THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 264, No. 2, Issue of January 15, pp. 1027-1035, 1989
Printed in U.S.A.

Novel *O*-Linked Carbohydrate Chains in the Cellulase Complex (Cellulosome) of *Clostridium thermocellum*

3-*O*-METHYL-*N*-ACETYLGLUCOSAMINE AS A CONSTITUENT OF A GLYCOPROTEIN*

(Received for publication, July 15, 1988)

Gerrit J. Gerwig, Pieter de Waard, Johannis P. Kamerling, and Johannes F. G. Vliegenthart‡
From the Department of Bio-Organic Chemistry, Utrecht University, Transitorium III, NL-3508 TB Utrecht, The Netherlands

Ely Morgenstern and Raphael Lamed
From the Department of Biotechnology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel

Edward A. Bayer
From the Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel

Alkaline borohydride treatment of the cellulosome of *Clostridium thermocellum* yielded two major oligosaccharide-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-OMe-D-Glc*p*NAc-α(1→2)

The compounds, isolated via gel permeation chromatography and high performance liquid chromatography, were analyzed by monosaccharide analysis, methylation analysis, gas-liquid chromatography/mass spectrometry, fast atom bombardment/mass spectrometry, and one- and two-dimensional 500-MHz (COSY, HOHAHA, ROESY) ^{1}H NMR spectroscopy. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis combined with blotting technology indicated that the tetrasaccharide is mainly associated with one of the cellulosome subunits.

The search for new renewable raw materials for the production of energy and food has led to increasing interest in the enzymatic hydrolysis of cellulosic materials (1). Cellulolytic enzymes are produced by a large number of microorganisms (2). Among these, the anaerobic thermophilic cellulolytic bacterium, *Clostridium thermocellum*, is known for its production of very high levels of an extracellular cellulase system of particularly high specific activity (3, 4). Most of the cellulases in this organism are organized into a multicomponent complex, termed the cellulosome (5, 6). The cellulosome occurs in both extracellular and cell-surface forms and consists of at least 14 polypeptide subunits, most of which appear to be cellulases (7). Some of the polypeptides are glycosylated (8–11); in particular, the 210-kDa component (the S1 subunit) has been estimated to contain up to 40% covalently bound carbohydrate (6).

To gain more insight into the mechanisms involved in the enzymatic hydrolysis of cellulose by *C. thermocellum*, detailed information of the molecular structures of its cellulase system is a prerequisite. This study describes the structural analysis of novel *O*-linked carbohydrate chains of the cellulosome.

EXPERIMENTAL PROCEDURES

Isolation of the Cellulosome—The cellulosome was isolated from *C. thermocellum* strain YS as described earlier (5). Before structural analysis, samples were intensively dialyzed against water (48 h running tap water, 24 h bidistilled water) followed by lyophilization.

SDS-PAGE[1] *and Blotting*—SDS-PAGE of cellulosome samples (10 μg) was carried out on 6% gels using previously described conditions (12). The separated proteins were either stained using Coomassie Brilliant Blue or transferred electrophoretically onto nitrocellulose sheets. The presence of vicinal hydroxyl groups on sugars was determined by periodate oxidation of the blotted material using the enzyme hydrazide technique (13). The presence of galactose residues was determined enzymatically on blots by combining galactose oxidase with avidin-biotin technology (14). A biotinylated form of isolectin GSI-B_4 from *Griffonia simplicifolia* (Sigma) was employed to detect terminal α-galactosyl residues on the blotted material (15).

Monosaccharide Analysis—Samples (0.1–1.0 mg) were subjected to methanolysis (0.5 ml) (1.0 M methanolic HCl, 24 h, 85 °C) followed by GLC of the trimethylsilylated (*N*-reacetylated) methyl glycosides on a capillary SE-30 fused silica column (25 m × 0.32 mm, Pierce) using a Varian 3700 gas chromatograph (temperature program 130–220 °C at 4 °C/min) (16). Identification of the monosaccharide derivatives was confirmed by GLC-MS (17). The absolute configuration of the monosaccharides was determined by GLC of the trimethylsilylated (*N*-reacetylated) (−)-2-butyl glycosides (18, 19).

PNGase-F Treatment—A cellulosome sample (20 μl, 2.2 mg/ml) was brought to 0.5% SDS and 0.1 M β-mercaptoethanol, and the solution was boiled for 3 min. The sample was then diluted to 60 μl in 0.2 M sodium phosphate buffer (pH 8.5), 10 mM 1,10-phenanthroline hydrate and 1.25% Nonidet P-40 (final concentration in each

* This investigation was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence and reprint requests should be addressed.

[1] The abbreviations used are: SDS-PAGE, sodium dodecyl sulfate/polyacrylamide gel electrophoresis; GLC-MS, gas-liquid chromatography/mass spectrometry; HPLC, high performance liquid chromatography; FAB-MS, fast atom bombardment/mass spectrometry; DQF ^{1}H-^{1}H COSY, double-quantum filtered ^{1},H-^{1}H correlation spectroscopy; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; ROESY, rotating frame nuclear Overhauser enhancement spectroscopy; NOE, nuclear Overhauser enhancement; Xyl, xylose; 3-*O*Me-Glc*p*NAc,2-acetamido-2-deoxy-3-*O*-methyl-glucopyranose; GalOH, galactitol; PNGase-F, peptide-N^4-(*N*-acetyl-β-glucosaminyl) asparagine amidase-F (EC 3.5.1.52).

case). An aliquot (2.5 μl) of PNGase-F (250 units/ml, Genzyme) was then added, and the reaction was allowed to proceed overnight at 37 °C. The relative mobility of the glycoproteins in the respective samples were examined by SDS-PAGE. Fetuin was used as a positive control.

Alkaline Borohydride Treatment—A cellulosome sample (20 mg) was treated with 10 ml of 0.1 M NaOH containing 1 M $NaBH_4$ (20). After 48 h of stirring at 37 °C under N_2, the solution was acidified to pH 5.0 with 4 M acetic acid at 0 °C and then applied to a column (15 × 1.2 cm) of Dowex 50W-X8, H^+ form (100–200 mesh, Fluka). The column was eluted with 50 ml of 0.01 M formic acid, and the eluate was lyophilized. Boric acid was removed by co-evaporation with methanol under reduced pressure. A small part (1/20) was used for monosaccharide analysis. The remaining material was fractionated on a column (65 × 1.8 cm) of Bio-Gel P-2 (200–400 mesh, Bio-Rad) using water as eluent (10.1 ml/h, 0.84-ml fractions). The eluate was monitored by refractive index detection with a Bischoff 8100 RI detector and by hexose determination with the phenol/sulfuric acid assay (21).

The high molecular mass fraction was subjected to a second β-elimination procedure using NaB^2H_4 and conditions as described above.

Thin Layer Chromatography—TLC was carried out on Kieselgel 60 F_{254} (0.2 mm) plastic sheets (Merck) in 1-butanol:ethanol:water (3:2:2, v/v) for 6 h at room temperature. The spots were visualized with a mixture of 0.2% orcinol in methanol and 20% H_2SO_4 in methanol (1:1, v/v) (2 min at 150 °C).

Methylation Analysis—Methylation analysis on 0.3-mg samples was carried out essentially as described (22). After permethylation and purification on a SepPak C_{18} column, the derivatized material was hydrolyzed with 4 M trifluoroacetic acid (0.2 ml) (4 h, 100 °C). The obtained mixture of partially methylated monosaccharides was reduced with $NaBH_4$ or NaB^2H_4 in water (0.4 ml) (3 h, room temperature) and finally acetylated with acetic anhydride (0.1 ml) (3 h, 121 °C). The partially methylated alditol acetates were analyzed by GLC and GLC-MS (23).

High Performance Liquid Chromatography—HPLC was carried out on a Kratos liquid chromatograph consisting of two Spectroflow 400 solvent delivery systems, a Spectroflow 450 solvent programmer, and a Rheodyne injection valve module using a Lichrosorb-10-NH_2 column (250 × 4.6 mm, Chrompack). Separations were performed isocratically with a mixture of acetonitrile:water (75:25, v/v) at a flow rate of 2.0 ml/min (24). The eluate was monitored by a Spectroflow 783 programmable absorbance detector at 205 nm.

