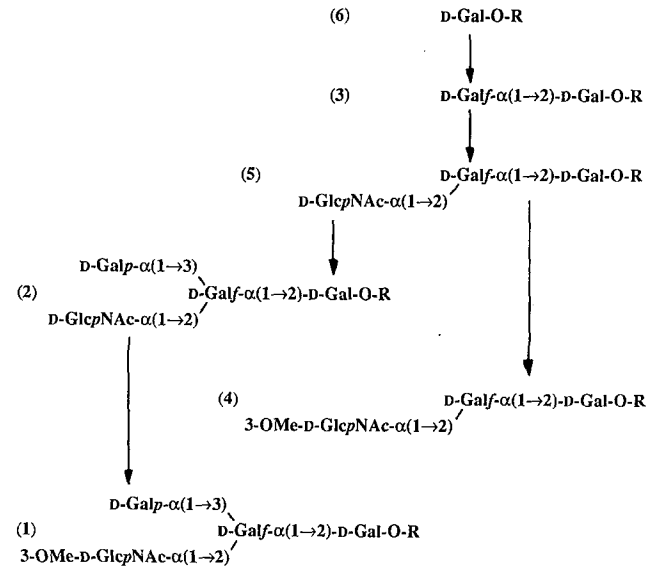


Fig. 7. *Hypothetical biosynthetic pathway of the oligosaccharide chains occurring in* C. thermocellum *cellulosome.* The numerals between brackets refer to the structures discussed in the Results. **R**, protein core

and cell-associated forms of the cellulosome of strain YS [5, and this study] we have characterized a series of novel glycoprotein glycans, which can be released under alkaline conditions, suggesting a glycopeptide linkage between Gal and Ser or Thr. No indications were obtained for the occurrence of N-linked oligosaccharides [5]. Two types of O-linked oligosaccharides can be distinguished, (1) D-Gal*p*-β(1 → 4)-D-Gal and (2) 3-*O*Me-D-Glc*p*NAc-α(1 → 2)-[D-Gal*p*-α(1 → 3)]-D-Gal*f*-α(1 → 2)-D-Gal together with partial structures of the tetrasaccharide (**1**–**6**). The relative amounts of compounds **1**–**6** were similar for both the cell-free and the cell-bound forms of strain YS. Investigation of the cell-free cellulosomes of strains ATCC 27405 and LQRI demonstrated similar patterns as present in strain YS, which excludes strain specificity. As reported earlier [5, 22, 23] similar types of carbohydrate structures may be present in cellulosome-like complexes of other cellulolytic bacteria, as indicated by agglutination experiments with the terminal α-D-Gal specific lectin GS-I from *Griffonia simplicifolia* and work is in progress to obtain more detailed structural information.

From a biosynthetic point of view it is interesting to consider the glycosylation mechanism of bacterial glycoproteins. So far, detailed information is only available concerning the *N*-glycosylation of the cell-surface glycoprotein of the genus *Halobacterium* of Archaebacteria [15]. Prokaryotes lack the subcellular organelles typically involved in eukaryotic glycoprotein biosynthesis. However, the transfer of glycosaminoglycan built up from repeating sulfated-pentasaccharide units (GalNAc-Asn linkage) and sulfated-tetrasaccharide elements (Glc-Asn linkage) in halobacteria still proceeds via lipid-linked precursors to Asn in an Asn–X–Ser/Thr-acceptor sequence. As compared to eukaryotes, it has been shown that in case of halobacteria other biosynthetic pathways are followed to reach the glycoprotein state. For instance, sulfation of the oligosaccharides is already completed at the lipid-diphosphate (GalNAc-Asn) or dolichol-monophosphate (Glc-Asn) level. Furthermore, there are clear indications that the sulfated tetrasaccharides are transferred to the protein core on the cell surface, whereby a transient methylation of the lipid-linked intermediate is involved at some stage in the biosynthesis of the halobacterial glycoprotein. Such methylation has not been found to be transient for the *Methanothermus fervidus* glycoprotein [24].

For the *O*-glycosylation of bacterial glycoproteins no information is available with respect to the type of precursors, i.e. nucleotide-activated or lipid-activated precursors, or to the site of glycosylation, i.e. inside the cell or at the cell surface. It is well known that the *O*-glycosylation of mucin-type glycoproteins in higher organisms does not proceed via an assembled lipid-linked oligosaccharide intermediate, transferred as one unit to Ser or Thr. Here, in a step-wise glycosylation, the various monosaccharide constituents are added by individual transfers via their nucleotide-activated derivatives [25]. Based on the observed parallels between the *N*-glycosylation of eukaryotic and prokaryotic glycoproteins, with respect to the involvement of lipid-linked intermediates, it is tempting to suggest that a parallel also occurs between both *O*-glycosylation mechanisms. In this context, the identification of the strongly related oligosaccharides **1**–**6** leads to the proposal of a biosynthetic pathway, as depicted in Fig. 7. The initiation could take place by transfer of a Gal*p* or Gal*f* residue to the protein core, followed by elongation through addition of the individual monosaccharides from their nucleotide-activated derivatives. According to the one-linkage/one-enzyme concept, four glycosyltransferases and one methyltransferase should be involved in the biosynthetic process. Because the trisaccharide alditol D-Gal*p*-α(1 → 3)-D-Gal*f*-α(1 → 2)-D-GalOH was not found in our structural studies, the biosynthetic route probably proceeds via **5** to **2** and then to **1**. A detour to **4** may occur if 3-*O*-methylation functions as a biosynthetic stop signal. Alternatively, the occurrence of a 3-*O*-methylated monosaccharide can indicate that the biosynthesis proceeds via a lipid-linked 3-*O*-methylated oligosaccharide, important for the transfer of the oligosaccharide to the glycoprotein as mentioned above for the *N*-glycosylation. However, then a trimming machinery, including a demethylation step, also has to be present. It is evident that *C. thermocellum* forms an excellent model system to get more information about the biosynthesis of O-linked bacterial glycoprotein glycans.

