JOURNAL OF BACTERIOLOGY, May 1995, p. 2451–2459 Vol. 177, No. 9
0021-9193/95/$04.00+0

OlpB, a New Outer Layer Protein of *Clostridium thermocellum*, and Binding of Its S-Layer-Like Domains to Components of the Cell Envelope

MARC LEMAIRE,[1] HÉLÈNE OHAYON,[2] PIERRE GOUNON,[2]
TSUCHIYOSHI FUJINO,[1]† AND PIERRE BÉGUIN[1]*

Unité de Physiologie Cellulaire and URA 1300 CNRS, Département des Biotechnologies,[1] and Station de Microscopie Electronique,[2] Institut Pasteur, 75724 Paris Cedex 15, France

Received 23 December 1994/Accepted 22 February 1995

Several proteins of *Clostridium thermocellum* possess a C-terminal triplicated sequence related to bacterial cell surface proteins. This sequence was named the SLH domain (for S-layer homology), and it was proposed that it might serve to anchor proteins to the cell surface (A. Lupas, H. Engelhardt, J. Peters, U. Santarius, S. Volker, and W. Baumeister, J. Bacteriol. 176:1224–1233, 1994). This hypothesis was investigated by using the SLH-containing protein ORF1p from *C. thermocellum* as a model. Subcellular fractionation, immunoblotting, and electron microscopy of immunocytochemically labeled cells indicated that ORF1p was located on the surface of *C. thermocellum*. To detect *C. thermocellum* components interacting with the SLH domains of ORF1p, a probe was constructed by grafting these domains on the C terminus of the MalE protein of *Escherichia coli*. The SLH domains conferred on the chimeric protein (MalE-ORF1p-C) the ability to bind noncovalently to the peptidoglycan of *C. thermocellum*. In addition, ^{125}I-labeled MalE-ORF1p-C was shown to bind to SLH-bearing proteins transferred onto nitrocellulose, and to a 26- to 28-kDa component of the cell envelope. These results agree with the hypothesis that SLH domains contribute to the binding of exocellular proteins to the cell surface of bacteria. The gene carrying ORF1 and its product, ORF1p, are renamed *olpB* and OlpB (for outer layer protein B), respectively.

Clostridium thermocellum, a gram-positive thermophilic anaerobic bacterium, produces a highly active cellulase system whose cellulolytic components form a high-molecular-weight complex, termed cellulosome (21). Previous studies showed that these components are organized around a scaffolding protein, named CipA, which promotes binding of the cellulosome to cellulose (10, 13, 27, 33, 37). The cellulosome is located in protuberances present on the cell surface and is subsequently released into the culture medium. The polypeptide sequences deduced from the genes encoding cellulosome components comprise typical signal peptides. This suggests that the components are secreted individually and that the cellulosome is assembled on the cell surface (4). It is therefore of interest to identify components of the cell envelope of *C. thermocellum* and to characterize their properties.

Sequencing of the DNA downstream from *cipA* revealed three open reading frames (ORFs), named ORF1, ORF2, and ORF3 (11). The gene carrying ORF3 is now termed *olpA* (31). The polypeptides encoded by these ORFs contain a C-terminal domain composed of three repeats of about 65 amino acids, which are similar to segments present in the S-layer proteins of several bacteria, e.g., *Acetogenium kivui* (GenBank accession no. M31069), *Bacillus sphaericus* (6), *Thermus thermophilus* (GenBank accession no. X57333), and *Bacillus brevis* 47 (MWP, GenBank accession no. M19115), and in the Ompα protein of *Thermotoga maritima* (GenBank accession no. X68276). This similarity led Lupas et al. (24) to term these segments SLH domains (for S-layer homology). Besides S-layer proteins, SLH domains also occur in several extracellular enzymes acting on carbohydrate polymers, e.g., a xylanase from *C. thermocellum* (GenBank accession no. M67438), an alkaline cellulase from *Bacillus* sp. strain KSM-635 (GenBank accession no. M27420), an endoxylanase from *Thermoanaerobacter saccharolyticum* (GenBank accession no. M97882), an α-amylase-pullulanase from *Bacillus* sp. strain XAL601 (GenBank accession no. D28467), and a pullulanase of *Thermoanaerobacterium thermosulfurigenes* EM1 (25). In addition, the OlpA protein of *C. thermocellum* carries an N-terminal domain (cohesin domain) which binds the anchoring determinant (dockerin domain) common to cellulosomal β-glycanases of *C. thermocellum* (10, 32). In several organisms, cell-bound enzymes are known to participate in the degradation of insoluble carbohydrates (16, 20, 22), and indeed, OlpA and the pullulanase have been located on the cell surface of *C. thermocellum* and *T. thermosulfurigenes* EM1, respectively (25, 31). These data suggest that SLH domains might be responsible for the anchoring of these proteins to the bacterial cell surface.