Gas-Liquid Chromatography/Mass Spectrometry—Combined GLC-MS was performed on a Carlo-Erba GC/Kratos MS 80/Kratos DS 55 system; electron energy, 70 eV; accelerating voltage, 2.7 kV; ionizing current, 100 μA; ion source temperature, 225 °C; capillary BP 1 WCOT fused silica column (25 m × 0.32 mm); oven temperature program, 130 °C during 2 min, 130–260 °C at 4 °C/min.

Fast Atom Bombardment/Mass Spectrometry—Positive-ion mass spectra were recorded on a VG Analytical ZAB-HF mass spectrometer. The primary beam was composed of xenon atoms with a maximum energy of approximately 7.6 keV (1 mA). The carbohydrate samples (100 μg) were dispersed in glycerol and, if necessary, acidified with a small amount of acetic acid. The sputtered ions were extracted and accelerated with a potential of 8 kV. The spectra were obtained with an ultraviolet chart recorder.

500-MHz ^1H NMR Spectroscopy—Oligosaccharide-alditols (0.1–0.5 mg) were repeatedly exchanged in ^2H_2O (99.96 atom % ^2H, Aldrich) with intermediate lyophilization. Resolution-enhanced ^1H NMR spectra were recorded on a Bruker AM-500 spectrometer (Department of NMR Spectroscopy, Utrecht University) operating at 500 MHz at a probe temperature of 300 K. Chemical shifts (δ) are expressed in parts/million relative to sodium 4,4-dimethyl-4-silapentane-1-sulfonate, but were actually measured by reference to internal acetone in ^2H_2O (δ = 2.225 ppm).

Two-dimensional double-quantum filtered ^1H–^1H correlation spectra (two-dimensional DQF ^1H-^1H COSY) were obtained by a three-pulse sequence, 90°-t_1-90°-90°-Acq (25–28). The spectral width was 4000.00 Hz in both dimensions. A 32-step phase cycling was used, and 512 measurements of 2048 data points were recorded.

Two-dimensional homonuclear Hartmann-Hahn spectra (two-dimensional HOHAHA) were obtained by the pulse sequence, 90°-t_1-MLEV-17-Acq (29–32). The spectral width was 4000.00 Hz in both dimensions. A number of 512 measurements of 2048 data points were recorded, with 32 scans per t_1 value. A MLEV-17 mixing time of 120 ms was used and the 90° ^1H pulse width was 27 μs.

Two-dimensional rotating frame nuclear Overhauser enhancement spectra (two-dimensional ROESY) were obtained by a pulse scheme described as $90_\phi°$-t_1-[spin lock]$_{\phi+\pi/2}$-Acq (t_2) (33). The spectral width was 3333.33 Hz in both dimensions. The COSY, HOHAHA, and ROESY experiments were performed using the time proportional phase increment method (26), and the data were displayed in the phase-sensitive mode. The HO^2H signal was suppressed by a presaturation during 1.0 s.

Synthesis of Methyl 2-Acetamido-2-deoxy-3-O-methyl-α-D-glucopyranoside—Methyl 2-phthalimido-2-deoxy-4,6-*O*-isopropylidene-β-D-glucopyranoside (40 mg) was methylated with methyl iodide (0.7 ml) in the presence of sodium hydride (10 mg) in tetrahydrofuran (1 ml) (16 h, room temperature). The product was purified over a Kieselgel 60 column (15 × 1 cm, Merck). After removal of the phthaloyl group with hydrazine-hydrate (0.5 ml) in ethanol (1 ml), the free NH_2 function was *N*-acetylated with acetic anhydride (0.1 ml) in methanol (1 ml) (34). Finally, the isopropylidene group was removed with 1.0 M methanolic HCl (1 ml) (16 h, 85 °C). After *N*-reacetylation, methyl 2-acetamido-2-deoxy-3-*O*-methyl-α-D-glucopyranoside was obtained as the main product as could be established by ^1H NMR spectroscopy. ^1H NMR data (360-MHz, ^2H_2O): H-1, δ = 4.723 ppm, $J_{1,2}$ = 3.6 Hz; H-2, δ = 3.982 ppm, $J_{2,3}$ = 10.2 Hz; H-3 and H-4, δ = 3.55 ppm, $J_{4,5}$ = 9.5 Hz; H-5, δ = 3.680 ppm, $J_{5,6a}$ = 2.4 Hz, $J_{5,6b}$ = 5.2 Hz; H-6a, δ = 3.870 ppm; H-6b, δ = 3.774 ppm, $J_{6a,6b}$ = −12.2 Hz; NAc, δ = 2.038 ppm; 3-OCH_3, δ = 3.510 ppm; 1-OCH_3, δ = 3.390 ppm.

Preparation of Reference Methyl α- and β-D-Galactofuranoside—D-Galactose (250 mg) was methanolyzed with 1.0 M methanolic HCl (2 ml) (24 h, 85 °C). Me α-Gal*f* and Me β-Gal*f* were isolated from the methylglycoside mixture by reversed-phase HPLC on a CP Spher C_{18} column (250 × 4.6 mm, Chrompack) eluted with water (flow rate 1.0 ml/min), monitored at 195 nm. After lyophilization, the compounds were characterized by ^1H NMR spectroscopy. ^1H NMR data (500-MHz, ^2H_2O): Me α-D-Gal*f* H-1, δ = 4.884 ppm, $J_{1,2}$ = 4.0 Hz; H-2, δ = 4.132 ppm, $J_{2,3}$ = 8.0 Hz; H-3, δ = 4.122 ppm, $J_{3,4}$ = 7.2 Hz; H-4, δ = 3.787 ppm, $J_{4,5}$ = 5.7 Hz; H-5, δ = 3.716 ppm, $J_{5,6a}$ = 4.0 Hz, $J_{5,6b}$ = 8.0 Hz; H-6a, δ = 3.712 ppm; H-6b, δ = 3.598 ppm, $J_{6a,6b}$ = −12.7 Hz; OCH_3, δ = 3.433 ppm. Me β-D-Gal*f* H-1, δ = 4.908 ppm, $J_{1,2}$ = 2.0 Hz; H-2, δ = 4.042 ppm, $J_{2,3}$ = 3.6 Hz; H-3, δ = 4.067 ppm, $J_{3,4}$ = 6.1 Hz; H-4, δ = 3.955 ppm, $J_{4,5}$ = 4.3 Hz; H-5, δ = 3.826 ppm, $J_{5,6a}$ = 4.5 Hz, $J_{5,6b}$ = 7.4 Hz; H-6a, δ = 3.715 ppm; H-6b, δ = 3.648 ppm, $J_{6a,6b}$ = −11.7 Hz; OCH_3, δ = 3.412 ppm.

RESULTS

In Table I, the monosaccharide composition of the dialyzed cellulosome is presented. In addition to Gal as the major component, Glc, GlcNAc, and an unknown monosaccharide residue were detected. GLC-MS indicated the unknown compound to be a 3-*O*-methylated *N*-acetylhexosamine (35). The *gluco*-configuration was established by comparison with synthetic 3-*O*Me-Glc*p*NAc on GLC and further proven by methylation analysis and ^1H NMR spectroscopy (see below). The total carbohydrate content was 6.8% (w/w). Determination of the absolute configuration of the monosaccharides gave the D configuration in each case.

In order to obtain information about the type of carbohydrate chains, the cellulosome was first incubated with PNGase-F. No alteration in the protein pattern on SDS-

TABLE I

Carbohydrate composition of the cellulosome of C. thermocellum

Monosaccharide	T_M[a]	Molar ratio[b]	
			nmol/mg
Xyl	0.52; 0.55	Trace	+
Man	0.73; 0.77	Trace	+
Gal	0.74; 0.79; 0.80; 0.84	27.2	283.4
Glc	0.88; 0.91	2.3	23.4
3-*O*Me-GlcNAc[c]	1.02	4.3	45.1
GlcNAc	1.04; 1.12; 1.19	1.0	10.4

[a] GLC retention times relative to internal mannitol (M).

