THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 268, No. 36, Issue of December 25, pp. 26956-26960, 1993
Printed in U.S.A.

The Nature of the Carbohydrate-Peptide Linkage Region in Glycoproteins from the Cellulosomes of *Clostridium thermocellum* and *Bacteroides cellulosolvens**

(Received for publication, May 5, 1993, and in revised form, August 16, 1993)

Gerrit J. Gerwig‡, Johannis P. Kamerling‡, Johannes F. G. Vliegenthart‡§, Ely Morag¶, Raphael Lamed¶, and Edward A. Bayer‖

From the ‡Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, P. O. Box 80.075, NL-3508 TB Utrecht, The Netherlands, the ¶Department of Biotechnology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel, and the ‖Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel

The cellulase complexes of two cellulolytic bacteria, *Clostridium thermocellum* and *Bacteroides cellulosolvens*, were subjected to extensive Pronase digestion. Glycopeptide fractions were isolated by gel permeation and fast protein liquid chromatography and analyzed by monosaccharide analysis, amino acid analysis, methylation analysis, and ^{1}H NMR spectroscopy. Alkaline borohydride-induced deglycosylation/amino acid conversion and periodate oxidation studies on the glycopeptide fraction of the *C. thermocellum* cellulosome demonstrated that the earlier established collection of carbohydrate moieties with 3-*O*-Me-D-Glc*p*NAc-α(1→2)-[D-Gal*p*-α(1→3)]-D-Gal*f*-α(1→2)-D-Gal (where 3-*O*-Me-D-Glc*p*NAc is 3-*O*-methyl-*N*-acetylglucopyranosamine, Gal*p* is galactopyranose, and Gal*f* is galactofuranose) as the major component, is *O*-linked to threonine via galactopyranose. Using the same approach for the glycopeptide fraction of the cellulase complex of *B. cellulosolvens*, it was found that the reported collection of carbohydrate moieties with D-Gal*f*-α(1→3)-D-Glc*p*NAc-α(1→2)-D-Gal*f*-α(1→2)-[D-Gal*f*-β(1→3)]-D-Gal as the major component, is *O*-linked mainly to threonine and partly to serine via galactopyranose. In both species, the hydroxyamino-acid-bound galactopyranose residue has probably an α-configuration. The carbohydrate chains appear as clusters located in highly Thr/Pro-rich peptide regions of the glycoproteins. The results are consistent with the notion that the glycosylation sites are localized in linker sequences which connect the various binding domains of the noncatalytic S1 subunit of the cellulosome.

During the course of our studies on the process of cellulose biodegradation, we have analyzed the primary structures of the covalently linked carbohydrate moieties in the multicomponent enzyme complexes of two different cellulolytic bacteria, namely, *Clostridium thermocellum* and *Bacteroides cellulosolvens* (1–3). The major components turned out to be very similar: [3-*O*-Me-]$_{+/-}$-D-Glc*p*NAc[1]-α(1→2)-[D-Gal*p*-α(1→3)]-D-Gal*f*-α(1→2)-D-Gal (58% of the carbohydrate chains) for *C. thermocellum*, and [D-Gal*f*-α(1→3)]$_{+/-}$-D-Glc*p*NAc-α(1→2)-D-Gal*f*-α(1→2)-[D-Gal*f*-β(1→3)]-D-Gal (70% of the carbohydrate chains) for *B. cellulosolvens* (where + means present; and − means not present). The minor components comprised partial structures of the major components, missing one or more terminal monosaccharide unit(s) (Table I).

Since for both species the aforementioned oligosaccharides could be released by alkaline borohydride treatment as alditols, *O*-glycosidic linkages to the peptide backbones were suggested. In this report, data are presented with respect to the nature of the monosaccharides and amino acids involved in the carbohydrate-peptide linkage and the nature of the glycosylated peptide backbone of these glycoproteins.

EXPERIMENTAL PROCEDURES

Isolation of Cellulase Complex—The extracellular cellulase complexes from *C. thermocellum* YS and *B. cellulosolvens* NRCC 2944 were isolated as reported elsewhere (4–6).

Pronase Digestion—A sample (50 mg/10 ml) of the desired cellulase complex was dissolved in 0.1 M Tris-HCl buffer, pH 7.6, containing 5 mM $CaCl_2$. Portions of 2 mg of Pronase (*Streptomyces griseus*, Boehringer Mannheim) were added at 0, 6, 24, and 48 h, and the incubation was carried out for 62 h at 40 °C under N_2. After heat inactivation of proteolytic activity (3 min at 100 °C), the solution was centrifuged, and the supernatant was passed through a Bio-Gel P-4 column (95 × 1.5 cm, 200–400 mesh, Bio-Rad) using water as eluent (15 ml/h). The eluate was monitored by refractive index detection and by hexose determination with a phenol/sulfuric acid assay (7). Carbohydrate-positive fractions were pooled and lyophilized. The residue was taken up in 5 ml of 0.1 M Tris-HCl buffer, pH 7.6, containing 5 mM $CaCl_2$, and the Pronase digestion/gel filtration procedure was repeated twice.