To test this hypothesis, the location of *C. thermocellum* ORF1p, the protein encoded by ORF1 (11), was investigated. The N-terminal region and the C-terminal SLH domains of ORF1p were fused with the *Escherichia coli* maltose-binding protein MalE. The resulting fusion proteins, termed MalE-ORF1p-N and MalE-ORF1p-C, respectively, were purified, and antibodies specific for each of the two regions of ORF1p were prepared. These antibodies were used to detect the presence of ORF1p in *C. thermocellum* by Western blotting (immunoblotting) and to localize it by electron microscopy of immunochemically labeled cells. In addition, MalE-ORF1p-C was used as a probe to detect components of the *C. thermocellum* cell envelope interacting with the SLH domains of ORF1p.

* Corresponding author. Mailing address: Unité de Physiologie Cellulaire, Département des Biotechnologies, Institut Pasteur, 28, rue du Dr. Roux, 75724 Paris Cedex 15, France. Phone: 33 1 45 68 88 19. Fax: 33 1 45 68 87 90. Electronic mail address: beguin@pasteur.fr.

† Permanent address: Nagoya Seiraku Co. Ltd., Nagoya 468, Japan.

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Relevant features	Source or reference
Strains		
Escherichia coli	TG1 [Δ(*lac-pro*) *thi supE hsdD5*/F′ *tra-36 proA*$^+$ *B*$^+$ *lacI*^{q} *lacZ*ΔM15]	14
Clostridium thermocellum	NCIB 10682	
Plasmids		
pTZ19R		26
pTZ19U		26
pMal-c		New England Biolabs
pMal-cRI		New England Biolabs
pCT1474	pTZ19U derivative carrying the PCR product encoding the first N-terminal repeat of ORF1p	This study
pCT1475	pMal-c derivative encoding the fusion protein MalE-ORF1p-N	This study
pCT1416R	pTZ19R derivative carrying the *Pst*I-*Cla*I fragment which contains the 3′ end of ORF1	This study
pCT1472	pCT1416R derivative carrying the fragment encoding the SLH domains of ORF1p	This study
pCT1473	pMal-cRI derivative encoding the fusion protein MalE-ORF1p-C	This study

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions. Bacterial strains and plasmids used in this study are compiled in Table 1. *C. thermocellum* NCIB 10682 was grown anaerobically with gentle stirring at 60°C in complete CM3-3 medium (36) containing 5 g of cellobiose per liter (Fluka AG, Buchs, Switzerland). *E. coli* TG1 (14) was used as a cloning host and was grown in LB medium containing ticarcillin (100 μg/ml; Beecham).

DNA techniques. DNA manipulations were performed by the method of Ausubel et al. (2). Restriction enzymes were used as recommended by the suppliers.

Oligonucleotides were synthesized with a Cyclone Plus synthesizer (Milligene/Millipore Corp.). PCR amplification was performed as described previously (32) with 20 ng of cloned ORF1 DNA as a template. Single-stranded templates were sequenced by using the Taquence kit (U.S. Biochemical) and synthetic oligonucleotide primers.