[b] GlcNAc taken as 1.0.

[c] For quantification, the molar adjustment factor of GlcNAc was used.

PAGE was observed, indicating the absence of *N*-linked oligosaccharide chains.

On the other hand, monosaccharide analysis of the material obtained after alkaline borohydride treatment of the cellulosome (AB, Table II) demonstrated the occurrence of the constituents as mentioned above, but a part of Gal (30%) was converted into GalOH. TLC of the alkaline borohydride-reduced product showed two carbohydrate-positive spots at the origin and at R_F 0.33, respectively. The low molecular mass compound had the same R_F value as maltotetraose. The mixture was separated on Bio-Gel P-2 yielding a high molecular mass carbohydrate fraction I and a low molecular mass carbohydrate fraction II, respectively (Fig. 1).

Fraction I—Monosaccharide analysis of fraction I showed Gal as the major component (Table II), in addition to a small amount of 3-*O*Me-GlcNAc. Since fraction I did not contain GalOH, it was subjected to a second alkaline borohydride treatment with NaB^2H_4 and then separated on Bio-Gel P-2 (Fig. 2). Monosaccharide analysis of I.1 yielded Gal and GalOH in a molar ratio of 1:1 (Table II), indicating a disaccharide-alditol. Compound I.1 was permethylated and, in conjunction with monosaccharide analysis, identified by GLC-MS to be Gal*p*-(1→4)-GalOH (Fig. 3) (*c.f.* Ref. 37). Finally, the β-configuration of the glycosidic linkage was determined by one-dimensional 500-MHz ^1H NMR spectroscopy. The spectrum (not depicted) showed a doublet for an anomeric proton at $\delta = 4.506$ ppm with a coupling constant $J_{1,2} = 7.8$ Hz. Combining the data, the structure for the disaccharide-alditol is demonstrated to be: D-Gal*p*-β(1→4)-D-GalOH.

Fraction II—Monosaccharide analysis of fraction II yielded

TABLE II

Molar carbohydrate composition of alkaline borohydride-treated cellulosome (AB) and derived fractions

Monosaccharide	Molar ratio[a]				
	AB	I	I.1	II	II.1
Gal	2.5	1.0	1.0	1.8	1.9
Glc	+	+	−	0.2	+
GalOH[b]	1.0	−	1.0	1.0	1.0
3-*O*Me-GlcNAc[c]	0.6	0.1	−	0.6	0.6
GlcNAc	0.2	+	−	0.1	−

[a] GalOH taken as 1.0.

[b] Corrected for anhydro-GalOH ($T_M = 0.68$) (36).

[c] For quantification, the molar adjustment factor of GlcNAc was used.

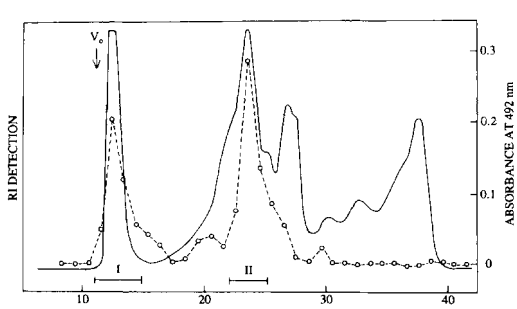


FIG. 1. **Elution profile on Bio-Gel P-2 of alkaline borohydride-treated cellulosome of *C. thermocellum*.** The column (65 × 1.8 cm) was eluted with bidistilled water. Fractions of 0.84 ml were collected at a flow rate of 10.1 ml/h. The eluate was monitored by refractive index detection (——) and by hexose determination with phenol/sulfuric acid (O- - -O). Fractions I and II were pooled.

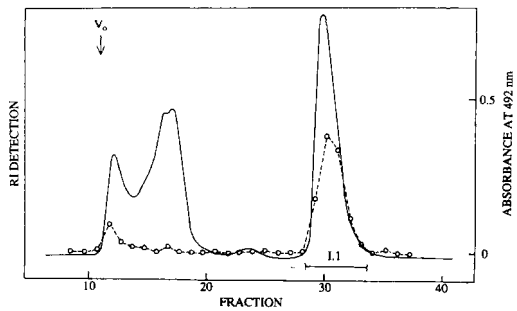


FIG. 2. **Elution profile on Bio-Gel P-2 of Fraction I derived from *C. thermocellum* cellulosome after a second alkaline borohydride treatment.** See Fig. 1 for conditions. Fraction I.1 was pooled.

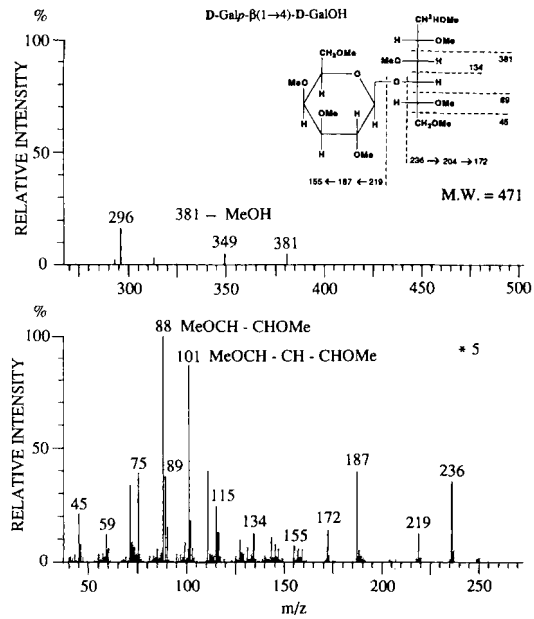


FIG. 3. **70-eV electron impact mass spectrum of permethylated Fraction I.1 derived from *C. thermocellum* cellulosome.**

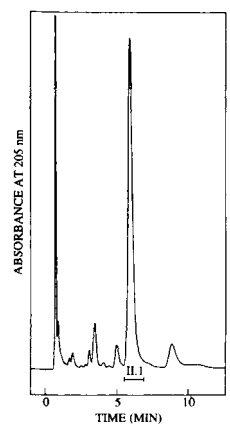


FIG. 4. **HPLC elution profile on Lichrosorb-NH_2 of Fraction II derived from *C. thermocellum* cellulosome.** The column (250 × 4.6 mm) was run isocratically with a mixture of acetonitrile/water (75:25, v/v) at a flow rate of 2.0 ml/min. The eluate was monitored at 205 nm. Fraction II.1 was pooled.

Gal, Glc, GlcNAc, 3-*O*Me-GlcNAc, and GalOH (Table II). After further purification by HPLC on Lichrosorb-NH_2 (Fig. 4), monosaccharide analysis of II.1 suggested the presence of a tetrasaccharide-alditol consisting of GalOH and, in addition,

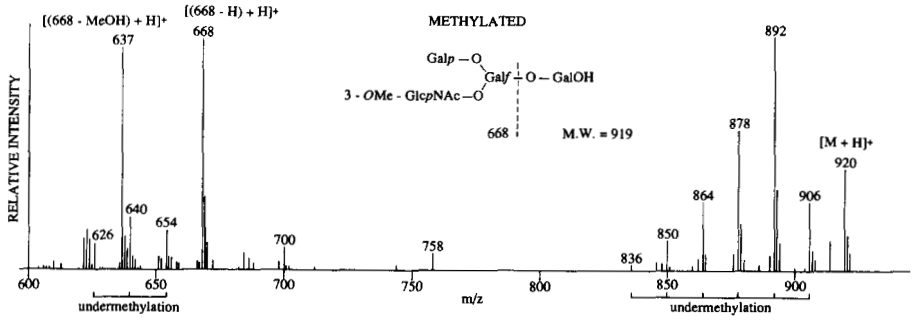


FIG. 5. **High mass region of the positive FAB-mass spectrum of methylated Fraction II.1 derived from *C. thermocellum* cellulosome.**