Fast Protein Liquid Chromatography—Glycopeptide material was fractionated on a Superose 12 column (30 × 1.4 cm, Pharmacia LKB Biotechnology Inc. FPLC system), using 0.1 M ammonium bicarbonate, pH 7.8, as eluent, and a flow rate of 0.5 ml/min with UV detection at 214 nm.

Monosaccharide Analysis—Monosaccharide composition and carbohydrate content determinations were carried out by gas-liquid chromatography on a capillary WCOT SE-30 fused silica column (25 m × 0.32 mm, Pierce Chemical Co.). Trimethylsilylated methyl glycosides were prepared by methanolysis (1.0 M methanolic HCl, 24 h, 85 °C), *N*-reacetylation, and trimethylsilylation (8).

Amino Acid Analysis—Samples (50–100 μg) were hydrolyzed with 6.0 M HCl for 22 h at 110 °C under N_2. Amino acid analyses were performed on an LKB 4151 Alpha Plus amino acid analyzer using a five-buffer lithium citrate system (9).

Methylation Analysis—Permethylation was accomplished essentially as previously described (10). Each glycopeptide sample (50 μg) was

* 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 should be addressed: Bijvoet Center, Dept. of Bio-Organic Chemistry, Utrecht University, P. O. Box 80.075, NL-3508 TB Utrecht, The Netherlands. Tel.: 31-30-532168; Fax: 31-30-540980.

[1] The abbreviations used are: 3-*O*-Me-Glc*p*NAc, 3-*O*-methyl-*N*-acetylglucopyranosamine; Gal*p*, galactopyranose; Gal*f*, galactofuranose; FPLC, fast protein liquid chromatography; GPct, glycopeptide fraction from *C. thermocellum* cellulosome; GPbc, glycopeptide fraction from *B. cellulosolvens* cellulosome.

TABLE I
O-Linked glycoprotein carbohydrates in the cellulase complexes of C. thermocellum and B. cellulosolvens

Structure	Molar percentage
C. thermocellum	
3-*O*-Me-D-Glc*p*NAc-α(1→2)-[D-Gal*p*-α(1→3)]-D-Gal*f*-α(1→2)-D-Gal	45
D-Glc*p*NAc-α(1→2)-[D-Gal*p*-α(1→3)]-D-Gal*f*-α(1→2)-D-Gal	13
3-*O*-Me-D-Glc*p*NAc-α(1→2)-D-Gal*f*-α(1→2)-D-Gal	5
D-Glc*p*NAc-α(1→2)-D-Gal*f*-α(1→2)-D-Gal	<5
D-Gal*f*-α(1→2)-D-Gal	12
D-Gal*p*-β(1→4)-D-Gal*p*[a]	~10
D-Gal	12
B. cellulosolvens	
D-Gal*f*-α(1→3)-D-Glc*p*NAc-α(1→2)-D-Gal*f*-α(1→2)-[D-Gal*f*-β(1→3)]-D-Gal	24
D-Glc*p*NAc-α(1→2)-D-Gal*f*-α(1→2)-[D-Gal*f*-β(1→3)]-D-Gal	46
D-Glc*p*NAc-α(1→2)-D-Gal*f*-α(1→2)-D-Gal	8
D-Gal*f*-β(1→3)-D-Gal D-Gal*f*-α(1→2)-D-Gal	15
D-Gal	7

[a] This disaccharide was found only in the cell-free cellulosome of *C. thermocellum* YS and is split off only after a second alkaline borohydride treatment (1).

dissolved in 0.1 ml of methyl sulfoxide using ultrasonication for 20 min at 40 °C. Fresh finely powdered NaOH (~1.5 mg) was added under an inert atmosphere, and the mixture was sonicated for 1 h at room temperature. After cooling to 0 °C, 0.2 ml of methyl iodide was added, and the sonication was continued for 1 h at a temperature not exceeding 20 °C. Then, 1 ml of water containing a few crystals of sodium thiosulfate was added, and the methylated product was isolated by extraction with chloroform (3 × 0.5 ml). The organic phase was washed with water (5 × 0.5 ml) and evaporated under a stream of N_2. The residue was hydrolyzed with 4 M trifluoroacetic acid (0.2 ml, 4 h, 100 °C), and, after co-evaporation with methanol, the obtained mixture of partially methylated monosaccharides was reduced with 1 mg of NaB^2H_4 in 0.1 ml of water (2 h at room temperature). After addition of acetic acid (~10 μl), the sample was acetylated with 0.1 ml of acetic anhydride/pyridine (1:1, v/v) for 30 min at 100 °C. The solvent was evaporated in a gentle stream of N_2 with intermediate addition of toluene, and the residue was dissolved in dichloromethane. The partially methylated alditol acetates were analyzed by gas-liquid chromatography/mass spectrometry using a Carlo Erba GC/Kratos MS 80/Kratos DS 55 system as described elsewhere (1).