We thank C. Versluis and A. C. van der Kerk-van Hoof (Laboratory of Analytical Chemistry, Utrecht University) for recording the mass spectra.

REFERENCES

1. Lamed, R., Setter, E., Kenig, R. & Bayer, E. A. (1983) *Biotechnol. Bioeng. Symp. 13*, 163–181.
2. Lamed, R. & Bayer, E. A. (1988) *Adv. Appl. Microbiol. 33*, 1–46.
3. Bayer, E. A., Setter, E. & Lamed, R. (1985) *J. Bacteriol. 163*, 552–559.
4. Lamed, R. & Bayer, E. A. (1988) *Methods Enzymol. 160*, 472–482.
5. Gerwig, G. J., de Waard, P., Kamerling, J. P., Vliegenthart, J. F. G., Morgenstern, E., Lamed, R. & Bayer, E. A. (1989) *J. Biol. Chem. 264*, 1027–1035.
6. Kamerling, J. P. & Vliegenthart, J. F. G. (1989) in *Clinical biochemistry – principles, methods, applications, vol. 1, mass spectrometry* (Lawson, A. M., ed.) pp. 175–263, Walter de Gruyter, Berlin, New York.
7. Gerwig, G. J., Kamerling, J. P. & Vliegenthart, J. F. G. (1978) *Carbohydr. Res. 62*, 349–357.
8. Gerwig, G. J., Kamerling, J. P. & Vliegenthart, J. F. G. (1979) *Carbohydr. Res. 77*, 1–7.
9. Kärkkäinen, J. (1970) *Carbohydr. Res. 14*, 27–33.
10. Lloyd, K. O., Kabat, E. A. & Licerio, E. (1968) *Biochemistry 7*, 2976–2990.
11. Mescher, M. F. (1981) *Trends Biochem. Sci. 6*, 97–99.
12. Paul, G. & Wieland, F. (1987) *J. Biol. Chem. 262*, 9587–9593.
13. Messner, P. & Sleytr, U. B. (1988) in *Crystalline bacterial cell surface layers* (Sleytr, U. B., Messner, P., Pum, D. & Sára, M., eds) pp. 11–16, Springer-Verlag, Berlin.
14. Sleytr, U. B. & Messner, P. (1988) *J. Bacteriol. 170*, 2891–2897.
15. Lechner, J. & Wieland, F. (1989) *Annu. Rev. Biochem. 58*, 173–194.
16. Ait, N., Creuzet, N. & Forget, P. (1979) *J. Gen. Microbiol. 113*, 399–402.
17. Ng, T. K. & Zeikus, J. G. (1981) *Biochem. J. 199*, 341–350.
18. Coughlan, M. P., Hon-Nami, K., Hon-Nami, H., Ljungdahl, L. G., Paulin, J. J. & Rigsby, W. E. (1985) *Biochem. Biophys. Res. Commun. 130*, 904–909.
19. Hon-Nami, K., Coughlan, M. P., Non-Nami, H. & Ljungdahl, L. G. (1986) *Arch. Microbiol. 145*, 13–19.
20. Mayer, F., Coughlan, M. P., Mori, Y. & Ljungdahl, L. G. (1987) *Appl. Environ. Microbiol. 53*, 2785–2792.
21. Ljungdahl, L. G., Coughlan, M. P., Mayer, F., Mori, Y., Hon-Nami, H. & Hon-Nami, K. (1988) *Methods Enzymol. 160*, 483–506.
22. Lamed, R. & Bayer, E. A. (1987) in *Biochemistry and genetics of cellulose degradation* (Aubert, J. P., Béguin, P. & Millet, J., eds) pp. 101–106, Academic Press, Orlando, FL.
23. Lamed, R., Naimark, J., Morgenstern, E. & Bayer, E. A. (1987) *J. Bacteriol. 169*, 3792–3800.
24. Hartmann, E. & König, H. (1989) *Arch. Microbiol. 151*, 274–281.
25. Carraway, K. L. & Hull, S. R. (1989) *Bioessays 10*, 117–121.