TABLE III

Methylation analysis of fraction II.1

Partially methylated alditol acetate	T_R[a]	Molar ratio[b]	Structural feature
1,3,4,5,6-Penta-*O*-methyl-2-mono-*O*-acetylgalactitol	0.69	0.5	→2)-GalOH
1,4,6-Tri-*O*-methyl-2,3,5-tri-*O*-acetylgalactitol	1.06	0.4	[undermethylation of →2)-GalOH]
2,3,4,6-Tetra-*O*-methyl-1,5-di-*O*-acetylgalactitol-(1-^{2}H)	1.00	1.0	Gal*p*-(1→
5,6-Di-*O*-methyl-1,2,3,4-tetra-*O*-acetylgalactitol-(1-^{2}H)	1.29	0.3	→2,3)-Gal*f*-(1→
2-*N*-Methylacetamido-2-deoxy-3,4,6-tri-*O*-methyl-1,5,-di-*O*-acetylglucitol-(1-^{2}H)	1.64	0.5	(3-*O*Me-)Glc*p*NAc-(1→

[a] GLC retention times on SE-30 relative to 2,3,4,6-tetra-*O*-methyl-1,5-di-*O*-acetylgalactitol.
[b] Calculated from peak areas, not corrected by response factors.

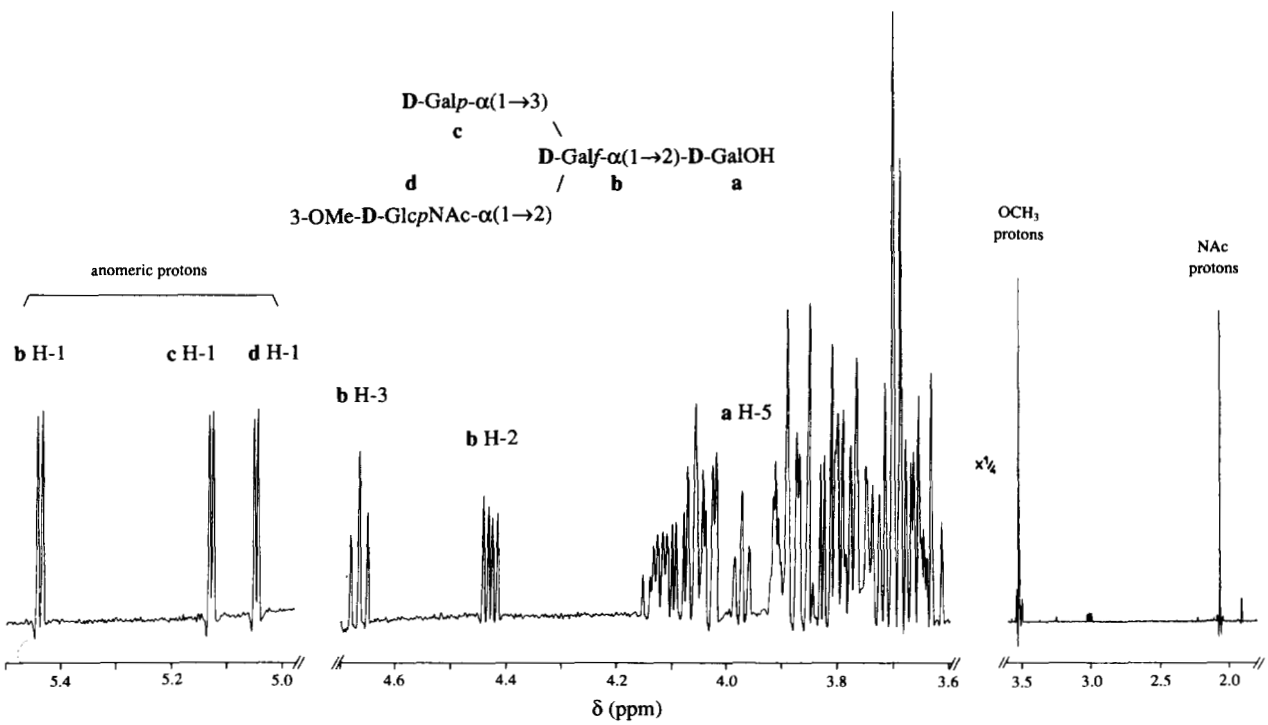


FIG. 6. **500-MHz ^{1}H NMR spectrum of Fraction II.1 derived from *C. thermocellum* cellulosome recorded in ^2H_2O at 300 K.** The denotation of the protons refer to the corresponding residues in the structure. The HO^{2}H signal (4.70–4.85 ppm) has been omitted. The relative intensity scale of the *O*-methyl and *N*-acetyl regions differs from that of the other part of the spectrum as indicated.

two Gal and one 3-*O*Me-GlcNAc residues (Table II).

The underivatized II.1 was investigated by positive FAB-MS (spectrum not depicted). The presence of the protonated molecular ion $(M + H)^+$ at *m/z* 724 and the cationized molecular ions $(M + Na)^+$ and $(M + K)^+$ at *m/z* 746 and at *m/z* 762, respectively, indicated a molecular mass of 723 Da, in accordance with the assumed tetrasaccharide-alditol structure. The positive FAB-mass spectrum of methylated II.1 (Fig. 5) showed, in addition to the protonated molecular ion $(M + H)^+$ at *m/z* 920, six peaks in the high mass region which originated from reproducible undermethylation. The undermethylation is mainly located in the GalOH residue, since the fragment ion at *m/z* 668, representing the permethylated part missing the alditol unit is accompanied by only three peaks derived from undermethylation.

Methylation analysis, including reduction with NaB^{2}H$_4$ of

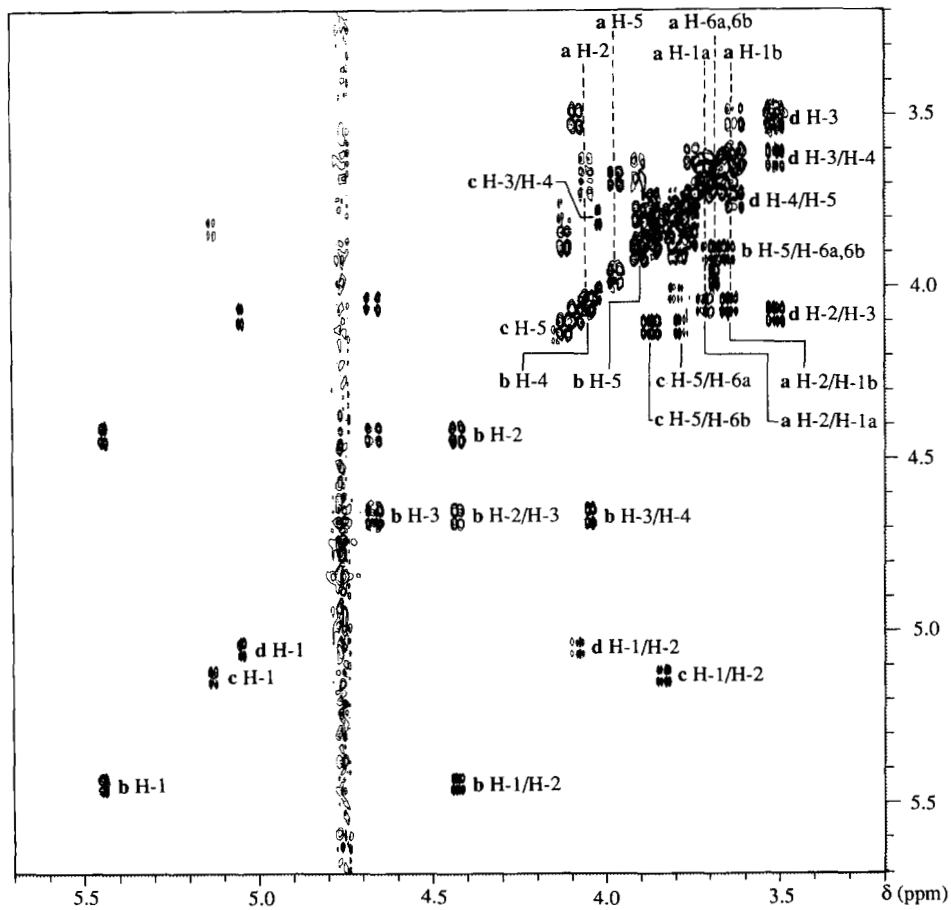


FIG. 7. **Two-dimensional DQF ^1H-^1H COSY spectrum (region 3.2–5.7 ppm) of Fraction II.1 derived from *C. thermocellum* cellulosome recorded in ^2H_2O at 300 K.** The denotation of the protons and cross-peaks refer to the corresponding residues in the structure (Fig. 6).