Construction of pCT1475. A 529-bp fragment encoding the first N-terminal repeat of ORF1p (11), flanked by *Eco*RI and *Xba*I sites and carrying an artificial stop codon, was synthesized by PCR (Fig. 1). The forward primer was 5′-GCG GAA TTC GCT GAA GCA ACT CCA AGT-3′, and the reverse primer was 5′-TAA TCT AGA CTG CCT CTA CAA CTA TAA G-3′. The fragment was digested and inserted between the *Eco*RI and *Xba*I sites of pTZ19U (26) to yield pCT1474. The amplified sequence was confirmed by DNA sequencing.

The *Eco*RI-*Xba*I fragment of pCT1474 was cloned between the *Eco*RI and *Xba*I sites of pMal-c, yielding pCT1475 (Fig. 1). In the protein encoded by pCT1475, the first N-terminal repeat of ORF1p was fused to the C terminus of MalE to yield MalE-ORF1p-N.

Construction of pCT1473. The *Pst*I-*Cla*I fragment containing the 3′ region of ORF1 was inserted between the *Pst*I and the *Acc*I sites of pTZ19R (26), yielding pCT1416R (Fig. 2). pCT1416R was digested completely with *Sma*I and partially with *Hin*cII. The *Hin*cII-*Sma*I segment encoding the SLH domains of ORF1p was cloned at the *Sma*I site of pTZ19U, yielding pCT1472. The *Xba*I fragment of pCT1472 was inserted at the *Xba*I site of pMal-cRI, yielding pCT1473. In the protein encoded by pCT1473, the C-terminal region of ORF1p, containing the SLH domains and the major part of the G/P/T/S-rich linker, was fused to the C terminus of MalE to yield MalE-ORF1p-C.

Purification of the fusion proteins. All operations were performed at 4°C. Protein concentrations were determined by using the Bio-Rad protein assay (7) with bovine serum albumin as a standard.

The fusion protein MalE-ORF1p-N was produced from *E. coli* TG1(pCT1475). Cells (200 ml) were grown at 42°C to a optical density at 600 nm (OD_{600}) of 0.4, and isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 0.3 mM. Cells were further cultivated at 42°C to OD_{600} = 1.5, harvested, resuspended in 1/10 of the original volume of buffer 1 (20 mM Tris-HCl [pH 7.4], 200 mM NaCl, 1 mM phenylmethylsulfonyl fluoride, 1 mM EDTA, 1 mM NaN_3), and disrupted by sonication. Cell debris were removed by centrifugation (15 min at 38,000 × *g*), and the supernatant was applied to an amylose affinity column (5 by 2.5 cm; New England Biolabs). Elution was performed with 10 mM maltose in buffer 1, and the pool of MalE-ORF1p-N was dialyzed overnight against buffer 2 (50 mM Tris-HCl [pH 7.5]). The dialyzed sample (3.4 mg of protein in 7.8 ml) was concentrated to 1 mg/ml with an Amicon ultrafiltration cell (membrane PM 10).

MalE-ORF1p-C and MalE-LacZα, produced from *E. coli* TG1(pCT1473) and *E. coli* TG1(pMal-cRI), respectively, were purified according to the same protocol. The purification of MalE-ORF1p-C and MalE-LacZα yielded 15.2 and 20 mg of protein per liter of culture, respectively.

Proteins were stored in aliquots at −80°C until used.