Deglycosylation of Glycopeptides—Glycopeptide samples (100 μg) were treated with 1 ml of 0.1 M NaOH, containing 1 M $NaBH_4$ for 48 h or 62 h at 37 °C under N_2. The solution was adjusted to pH 6.0 with 4 M acetic acid and concentrated. After removal of boric acid by co-evaporation with methanol, the residue was fractionated on a Bio-Gel P-2 column (45 × 1 cm), using water as eluent, and the void volume fraction containing the (glyco)peptide material was lyophilized.

O-Acyl Group Determination—A colorimetric *O*-acyl group microassay was performed essentially as previously described (11). Glycopeptide samples (50 μg/40 μl water) were added to 40 μl of alkaline hydroxylamine (0.35 M hydroxylamine hydrochloride, 2.5 M ammonium hydroxide, 1:1, v/v). In each case the mixture was shaken and allowed to react for 10 min at room temperature. Then, 40 μl of 0.74 M perchloric acid was added and, after shaking, 20 μl of 0.07 M ferric perchlorate in 0.4 M perchloric acid. The optical density was determined at 520 nm. Galactose penta-acetate (10 μg/40 μl water) and ethyl acetate (5 mM in methanol/water (1:1, v/v)) were used as positive controls.

Periodate Oxidation—Glycopeptide samples (100 μg) were dissolved in 0.5 ml of 0.1 M sodium acetate buffer, pH 5.5, containing 8 mM $NaIO_4$, and kept for 48 h at 5 °C in the dark. The excess of periodate was destroyed with 2 μl of ethylene glycol, and after 16 h at 5 °C, 0.5 ml of 0.2 M NaOH and 10 mg $NaBH_4$ were added. The resulting solution was kept for 6 h at room temperature, then adjusted to pH 6.0 with 4 M acetic acid, and concentrated. After removal of boric acid by co-evaporation with methanol, the residue was desalted on a Bio-Gel P-2 column (45 × 1 cm), using water as eluent, followed by lyophilization.

^1H NMR Spectroscopy—Glycopeptide samples (100–500 μg) were repeatedly exchanged in ^2H_2O (99.96 atom % ^2H, MSD Isotopes) with intermediate lyophilization. Resolution-enhanced 300- and 500-MHz ^1H NMR spectra were recorded as described elsewhere (1–3).

RESULTS

The cellulosomes of both *C. thermocellum* and *B. cellulosolvens* are very stable multicomponent enzyme complexes. So far, it has not been possible to separate under mild conditions the complexes into their component parts, due to the cohesive nature and extensive aggregation of the cellulosome (6, 12–16). In this regard, rather harsh conditions, *i.e.* SDS treatment at elevated temperature combined with electro-elution from SDS-polyacrylamide gel electrophoresis gels, were required to isolate minor amounts of the S1 glycoprotein subunit from the cellulosome of *C. thermocellum* (15). In the case of *B. cellulosolvens*, all efforts failed to obtain purified samples of the glycoprotein component from its cellulosome-like entity. Facing such difficulties in purifying individual cellulosomal components, it may be evident that the aimed study on the nature of the carbohydrate-peptide region in the cellulosomal glycoproteins could not be carried out on the level of specific glycopeptides generated from purified glycoproteins. Therefore, it was decided to use the complete cellulase complex of each bacterial species for the preparation of glycopeptide material.

C. thermocellum—A triple Pronase digestion of the *C. thermocellum* cellulase complex (cellulosome), with intermediate fractionation on Bio-Gel P-4, yielded a carbohydrate-containing product (GPct), eluting in the void volume, which was not further degradable with Pronase (carbohydrate and amino acid analyses, data not shown). Fractionation of GPct by FPLC on Superose 12 afforded a major glycopeptide fraction GPct1 (Fig. 1*A*). It should be noted that the native cellulosomes elute as a sharp peak in the void volume on the same column.

Monosaccharide analysis of GPct1 (Table II), having a carbohydrate content of 70% (w/w), revealed the presence of Gal, 3-*O*-Me-GlcNAc and GlcNAc, in a molar ratio essentially identical to that of the cellulosome and of the S1 subunit, suggesting that the earlier reported oligosaccharide heterogeneity (1, 2) is also present in the isolated glycopeptide fraction. In Table III, the amino acid analysis data of GPct1 (Fig. 2*A*, *a*) are shown, together with the amino acid composition of the cellulosome and the S1 subunit glycoprotein. Threonine (37%, w/w) was the major constituent of GPct1, together with proline (25%, w/w). Serine was only present in a minor amount (4%, w/w).