II.1 yielded the partially methylated alditol acetates presented in Table III. Two nonlabeled derivatives were detected, namely 1,3,4,5,6-penta-*O*-methyl-2-mono-*O*-acetylgalactitol representing the original GalOH residue substituted at C-2, and 1,4,6-tri-*O*-methyl-2,3,5-tri-*O*-acetylgalactitol. The latter alditol stems from undermethylation of the 2-substituted GalOH unit in line with the occurrence of *m/z* 892 in the FAB-mass spectrum. The finding of terminal Gal*p* and terminal 3-*O*Me-Glc*p*NAc residues indicated a branched tetrasaccharide-alditol. The detection of 5,6-di-*O*-methyl-1,2,3,4-tetra-*O*-acetylgalactitol demonstrated a branching Gal*f* residue substituted at C-2 and C-3.

500-MHz ^1H NMR spectroscopy was applied to identify the attachment sites of the 2 terminal monosaccharide units at the branching point and the configurations of the glycosidic linkages. The one-dimensional ^1H NMR spectrum of II.1 is presented in Fig. 6. For a complete interpretation of the spectrum, two-dimensional DQF ^1H-^1H COSY (Fig. 7), two-dimensional HOHAHA (Fig. 8), and two-dimensional ROESY (Fig. 9) experiments were carried out. The ^1H NMR chemical shift values are listed in Table IV.

A temperature experiment showed no signals under the HO^{2}H signal at $\delta = 4.70$–4.85 ppm. Starting with the anomeric signal of D-Gal*f* (residue **b**) at $\delta = 5.440$ ppm ($J_{1,2} = 4.6$ Hz), the COSY spectrum demonstrated connectivities for H-1/H-2, H-2/H-3, H-3/H-4, and H-5/H-6a,6b. The absence of a cross-peak for H-4/H-5 is probably due to small coupling. However, this connectivity can be observed in the two-dimensional HOHAHA spectrum on the **b**H-1 track, showing the subspectrum of Gal*f*H-1, H-2, H-3, H-4, and H-5. The downfield positions of Gal*f*H-2 ($\delta = 4.431$ ppm, $J_{2,3} = 8.2$ Hz) and H-3 ($\delta = 4.665$ ppm, $J_{3,4} = 7.2$ Hz) are due to the monosaccharide substituents at C-2 and C-3 (see above). The α-anomeric configuration of Gal*f* was concluded from comparison of the coupling constant of H-1 ($J_{1,2} = 4.6$ Hz) with those of the anomeric signals of methyl α-D-Gal*f* (H-1, $\delta = 4.884$ ppm, $J_{1,2} = 4.0$ Hz) and methyl β-D-Gal*f* (H-1, $\delta = 4.908$ ppm, $J_{1,2} = 2.0$ Hz).

The coupling constant of the H-1 signal of D-Gal*p* (residue **c**) at $\delta = 5.131$ ppm, being $J_{1,2} = 3.6$ Hz, indicated the α-configuration. The COSY spectrum gave rise to cross-peaks correlating H-1/H-2, H-3/H-4, H-5/H-6a, and H-5/H-6b. The H-2/H-3 cross-peak could not be clearly observed because of the close chemical shift values of both protons. The very small coupling constant between the equatorial H-4 and axial H-5 atom in Gal*p* is responsible for the absence of the H-4/H-5 connectivity. The HOHAHA spectrum showed on the **c**H-1 track the subspectrum of α-D-Gal*p* H-1—H-5, and on the **c**H-5 track that of α-D-Gal*p* H-4— H-6a,b.

Also, 3-*O*Me-D-Glc*p*NAc (residue **d**) had α-configuration, as could be deduced from the coupling constant of the anomeric signal at $\delta = 5.049$ ppm, being $J_{1,2} = 3.5$ Hz. In the COSY spectrum, cross-peaks were observed correlating H-1/H-2, H-2/H-3, H-3/H-4, and H-4/H-5. The HOHAHA spectrum showed on the **d**H-1 track the complete subspectrum of 3-*O*Me-α-D-Glc*p*NAc.

Singlets of equal intensity at $\delta = 3.518$ ppm and at $\delta = 2.063$ ppm originate from the OCH_3 and *N*-acetyl groups of 3-*O*Me-Glc*p*NAc, respectively. In the synthetic methyl glycoside of 3-*O*Me-α-D-Glc*p*NAc, the values for the OCH_3 at C-3 and the *N*-acetyl group are $\delta = 3.510$ ppm and $\delta = 2.038$ ppm, respectively.

The GalOH unit (residue **a**) showed in the COSY spectrum a cross-peak for H-5 at $\delta = 3.976$ ppm and H-6a,6b at $\delta =$

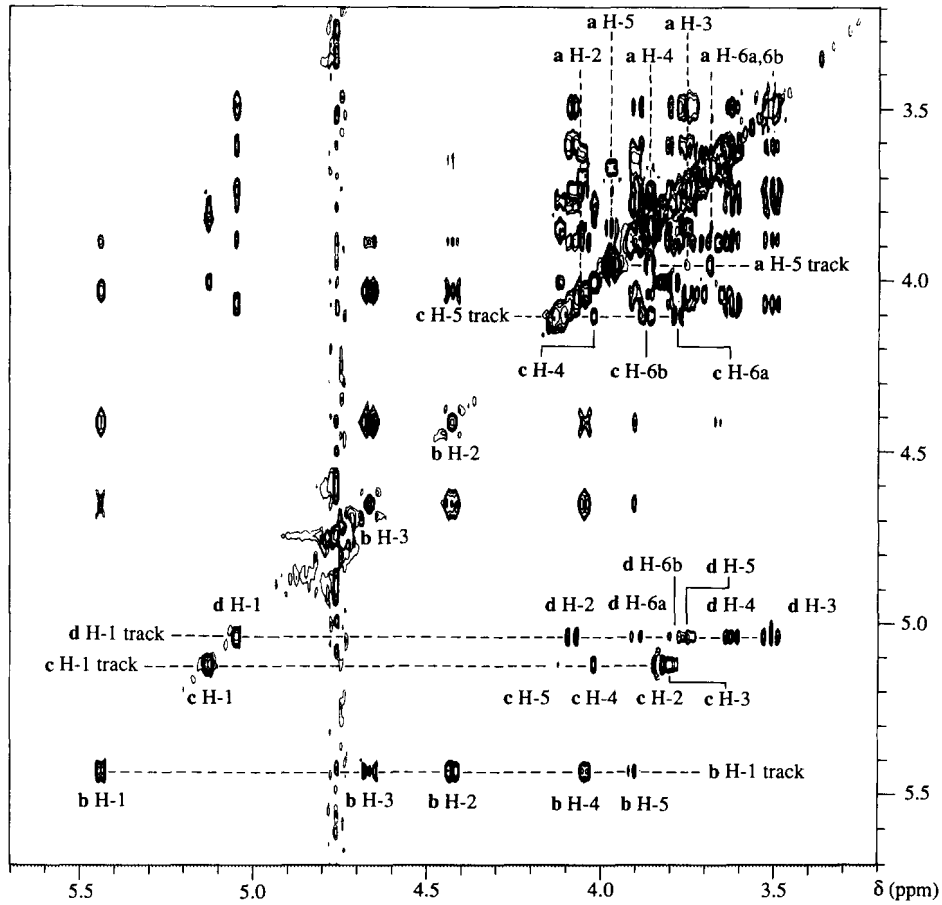


FIG. 8. **Two-dimensional HOHAHA spectrum (region 3.2–5.7 ppm) of Fraction II.1 derived from *C. thermocellum* cellulosome recorded in ^2H_2O at 300 K.** The denotation of the protons and cross-peaks refer to the corresponding residues in the structure (Fig. 6).