Preparation of ORF1p-specific antibodies. Purified MalE-ORF1p-N and MalE-ORF1p-C were used to immunize rabbits as described previously (23), except that the final injection was performed 1 week after the third subcutaneous injection. Anti-MalE antibodies were removed by treating the antisera with an equal volume of a crude extract [2.6 mg of protein per ml of *E. coli* TG1(pMalcRI) cells] that had been induced for 2 h in the presence of 0.3 mM IPTG. Precipitated proteins were removed by centrifugation. Antibodies were further purified by immunoadsorption using the respective immunogens purified

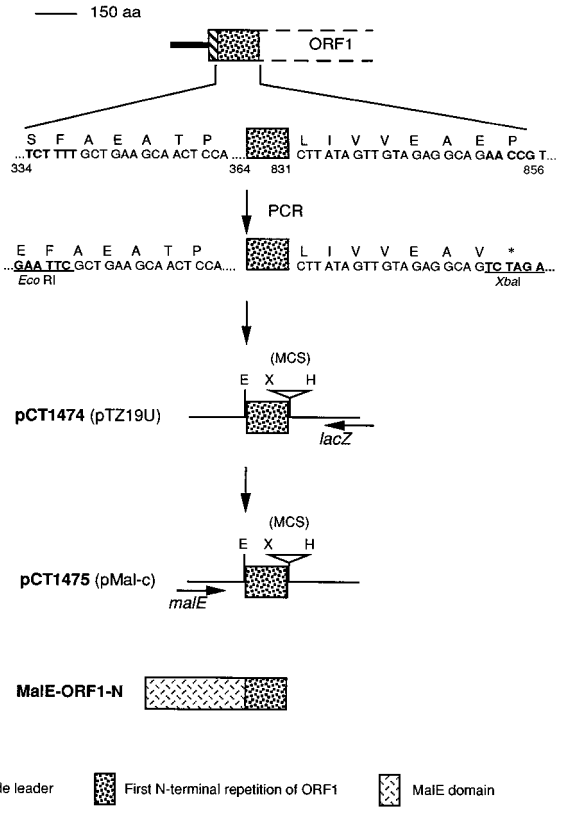

FIG. 1. Construction and structure of pCT1475, encoding the chimeric protein MalE-ORF1p-N. Numbers refer to the nucleotide sequence published by Fujino et al. (11). Nucleotides that were changed in the PCR-amplified sequence are shown in boldface type. The DNA of the vector is indicated by a thin line, and the directions of the *lacZ* and *malE* genes are indicated by arrows. MalE-ORF1p-N is drawn to scale. MCS, multiple cloning site; Ec, *Eco*RI; X, *Xba*I; H, *Hin*dIII. aa, amino acids.