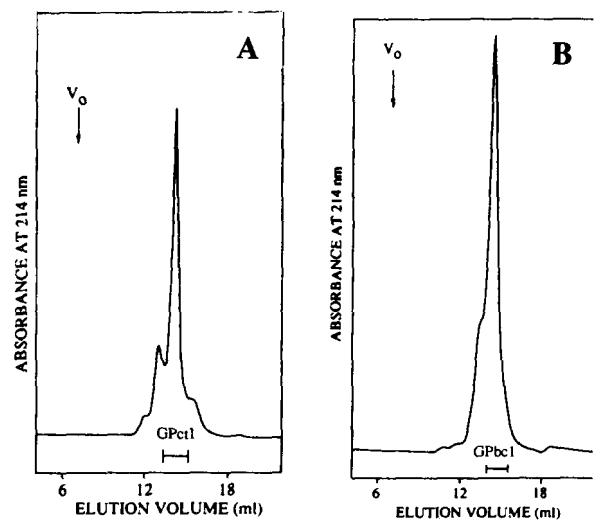

FIG. 1. **FPLC elution profile on Superose 12 of glycopeptide fractions GPct (*A*) and GPbc (*B*).** The column (30 × 1.4 cm) was eluted with 0.1 M ammonium bicarbonate, pH 7.8, at a flow rate of 0.5 ml/min. The eluate was monitored at 214 nm. In each case, the fraction indicated was isolated.

TABLE II
Carbohydrate content and composition of the C. thermocellum and B. cellulosolvens cellulase complexes and derived samples

Monosaccharide	Molar ratio[a]					
	C. thermocellum			*B. cellulosolvens*		
	Cellulase complex	S1	GPct1	Cellulase complex	GPbc	GPbc1
Gal	6.3	6.0	6.5	4.4	4.1	4.0
3-*O*-Me-GlcNAc	1.0	1.0	1.0			
GlcNAc	0.3	0.5	0.3	1.0	1.0	1.0
Content (%, w/w)	7	15	70	4.5	ND[b]	65

[a] 3-*O*-Me-GlcNAc (*C. thermocellum*) or GlcNAc (*B. cellulosolvens*) taken as 1.0.
[b] Not determined.

TABLE III
Amino acid composition of the C. thermocellum and B. cellulosolvens cellulase complexes and derived samples (residues/100)

Amino acid	*C. thermocellum*				*B. cellulosolvens*		
	Cellulase complex	S1	GPct1	GPct1/62[a]	Cellulase complex	GPbc1	GPbc1/48[b]
Asx	13.0	13.3	11.0	18.3	12.3	5.7	6.8
Thr	7.8	9.0	37.5	9.1	7.2	26.3	16.6
Ser	5.6	5.8	4.0	5.0	6.3	8.1	3.8
Glx	9.7	8.1	4.1	7.4	10.3	2.6	3.7
Pro	8.2	7.5	24.6	24.2	5.1	23.5	25.4
Gly	9.1	10.3	1.2	6.0	8.6	4.5	4.2
Ala	7.6	7.9	4.1	6.7	8.3	5.2	10.3
Val	7.7	10.5	6.9	5.2	7.1	6.3	6.1
Met	0.5	0.1			1.4		
Ile	6.1	6.7	1.6	1.2	6.2	0.8	1.0
Leu	6.1	4.3	+	0.5	6.7	1.8	1.8
Tyr	4.2	2.6			3.8	0.5	0.4
Phe	4.5	5.3			3.6		
Lys	5.7	5.5	0.7	1.1	8.5	6.4	6.3
His	1.1	0.3			1.0		
Arg	3.0	2.2			2.7		
Trp	ND[d]	ND	ND	ND	ND	ND	ND
Cys	ND	ND	ND	ND	ND	ND	ND
α-Aminobutyric acid + $GlcNH_2$[c]	0.2	0.6	4.3	15.3	0.4	8.3	13.6

[a] GPct1/62 means GPct1 treated for 62 h with alkaline borohydride.
[b] GPbc1/48 means GPbc1 treated for 48 h with alkaline borohydride.
[c] α-Aminobutyric acid and $GlcNH_2$ elute in a common peak and the recorded values reflect the uncorrected sum of the two residues.
[d] Not determined.

To estimate the amino acid(s) involved in the glycan-peptide linkage, GPct1 was subjected to alkaline borohydride treatment. Usually, deglycosylations are performed for 48 h, but even after 62 h the deglycosylation was found to be incomplete, as monitored by monosaccharide analysis. After alkaline borohydride treatment for 62 h, the corresponding Bio-Gel P-2 void volume peptide fraction had a carbohydrate content of 20% (w/w), indicating that deglycosylation had proceeded to about 70% completion. As compared to the monosaccharide composition of GPct1, the partially deglycosylated fraction, denoted GPct1/62, showed a relatively increased amount of galactose. This observation fits the earlier finding (1) that, besides a nonpreferential release of the collection of carbohydrate chains discussed in this paper, the unique carbohydrate chain Gal*p*-β(1→4)-Gal*p* (see Table I) is not released at all during a single alkaline borohydride treatment. Amino acid analysis of GPct1/62 (Fig. 2*A*, *b*) showed that more than 70% of Thr was converted into α-aminobutyric acid (Table III). The other amino acids did not show significant changes, and evidently no serine was converted into alanine. From the carbohydrate and amino acid analysis data, it could be calculated that the carbohydrate part and Thr occur in a near-equimolar ratio. Based on the various findings, it can be concluded that in the native material the carbohydrate part is glycosidically linked to the hydroxyl function of threonine.