3.696 ppm. The observed chemical shifts are identical with those in free GalOH (H-1a,1b and H-6a,6b, δ = 3.696 ppm; H-3 and H-4, δ = 3.68 ppm; H-2 and H-5, δ = 3.97 ppm). Further connectivities were detected for H-2/H-1a and H-2/H-1b. The finding of the H-2 signal (δ = 4.060 ppm) at a more downfield position than in free GalOH is in agreement with a substitution at C-2. The HOHAHA spectrum showed on the **a**H-5 track the subspectrum of GalOH H-3, H-4, H-5, and H-6a,6b.

Finally, a two-dimensional ROESY spectrum, providing information about through-space nuclear Overhauser effect (NOE) connectivities, was recorded.

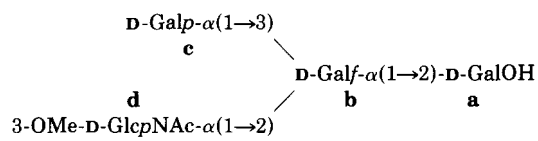
The (1→2) linkage between 2,3-substituted α-D-Gal*f* (**b**) and GalOH (**a**) was confirmed by a NOE connectivity between **b**H-1 and **a**H-2, marked **A.** It has to be noted that the α-configuration of D-Gal*f* has H-1, H-2, and H-4 atoms at the same side of the five-membered ring which is supported by H-1/H-2 and H-1/H-4 connectivities in the ROESY spectrum.

The observed through-space NOE connectivity between **c**H-1 and **b**H-3, marked **B**, indicated a (1→3) linkage between the α-D-Gal*p* (**c**) and α-D-Gal*f* (**b**) units. The absence of a connectivity between **c**H-1 and **b**H-2 supports this assignment. The inter-residual connectivities **c**H-1/**b**H-4 and **c**H-1/**b**H-5 indicated the spatial proximity of these protons, which is in agreement with the linkage of terminal α-D-Gal*p*(**c**) to Gal*f* (**b**) at C-3.

A through-space NOE connectivity between **d**H-1 and **b**H-2, marked **C,** could be detected, in accordance with a (1→2) linkage between the terminal 3-*O*Me-α-D-Glc*p*NAc (**d**) and α-D-Gal*f* (**b**) units. The absence of a NOE connectivity between **d**H-1 and **b**H-3 supports this assignment. The observed inter-residual connectivity **b**H-1/**d**H-1 is only possible if terminal 3-*O*Me-α-Glc*p*NAc (**d**) is linked to Gal*f* (**b**) at C-2.

Other observed intraresidual connectivities are in agreement with the proposed configurations of the individual monosaccharide units.

In conclusion, the various data indicate the structure of the tetrasaccharide-alditol to be:



The noncovalent association of the cellulosomal subunits into the multicomponent complex is remarkably stable, such that isolation of individual subunits in large quantities has presented insurmountable difficulties (5). We therefore reverted to other methodologies, in order to determine which subunit(s) contain carbohydrate chains. In this context, three different carbohydrate-specific staining techniques were carried out on SDS-PAGE-separated cellulosome samples. All three techniques demonstrated that the majority of label is associated with the highest molecular weight component, the M_r = 210,000 S1 subunit (Fig. 10). For general detection of sugars, a procedure based on periodate oxidation was employed (Fig. 10, *lane B*). Since the structures proposed for the cellulosomal disaccharide and tetrasaccharide are both rich in vicinal hydroxyl groups, the latter staining procedure would be expected to detect both oligosaccharide types. Likewise, the galactose oxidase-mediated staining procedure (Fig. 10, *lane C*) would also be expected to detect both types of carbohydrate structures, since both contain terminal galactosyl residues. The only staining procedure used in this study which

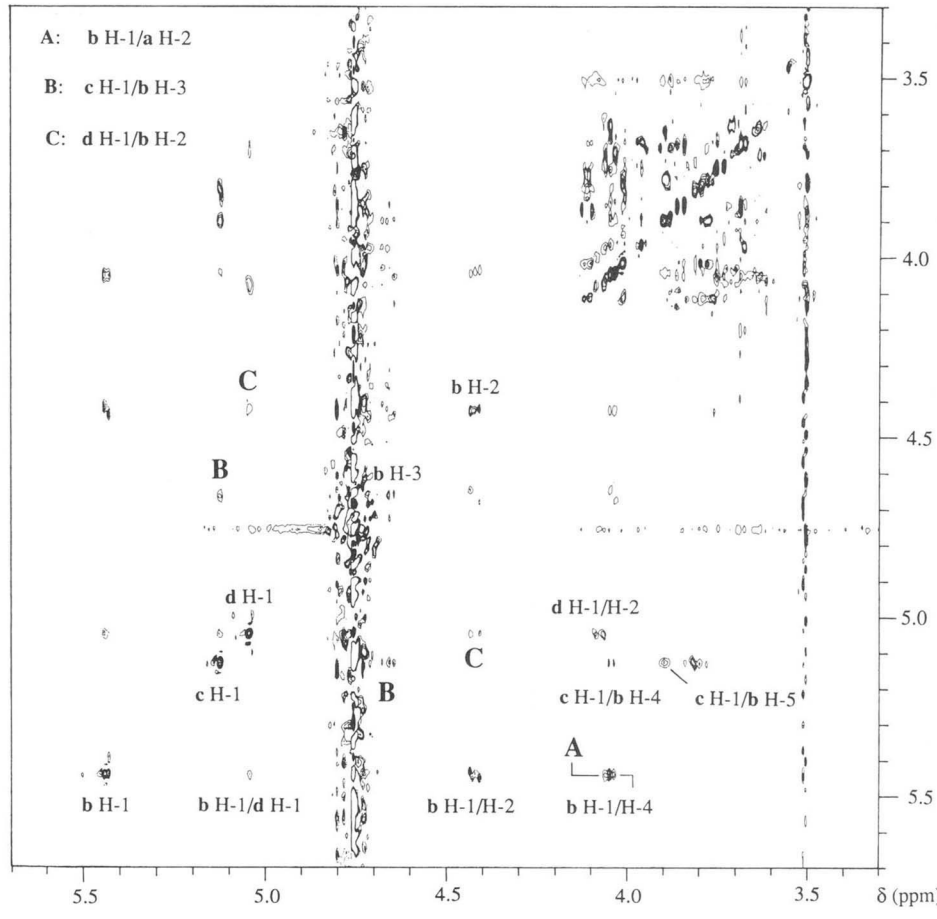


FIG. 9. **Two-dimensional ROESY spectrum (region 3.3–5.7 ppm) of Fraction II.1 derived from *C. thermocellum* cellulosome recorded in ^2H_2O at 300 K.** The denotation of the protons and cross-peaks refer to the corresponding residues in the structure (Fig. 6).

TABLE IV

^1H NMR data of the constituent monosaccharides for the tetrasaccharide-alditol derived from C. thermocellum cellulosome

Protons	Chemical shifts in			
	Gal*f* (**b**)	Gal*p* (**c**)	3-*O*Me-Glc*p*NAc (**d**)	GalOH (**a**)
	ppm		*ppm*	
H-1a	5.440	5.131	5.049	3.714
H-1b				3.648
H-2	4.431	3.837	4.087	4.060
H-3	4.665	3.803	3.517	3.760
H-4	4.048	4.024	3.632	3.857
H-5	3.908	4.124	3.752	3.976
H-6a	3.710	3.788	3.899	3.696
H-6b	3.660	3.875	3.790	3.696
NAc			2.063	
OCH_3			3.518	

would be expected to differentiate between the two oligosaccharides is that employed in Fig. 10, *lane D*. *G. simplicifolia* GS-I isolectin B_4 is specific for terminal α-D-galactosyl moieties and as such would presumably label the tetrasaccharide but not the disaccharide. The major portion of the tetrasaccharide thus appears to be associated with the S1 subunit in the cellulosome of *C. thermocellum* strain YS.