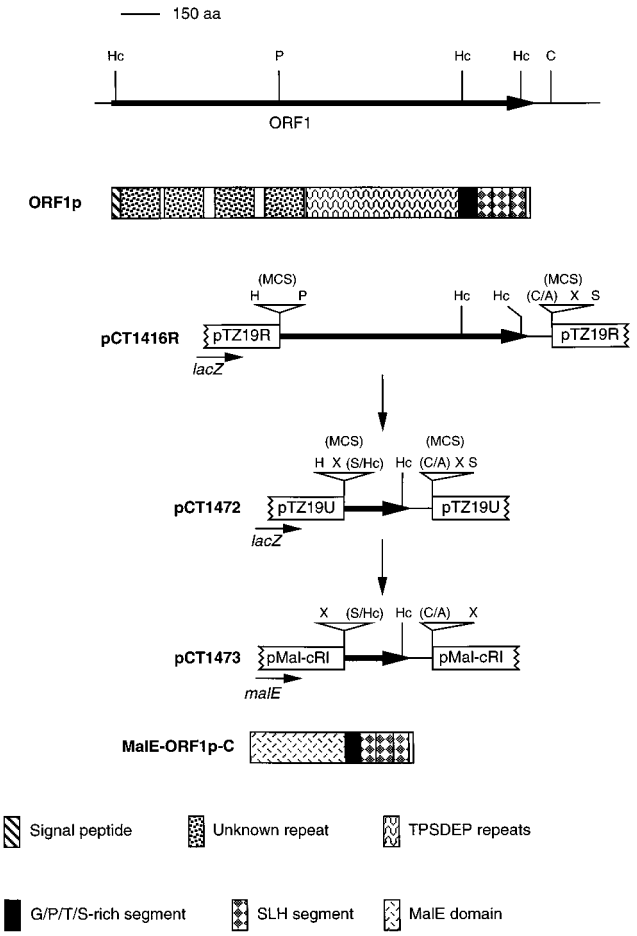

FIG. 2. Construction and structure of pCT1473, encoding the chimeric protein MalE-ORF1p-C. The position and orientation of ORF1 are indicated by a thick arrow. *C. thermocellum* DNA lying outside ORF1 is represented by a thin line. Vectors are indicated by boxes, and the directions of the *lacZ* and *malE* genes are indicated by arrows. The ORF1p polypeptide is drawn to scale and aligned with ORF1. MCS, multiple cloning site; Hc, *Hin*cII; P, *Pst*I; C, *Cla*I; X, *Xba*I; H, *Hin*dIII; S, *Sma*I; C/A, fusion of the *Cla*I site with the *Acc*I site; S/Hc, fusion of the *Sma*I site with the *Hin*cII site. aa, amino acids.

by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto nitrocellulose (31). The purified antibody fractions, recognizing the N-terminal reiterated segment or the C-terminal SLH segments of ORF1p, were termed AS-ORF1p-N and AS-ORF1p-C, respectively.

SDS-PAGE and blotting techniques. Proteins were separated by SDS-PAGE (17). Separated proteins were transferred onto nitrocellulose as described previously (2). Blots were probed either with AS-ORF1p-N or AS-ORF1p-C (dilution, 1/2,000), using the ECL Western blotting analysis system (Amersham), or with ^{125}I-labeled MalE-ORF1p-N, MalE-ORF1p-C, and MalE-LacZα (37).

Fractionation of *C. thermocellum* cultures. Unless otherwise stated, all operations were performed at 4°C. The fractionation procedure is summarized in Fig. 3. A 500-ml culture of *C. thermocellum*, grown to OD_{600} = 1.3, was centrifuged for 10 min at 10,800 × *g* (step 1). Proteins from 50 ml of the culture supernatant fraction were precipitated by slow addition of 2 volumes of cold acetone (step 2). The precipitate was solubilized in 3 ml of buffer 2, and insoluble salts, which crystallized upon standing on ice, were removed by centrifugation. The supernatant was dialyzed overnight against 1 liter of buffer 2, and the volume of the dialyzed sample was brought to 5 ml with the same buffer. This sample was called fraction F_0.

Cells were washed with 50 ml of 150 mM NaCl and resuspended in 50 ml of buffer 3 (50 mM Tris-HCl, 150 mM NaCl, 1% SDS, pH 7.5). The suspension was heated at 100°C in a water bath for 15 min and centrifuged for 10 min at 12,000 × *g* (step 3). The supernatant constituted fraction F_1.

The pellet was resuspended with 50 ml of buffer 3, yielding fraction C_1. A 25-ml aliquot of C_1 was heated at 100°C in a water bath for 15 min and centrifuged for 10 min at 12,000 × *g* (step 4). The supernatant constituted fraction F_2.

The pellet was resuspended in 25 ml of buffer 2, constituting fraction C_2. A 20-ml aliquot of C_2 was sonicated and centrifuged for 20 min at 40,000 × *g* (step 5). The supernatant constituted fraction F_3.