To get information on the ring form of the galactose residue involved in the linkage to Thr (see Fig. 3*A*, major carbohydrate moiety), GPct1 was subjected to periodate oxidation, followed by reduction. Monosaccharide analysis of the degraded material showed the presence of only arabinose and 3-*O*-Me-GlcNAc (Fig. 4*A*), but because of the heterogeneity of the carbohydrate chains linked to Thr, quantification is not reliable. Therefore, at this stage a discrimination between Gal*p* and Gal*f* was not possible. Subsequently, the degraded material was treated with alkaline borohydride (1). Monosaccharide analysis of the released carbohydrate revealed the presence of arabinose and 3-*O*-Me-GlcNAc, but no arabinitol could be detected, indicating that the Thr-linked galactose residue in the native material has the pyranose ring form. Methylation analysis of GPct1 demonstrated that, besides 2,3,4,6-tetra-*O*-methyl-1,5-di-*O*-acetylgalactitol, 5,6-di-*O*-methyl-1,2,3,4-tetra-*O*-acetylgalactitol, and 2-(*N*-methyl)acetamido-2-deoxy-3,4,6-tri-*O*-methyl-1,5-di-*O*-acetylglucitol, also 3,4,6-tri-*O*-methyl-1,2,5-tri-*O*-acetylgalactitol was present, thus supporting this conclusion.

The 300-MHz ^1H NMR spectrum at 300 K of GPct1 showed the presence of NAc and OCH_3 signals at δ 2.09 and δ 3.53, respectively. In spite of the appearance of broad resonances in the anomeric region, a relatively intense signal at δ 5.08 was observed, and it is suggested that this signal stems from the H-1 atom of threonine-bound α-galactopyranose. Temperature elevation to 343 K did not give any spectral improvement. No indications were obtained for the presence of *O*-acyl groups, in agreement with colorimetric studies (11, 17).

B. cellulosolvens—The cellulase complex of *B. cellulosolvens* was subjected to a triple Pronase digestion with intermediate

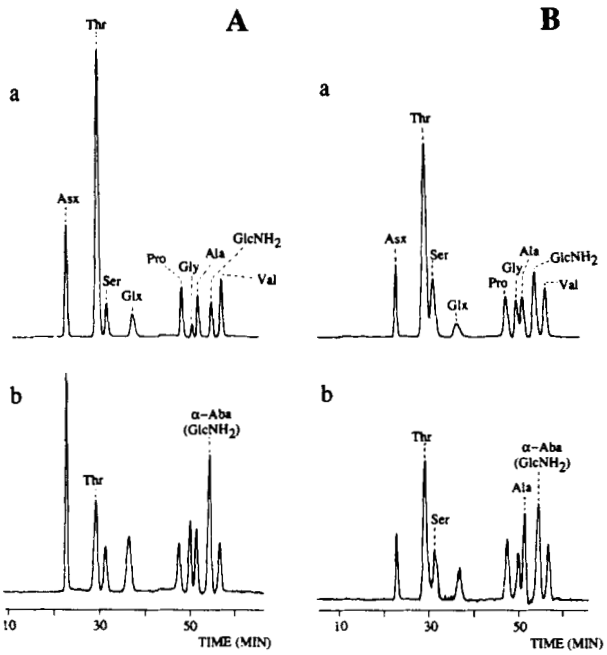

FIG. 2. **Relevant part of the amino acid analysis of (partially) deglycosylated glycopeptide fractions derived from *C. thermocellum* (*A*) and *B. cellulosolvens* (*B*) cellulase complex.** *A*, fraction GPct1 (*a*); GPct1/62, fraction GPct1, partially deglycosylated by alkaline borohydride treatment for 62 h (*b*); *B*, fraction GPbc1 (*a*); GPbc1/48, fraction GPbc1, partially deglycosylated by alkaline borohydride treatment for 48 h (*b*). Note that α-aminobutyric acid (*α-Aba*) and glucosamine ($GlcNH_2$) have identical elution volumes under the conditions used.

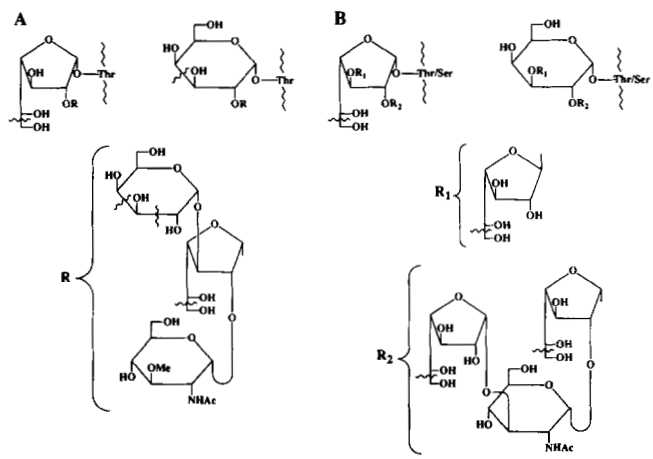

FIG. 3. **Schematic presentation of the *O*-linked tetrasaccharide in *C. thermocellum* cellulase complex (*A*) and pentasaccharide in *B. cellulosolvens* cellulase complex (*B*) to illustrate the effect of periodate oxidation in case an amino-acid-bound galactofuranose or a galactopyranose is expected.** (~) Linkages sensitive to periodate oxidation. In both species, a pyranose ring was demonstrated (see text).