DISCUSSION

Both carbohydrate chains reported in this paper were released from the cellulosome under alkaline borohydride conditions, usually applied for the cleavage of *O*-linked chains. It is remarkable that the disaccharide is split off only after a second alkaline borohydride treatment and not during the first one. Maybe, due to the complexity of the cellulosome (consisting of several polypeptide subunits), the disaccharide

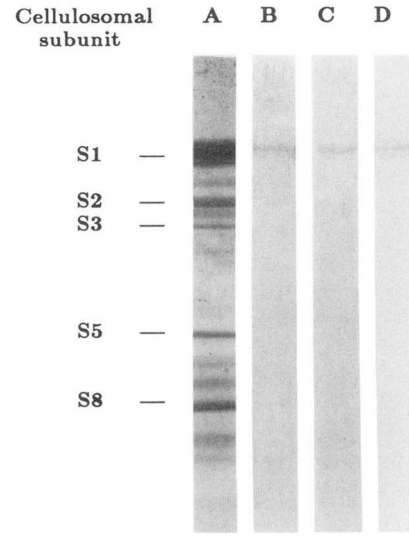


FIG. 10. **Carbohydrate-specific staining of blotted cellulosome preparations.** Samples of the cellulosome (10 μg) were resolved on 6% polyacrylamide gels. Gels were either stained with Coomassie Brilliant Blue (*lane A*) or blotted onto nitrocellulose membrane filters (*lanes B* through *D*). The blotted samples were oxidized either chemically with periodate (*lane B*) or enzymatically with galactose oxidase (*lane C*). A third blotted sample (*lane D*) was treated with the *Griffonia* lectin. In each case, an alkaline phosphatase-based detection system was employed, mediated by avidin-biotin technology. The positions of several of the confirmed cellulosomal subunits are designated; S1 (210 kDa), S2 (170 kDa), S3 (150 kDa), S5 (98 kDa), and S8 (75 kDa).

conjugate is protected by the tertiary structure and is achievable only after the first purification. This phenomenon is under current investigation.

From the SDS-PAGE experiments, it is clear that in *C. thermocellum* strain YS, the tetrasaccharide is a component of the S1 subunit of the cellulosome. The localization of the disaccharide in cellulosome subunits is less certain at this point. Although most of the carbohydrate generally appears to be associated with the S1 subunit, it remains to be determined whether the disaccharide is associated with S1 exclusively. The precise location of the disaccharide awaits the definition and selective employment of an appropriate (*e.g.* β-galactose-specific) lectin or antibody.

The presence of terminal α-galactopyranose in the cellulosomal tetrasaccharide is intriguing in view of earlier results which demonstrated that the *Griffonia* lectin is the only lectin (out of more than 20 examined) which interacts with cells of *C. thermocellum* (38). The same lectin was also shown to interact with a variety of evolutionarily divergent cellulolytic bacteria, but failed to interact with noncellulolytic strains (39). The possibility thus arises that similar types of carbohydrate structures may be associated with cellulosome-like complexes in other cellulolytic microorganisms, the implication being that such carbohydrates play a specific direct or indirect role in bacterial degradation of cellulose. A requisite and logical sequel to this study would therefore include a comparative structural analysis of oligosaccharides from other cellulolytic strains.

Both carbohydrate chains have not been found earlier in glycoprotein material. The occurrence of galactofuranose residues has been demonstrated in fungal and bacterial polysaccharides, in glycolipids, and in a few glycoproteins (40–50). Terminal galactofuranose residues occur in a cell wall glycoprotein from the fungus *Pithomyces chartarum* (46). An exocellular glycopeptide from *Penicillium charlesii* contains several poly-β(1→5)-linked galactofuranosyl chains (47). Also, the presence of galactofuranose residues in *N*-linked high mannose type oligosaccharides has been reported; *e.g.* in the protozoa *Crithidia fasciculata, Crithidia harmosa, Leptomonas samueli,* and *Herpetomonas samuelpessoai* (48) and in the fungus *Ascobolus furfuraceus* (49). Recently, it was clearly demonstrated that galactofuranose-containing, high mannose type oligosaccharides with compositions $Gal_{1-3}Man_9GlcNAc$ and $Gal_1Man_{7,8}GlcNAc$ occur in glycoproteins of the parasitic protozoan *L. samueli.* In all cases, the galactose units appeared to be linked to mannose in the terminal position (50).

O-Methylated monosaccharides have been reported as constituents of (bacterial) polysaccharides (51), glycolipids (52–56), and glycoproteins. In the last category of complex carbohydrates, 3-*O*-methyl-D-mannose has been found in glycoprotein material of the fungus *Mucor rouxii* (57), in hemocyanins of the molluscs *Lymneae stagnalis* (58–61) and *Acila castrensis* and *Stenoplex conspicua* (62), in hemoglobin of *Planorbis corneus* (63), and 3-*O*-methyl-D-galactose in hemocyanins of the molluscs *Helix pomatia* and *Lymneae stagnalis* (58–62). Recently, 3-*O*-methyl galacturonic acid has been found in a cell-surface glycoprotein of *Halobacterium halobium,* which contained galactofuranose, too (64).

This is the first report demonstrating the natural occurrence of 3-*O*-methyl-*N*-acetyl-D-glucosamine as a glycoprotein constituent. Three other *O*-methylated hexosamines have been reported: 6-*O*-methyl-glucosamine in lipopolysaccharide of strains of *Rhodopseudomonas palustris* (65), 3-*O*-methyl-galactosamine in oyster glycolipid (66) and in the spermatozoa glycolipid of a fresh water bivalve, *Hyriopsis schlegelii* (67), and 4-*O*-methyl-*N*-acetylglucosamine in a phosphonoglycosphingolipid of the skin of the sea gastropod, *Aplysia kurodai* (68).

Acknowledgments—We thank Dr. F. A. W. Koeman for the kind gift of methyl 2-phthalimido-2-deoxy-4,6-*O*-isopropylidene-β-D-glucopyranoside and C. Versluis and A. C. van der Kerk-van Hoof (Laboratory of Analytical Chemistry, Utrecht University) for recording the mass spectra.

REFERENCES

1. Mandels, M. (1982) in *Annual Report on Fermentation Processes* (Tsao, G. T., ed) Vol. 5, pp. 35–78, Academic Press, Orlando, FL
2. Enari, T-. M. (1983) in *Microbial Enzymes and Biotechnology* (Fogarty, W. M., ed) pp. 183–223, Applied Scientific Publishers, Essex, Great Britain
3. Ng, T. K., and Zeikus, J. G. (1981) *Appl. Environ. Microbiol.* **42,** 231–240
4. Johnson, E. A., Sakajoh, M., Halliwell, G., Madia, A., and Demain, A. L. (1982) *Appl. Environ. Microbiol.* **43,** 1125–1132
5. Lamed, R., Setter, E., Kenig, R., and Bayer, E. A. (1983) *Biotechnol. Bioeng. Symp.* **13,** 163–181
6. Lamed, R., and Bayer, E. A. (1988) *Adv. Appl. Microbiol.* **33,** 1–46
7. Bayer, E. A., Setter, E., and Lamed, R. (1985) *J. Bacteriol.* **163,** 552–559
8. Ait, N., Creuzet, N., and Forget, P. (1979) *J. Gen. Microbiol.* **113,** 399–402
9. Ng, T. K., and Zeikus, J. G. (1981) *Biochem. J.* **199,** 341–350
10. Hon-Nami, K., Coughlan, M. P., Hon-Nami, H., and Ljungdahl, L. G. (1986) *Arch. Microbiol.* **145,** 13–19
11. Bayer, E. A., and Lamed, R. (1986) *J. Bacteriol.* **167,** 828–836
12. Lamed, R., Setter, E., and Bayer, E. A. (1983) *J. Bacteriol.* **156,** 828–836
13. Bayer, E. A., Ben-Hur, H., and Wilchek, M. (1987) *Anal. Biochem.* **161,** 123–131
14. Bayer, E. A., Ben-Hur, H., and Wilchek, M. (1988) *Anal. Biochem.* **170,** 271–281
15. Wilchek, M., and Bayer, E. A. (1988) *Anal. Biochem.* **171,** 1–32
16. Kamerling, J. P., and Vliegenthart, J. F. G. (1982) *Cell Biol. Monogr.* **10,** 95–125
17. Kamerling, J. P., Gerwig, G. J., Vliegenthart, J. F. G., and Clamp, J. R. (1975) *Biochem. J.* **151,** 491–495
18. Gerwig, G. J., Kamerling, J. P., and Vliegenthart, J. F. G. (1978) *Carbohydr. Res.* **62,** 349–357
19. Gerwig, G. J., Kamerling, J. P., and Vliegenthart, J. F. G. (1979) *Carbohydr. Res.* **77,** 1–7
20. Korrel, S. A. M., Clemetson, K. J., Van Halbeek, H., Kamerling, J. P., Sixma, J. J., and Vliegenthart, J. F. G. (1984) *Eur. J. Biochem.* **140,** 571–576
21. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956) *Anal. Chem.* **28,** 350–356
22. Waeghe, T. J., Darvill, A. G., McNeil, M., and Albersheim, P. (1983) *Carbohydr. Res.* **123,** 281–304
23. Jansson, P.-E., Kenne, L., Liedgren, H., Lindberg, B., and Lönngren, J. (1976) *Chem. Commun. Univ. Stockholm* **8,** 1–75
24. Turco, S. J. (1981) *Anal. Biochem.* **118,** 278–283
25. Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) *J. Chem. Phys.* **64,** 2229–2246
26. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113,** 967–974
27. Wagner, G., and Zuiderweg, E. R. P. (1976) *Biochem. Biophys. Res. Commun.* **113,** 854–860
28. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117,** 479–485
29. Bax, A., and Davis, D. G. (1985) *J. Magnetic Res.* **65,** 335–360
30. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magnetic Res.* **53,** 521–526
31. Davis, D. G., and Bax, A. (1985) *J. Am. Chem. Soc.* **107,** 2820–2821
32. Edwards, M. W., and Bax, A. (1986) *J. Am. Chem. Soc.* **108,** 918–923
33. Bax, A., and Davis, D. G. (1985) *J. Magnetic Res.* **63,** 207–213
34. Whitfield, D. M., Ruzicka, C. J., Carver, J. P., and Krepinsky, J. J. (1987) *Can. J. Chem.* **65,** 693–703
35. Coduti, P. L., and Bush, C. A. (1977) *Anal. Biochem.* **78,** 21–38

36. Gerwig, G. J., Kamerling, J. P., and Vliegenthart, J. F. G. (1984) *Carbohydr. Res.* **129,** 149–157
37. Kärkkäinen, J. (1970) *Carbohydr. Res.* **14,** 27–33
38. Lamed, R., Naimark, J., Morgenstern, E., and Bayer, E. A. (1987) *J. Bacteriol.* **169,** 3792–3800
39. Lamed, R., and Bayer, E. A. (1987) in *Biochemistry and Genetics of Cellulose Degradation* (Aubert, J. P., Beguin, P., and Millet, J., eds.) pp. 101–106, Academic Press, Orlando, FL
40. Barreto-Bergter, E. M., Travassos, L. R., and Gorin, P. A. J. (1980) *Carbohydr. Res.* **86,** 273–285
41. Lederkremer, R. M., Casal, O. L., Alves, M. J. M., and Colli, W. (1980) *FEBS Lett.* **116,** 25–29
42. Nakajima, T., Yoshida, M., Nakamura, M., Hiura, N., and Matsuda, K. (1984) *J. Biochem.* **96,** 1013–1020
43. Richards, J. C., Perry, M. B., and Kniskern, P. J. (1984) *Can. J. Biochem. Cell Biol.* **62,** 666–677
44. McIntire, F. C., Bush, C. A., Wu, S.-S., Li, S.-C., Li, Y.-T., McNeil, M., Tjoa, S. S., and Fennessey, P. V. (1987) *Carbohydr. Res.* **166,** 133–143
45. McNeil, M., Wallner, S. J., Hunter, S. W., and Brennan, P. J. (1987) *Carbohydr. Res.* **166,** 299–308
46. Jack, W., and Sturgeon, R. J. (1976) *Carbohydr. Res.* **49,** 335–340
47. Gander, J. E., Jentoft, N. H., Drewes, L. R., and Rick, P. D. (1974) *J. Biol. Chem.* **249,** 2063–2072
48. Mendelzon, D. H., and Parodi, A. J. (1986) *J. Biol. Chem.* **261,** 2129–2133
49. Groisman, J. F., and de Lederkremer, R. M. (1987) *Eur. J. Biochem.* **165,** 327–332
50. Moraes, C. T., Bosch, M., and Parodi, A. J. (1988) *Biochemistry* **27,** 1543–1549
51. Aspinall, G. O. (ed) (1983) *The Polysaccharides*, Vol. 2, Academic Press, Orlando, FL
52. Itasaka, O., and Hori, T. (1979) *J. Biochem.* (*Tokyo*) **85,** 1469–1481
53. Gerwig, G. J., Kamerling, J. P., Vliegenthart, J. F. G., Homan, W. L., van Egmond, P., and van den Ende, H. (1984) *Carbohydr. Res.* **127,** 245–251
54. Gastambide-Odier, M., and Villé, C. (1970) *Bull. Soc. Chim. Biol.* **52,** 679–693
55. Mehra, V., Brennan, P. J., Rada, E., Convit, J., and Bloom, B. R. (1984) *Nature* **308,** 194–196
56. Fournie, J.-J., Riviere, M., and Puzo, G. (1987) *Eur. J. Biochem.* **168,** 181–183
57. Lederkremer, G. Z., and Parodi, A. J. (1984) *J. Biol. Chem.* **259,** 12514–12518
58. Hall, R. L., Wood, E. J., Kamerling, J. P., Gerwig, G. J., and Vliegenthart, J. F. G. (1977) *Biochem. J.* **165,** 173–176
59. van Kuik, J. A., van Halbeek, H., Kamerling, J. P., and Vliegenthart, J. F. G. (1985) *J. Biol. Chem.* **260,** 13984–13988
60. Van Kuik, J. A., Sijbesma, R. P., Kamerling, J. P., Vliegenthart, J. F. G., and Wood, E. J. (1986) *Eur. J. Biochem.* **160,** 621–625
61. Van Kuik, J. A., Sijbesma, R. P., Kamerling, J. P., Vliegenthart, J. F. G., and Wood, E. J. (1987) *Eur. J. Biochem.* **169,** 399–411
62. Van Kuik, J. A. (1987) *Structural Determination of the Carbohydrate Chains from Arthropod and Mollusc Hemocyanin by Means of 500-MHz ^1H NMR Spectroscopy.* Ph.D thesis, University of Utrecht
63. Wood, E. J., and Chaplin, M. F. (1983) *Life Chem. Rep. Suppl.* **1,** 141–142
64. Paul, G., and Wieland, F. (1987) *J. Biol. Chem.* **262,** 9587–9593
65. Mayer, H., Framberg, K., and Weckesser, J. (1974) *Eur. J. Biochem.* **44,** 181–187
66. Matsubara, T., and Hayashi, A. (1978) *J. Biochem.* **83,** 1195–1197
67. Hori, T., Sugita, M., Ando, S., Kuwahara, M., Kumauchi, K., Sugie, E., and Itasaka, O. (1981) *J. Biol. Chem.* **256,** 10979–10985
68. Araki, S., Abe, S., Odani, S., Ando, S., Fujii, N., and Satake, M. (1987) *J. Biol. Chem.* **262,** 14141–14145