fractionation on Bio-Gel P-4, ultimately affording a carbohydrate-positive void volume fraction GPbc, which was not further degradable with Pronase (carbohydrate and amino acid analyses, data not shown). Monosaccharide analysis of GPbc showed a molar ratio for Gal and GlcNAc, quite similar to that of the cellulase complex (Table II). GPbc was subfractionated by FPLC on Superose 12 (Fig. 1*B*), yielding fraction GPbc1.

Monosaccharide analysis of GPbc1 (Table II), having a carbohydrate content of 65% (w/w), showed a Gal/GlcNAc composition, essentially identical to that of the cellulase complex in this bacterium. Consequently, in parallel to the results obtained for *C. thermocellum*, this suggests that the earlier reported oligosaccharide heterogeneity (3) also occurs in the isolated glycopeptide fraction GPbc1. In Table III, the amino acid analysis data of GPbc1 (Fig. 2*B*, *a*) are shown, together with the amino acid composition of the cellulase complex. Threonine (26%, w/w) was the major constituent of GPbc1, together with proline (24%, w/w) and serine (8%, w/w). Similar to the case of glycopeptide fraction GPct1 from the cellulosome of *C. thermocellum*, the high proline content of the glycopeptides should be noted.

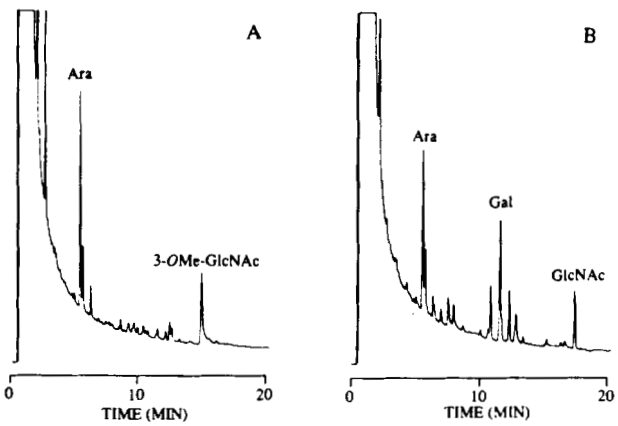

FIG. 4. **Monosaccharide analysis of periodate treated/reduced *C. thermocellum* glycopeptide GPct1 (*A*) and *B. cellulosolvens* glycopeptide GPbc1 (*B*).**

To obtain direct evidence for the amino acid(s) involved in the carbohydrate-peptide linkage, the glycopeptide fraction GPbc1 was subjected to alkaline borohydride treatment (48 h). Monosaccharide analysis of the Bio-Gel P-2 void volume peptide fraction showed that the deglycosylation had proceeded to about 50% completion. The monosaccharide composition of the partially deglycosylated glycopeptide fraction, denoted GPbc1/48, was similar to that of the starting material GPbc1, indicating a non-preferential release of the carbohydrate chains during alkaline borohydride treatment. Amino acid analysis (Fig. 2*B*, *b*, Table III) demonstrated that about 40% of threonine was converted into α-aminobutyric acid, and about 50% of serine into alanine, indicating that both hydroxyamino acids are glycosylated. In view of the possible incompleteness of the carbohydrate release, a definite conclusion cannot be drawn with respect to the percentage of glycosylation. Note that threonine is present at levels 3–4-fold those of serine. Other amino acids did not show significant changes after partial deglycosylation.

To get definite information on the ring form of the galactose residue involved in the linkage to Thr/Ser (see Fig. 3*B*, major carbohydrate moiety), GPbc1 was subjected to periodate oxidation, followed by reduction. Monosaccharide analysis of the degraded material showed arabinose, Gal, and GlcNAc (Fig. 4*B*), indicating that the Thr/Ser-linked galactose residue in the native material has the pyranose ring form. Methylation analysis of GPbc1 demonstrated that besides 2,3,5,6-tetra-*O*-methyl-1,4-di-*O*-acetylgalactitol, 3,5,6-tri-*O*-methyl-1,2,4-tri-*O*-acetylgalactitol, 2-(*N*-methyl)acetamido-2-deoxy-4,6-di-*O*-methyl-1,3,5-tri-*O*-acetylglucitol, and 2-(*N*-methyl)acetamido-2-deoxy-3,4,6-tri-*O*-methyl-1,5-di-*O*-acetylglucitol, also 4,6-di-*O*-methyl-1,2,3,5-tetra-*O*-acetylgalactitol was present, supporting this conclusion.

The 500-MHz ^1H NMR spectrum at 300 K of GPbc1 showed in the anomeric region a number of broad resonances, which could be related with the ^1H NMR data of the isolated oligosaccharide-alditols previously reported (3). In addition, an anomeric signal at δ 4.97 was observed with a coupling constant of ~3.5 Hz. The homonuclear Hartmann-Hahn cross-peaks with

this signal suggest a galactose residue, and it is proposed that this residue represents the Thr/Ser-bound galactopyranose in α-configuration. *O*-Acyl groups could not be detected, supporting data from colorimetric determinations (11, 17).

DISCUSSION

Previously, evidence was reported which indicated the presence of glycosylated polypeptide chains in the cellulase complexes of *C. thermocellum* and *B. cellulosolvens* (5, 6). In recent structural studies (1–3) it was shown that the oligosaccharides can be released by alkaline borohydride treatment of the cellulase complexes but the glycoprotein-containing material turned out to be resistant to treatment with peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F (EC 3.5.1.52). Based on these findings, it was suggested that the carbohydrates were originally linked to the polypeptide backbone via an alkali-labile *O*-glycosidic bond (1–3). Furthermore, it was found that the series of oligosaccharide-alditols, derived from the cellulase complexes of the two bacterial species, have a striking similarity with respect to the core structure (3).

Using a glycopeptide fraction, it has now been demonstrated that the oligosaccharides in the cellulosome of *C. thermocellum* are *O*-glycosidically linked via galactopyranose to threonine. In a similar way, investigation of a glycopeptide fraction from the cellulase complex of *B. cellulosolvens* has shown that the oligosaccharides are *O*-glycosidically bound via galactopyranose mainly to threonine and to a small extent to serine. In both bacterial species the hydroxyamino acid-bound galactopyranose has probably α-configuration. The difference in chemical shifts between the anomeric signals of the hydroxyamino acid-bound Gal residue in the *C. thermocellum* and *B. cellulosolvens* preparations is probably due to an influence of the peptide part of the species-specific glycopeptide, as has been observed earlier for the anomeric signal of the asparagine-bound GlcNAc residue in *N*-linked glycopeptides (18).

A Gal-Ser(Thr) linkage has first been reported to occur in peptides from the cuticle collagen of the earthworm *Lubricus terrestris* (19). Also, in the plant glycoproteins extensin (20) and potato lectin (21) this glycan-peptide linkage was found. Only for potato lectin the galactose residue was demonstrated to occur in the α-pyranose ring form, based on its sensitivity to exo-α-galactosidase. The S-layer glycoproteins of the archaebacteria *Halobacterium halobium* (22) and *Haloferax Volcanii* (23) contain Glcp-(1→2)-Galp linked to threonine residues. However, the anomeric configuration has not been clarified.

As evident from monosaccharide analysis, the *C. thermocellum* glycopeptide fraction corresponds with the S1 glycoprotein subunit. The glycopeptide fraction contained in total more than 60% (w/w) Thr and Pro, suggesting that the O-linked carbohydrate chains in the 210-kDa glycoprotein may appear as a cluster in a highly Thr/Pro rich peptide region in which the majority of the Thr residues appear to be glycosylated. The glycoprotein material of *B. cellulosolvens* also seems to have highly glycosylated Thr/Pro peptide regions. The occurrence of Thr/Pro regions have also been reported for a number of cellulases and xylanases from *C. thermocellum* and other bacterial species. In the case of *Cellulomonas fimi*, two extracellular endoglucanases turned out to be glycoproteins, having Thr/Pro rich regions for which an *O*-glycosylation was suggested (24–26). Likewise, endoglucanases produced by *Bacteroides succinogenes* (27), *Thermonospora fusca* (28), and *Pseudomonas fluorescens* ssp. *cellulosa* (29) also contain a domain rich in Pro and *O*-glycosylated hydroxyamino acids. For endoglucanases from *C. thermocellum*, obtained via gene expression in *Escherichia coli*, the presence of a conserved Thr/Pro region was demonstrated by sequence analysis (30, 31).

Thr/Pro-containing peptide regions generally serve as linker sequences which connect discrete cellulose-binding domains and catalytic domains of glucanases (32). Recent cloning of portions of the noncatalytic S1 subunit from *C. thermocellum* has demonstrated similar Thr/Pro-rich regions which interlink a principal cellulose-binding domain and several subunit-binding domains (33, 34). The results presented here, together with the sequence data (33, 34), indicate that these linker sequences are the major sites of glycosylation in the cellulosome. The striking similarities in the nature and position of the carbohydrate moieties in two such unrelated, genetically distinct cellulolytic microorganisms may imply similar structural elements in the composition of the carrier protein and a common important functional role in cellulolysis.

Acknowledgment—We thank H. J. L. Ravestein (Faculty of Biology, Instrumental Analysis, Utrecht University) for performing the amino acid analyses.

REFERENCES

1. Gerwig, G. J., de Waard, P., Kamerling, J. P., Vliegenthart, J. F. G., Morgenstern, E., Lamed, R., and Bayer, E. A. (1989) *J. Biol. Chem.* **264,** 1027–1035
2. Gerwig, G. J., Kamerling, J. P., Vliegenthart, J. F. G., Morag (Morgenstern), E., Lamed, R., and Bayer, E. A. (1991) *Eur. J. Biochem.* **196,** 115–122
3. Gerwig, G. J., Kamerling, J. P., Vliegenthart, J. F. G., Morag (Morgenstern), E., Lamed, R., and Bayer, E. A. (1992) *Eur. J. Biochem.* **205,** 799–808
4. Lamed, R., Setter, E., Kenig, R., and Bayer, E. A. (1983) *Biotechnol. Bioeng. Symp.* **13,** 163–181
5. Lamed, R., Naimark, J., Morgenstern, E., and Bayer, E. A. (1987) *J. Bacteriol.* **169,** 3792–3800
6. Lamed, R., Morag (Morgenstern), E., Mor-Yosef, O., and Bayer, E. A. (1991) *Current Microbiol.* **22,** 27–33
7. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956) *Anal. Chem.* **28,** 350–356
8. Kamerling, J. P., and Vliegenthart, J. F. G. (1989) in *Mass Spectrometry* (Lawson, A. M., ed) pp. 176–263, Walter de Gruyter, New York
9. De Baaij, J. A., Janssen, F. W., and Voortman, G. (1986) *Sci. Tools* **33,** 17–32
10. Ciucanu, I., and Kerek, F. (1984) *Carbohydr. Res.* **131,** 209–217
11. Schauer, R. (1978) *Methods Enzymol.* **50C,** 64–89
12. Bayer, E. A., Setter, E., and Lamed, R. (1985) *J. Bacteriol.* **163,** 552–559
13. Murray, W. D., Sowden, L. C., and Colvin, J. R. (1986) *Lett. Appl. Microbiol.* **3,** 69–72
14. Hon-nami, K., Coughlan, M. P., Hon-Nami, H., and Ljungdahl, L. G. (1986) *Arch. Microbiol.* **145,** 13–19
15. Lamed, R., and Bayer, E. A. (1988) *Adv. Appl. Microbiol.* **33,** 1–46
16. Lamed, R., and Bayer, E. A. (1988) in *Biochemistry and Genetics of Cellulose Degradation* (Aubert, J. P., Beguin, P., and Millet, J., eds) pp. 101–116, Academic Press, Orlando, FL
17. Hestrin, S. (1949) *J. Biol. Chem.* **180,** 249–261
18. Vliegenthart, J. F. G., Dorland, L., and Van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem.* **41,** 209–374
19. Muir, L., and Lee, Y. C. (1970) *J. Biol. Chem.* **245,** 502–509
20. Lamport, D. T. A., Katona, L., and Roerig, S. (1973) *Biochem. J.* **133,** 125–131
21. Allen, A. K., Desai, N. N., Neuberger, A., and Creeth, J. M. (1978) *Biochem. J.* **171,** 665–674
22. Lechner, J., and Wieland, F. (1989) *Annu. Rev. Biochem.* **58,** 173–194
23. Sumper, M., Berg, E., Mengele, R., and Strobel, I. (1990) *J. Bacteriol.* **172,** 7111–7118
24. McGavin, M., and Forsberg, C. W. (1989) *J. Bacteriol.* **171,** 3310–3315
25. Ghangas, G. S., and Wilson, D. B. (1988) *Appl. Environ. Microbiol.* **54,** 2521–2526
26. Gilbert, H. J., Hall, J., Hazlewood, G. P., and Ferreira, L. M. A. (1990) *Mol. Microbiol.* **4,** 759–767
27. Gilkes, N. R., Warren, R. A. J., Miller, R. C., Jr., and Kilburn, D. G. (1988) *J. Biol. Chem.* **263,** 10401–10407
28. Warren, R. A. J., Beck, C. F., Gilkes, N. R., Kilburn, D. G., Langsford, M. L., Miller, R. C., Jr., O'Neill, G. P., Scheufens, M., and Wong, W. K. K. (1986) *Proteins* **1,** 335–341
29. O'Neill, G. P., Goh, S. H., Warren, R. A. J., Kilburn, D. G., and Miller, R. C., Jr. (1986) *Gene* **44,** 325–330
30. Beguin, P., Cornet, P., and Aubert, J.-P. (1985) *J. Bacteriol.* **162,** 102–105
31. Grepinet, O., and Beguin, P. (1986) *Nucleic Acids Res.* **14,** 1791–1799
32. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. J. (1991) *Microbiol. Rev.* **55,** 303–315
33. Fujino, T., Beguin, P., and Aubert, J.-P. (1992) *FEMS Microbiol. Lett.* **94,** 165–170
34. Poole, D. M., Morag, E., Lamed, R., Bayer, E. A., Hazlewood, G. P., and Gilbert, H. J. (1993) *FEMS Microbiol. Lett.*, in press