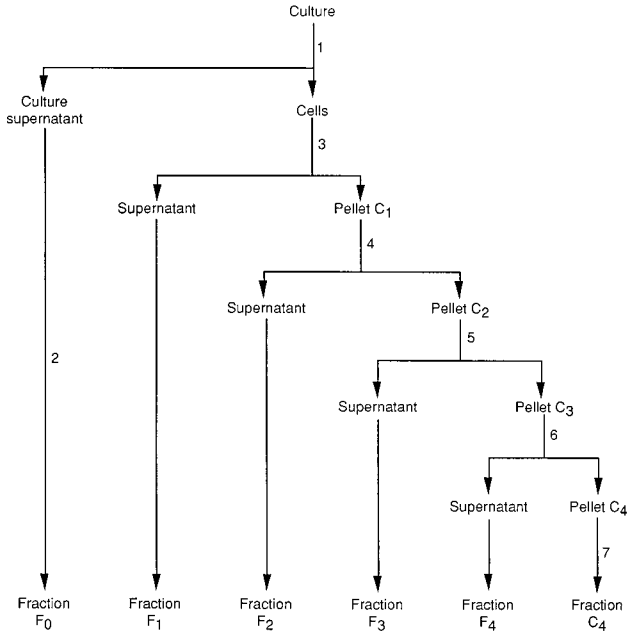

FIG. 3. Scheme of the fractionation procedure of *C. thermocellum* culture. 1, centrifugation at 10,800 × *g* for 5 min; 2, acetone precipitation and dialysis of the solubilized precipitate; 3, suspension in buffer 3, heating at 100°C for 15 min, and centrifugation at 12,000 × *g* for 10 min; 4, same as 3; 5, suspension in buffer 2, sonication, and centrifugation at 40,000 × *g* for 20 min; 6, suspension in buffer 3, heating at 100°C for 15 min, and centrifugation at 40,000 × *g* for 20 min; 7, suspension in buffer 2.

The pellet was resuspended in 20 ml of buffer 3, constituting fraction C_3. A 10-ml aliquot of C_3 was heated at 100°C in a water bath for 15 min and centrifuged for 20 min at 40,000 × *g* (step 6). The supernatant constituted fraction F_4.

The pellet was resuspended in 10 ml of buffer 2, constituting fraction C_4 (step 7).

All fractions were concentrated 10-fold with respect to the original culture volume.

Interaction of MalE-ORF1p-C with the peptidoglycan of *C. thermocellum*. The C_4 fraction (10 ml), containing the peptidoglycan, was washed five times with 20 mM sodium phosphate buffer (pH 7.5), and the pellet was resuspended in 5 ml of the same buffer. Purified MalE-ORF1p-C was dialyzed overnight against the same buffer. A 200-μg aliquot of the protein was mixed with 300 μl of fraction C_4, and the reaction volume was brought to 1 ml with phosphate buffer. The suspension was incubated at 37°C for 4 h and centrifuged at 40,000 × *g* for 20 min. The supernatant (1 ml) yielded fraction S (soluble). The pellet was resuspended in 1 ml of phosphate buffer, vortexed, and centrifuged at 40,000 × *g* for 20 min. The supernatant constituted fraction W (wash). The pellet was resuspended in 1 ml of phosphate buffer, yielding fraction I (insoluble). A 10-μl portion of each fraction was analyzed by SDS-PAGE.

Electron microscopy. *C. thermocellum* cells were harvested during the exponential growth phase. Immunocytochemical labeling of whole cells and immunolabeling of Lowicryl HM20 thin sections were performed as described previously (31). For conventional electron microscopy, aliquot fractions were fixed in 2.5% glutaraldehyde in 0.1 M Na cacodylate (pH 7.2) for 1 h at room temperature, post-fixed with 1% OsO_4 in the same buffer for 1 h, dehydrated in an acetone series, and embedded in Epon.

Analysis of the peptidoglycan of *C. thermocellum*. Peptidoglycan (fraction C_4) was vacuum dried and hydrolyzed in 200 μl of 6 N HCl—0.2% phenol at 110°C for 20 h in sealed tubes. Aliquots of the hydrolyzate were analyzed by a Beckman 6300 amino acid analyzer. Glucosamine and diaminopimelate were quantified by appropriate standards. The same analysis was performed with MalE-ORF1p-C and fraction I, containing MalE-ORF1p-C bound to the peptidoglycan.

RESULTS

Engineering and purification of fusion proteins. ORF1p is a multidomain protein (11). The N-terminal part of the polypeptide sequence begins with a signal peptide followed by four

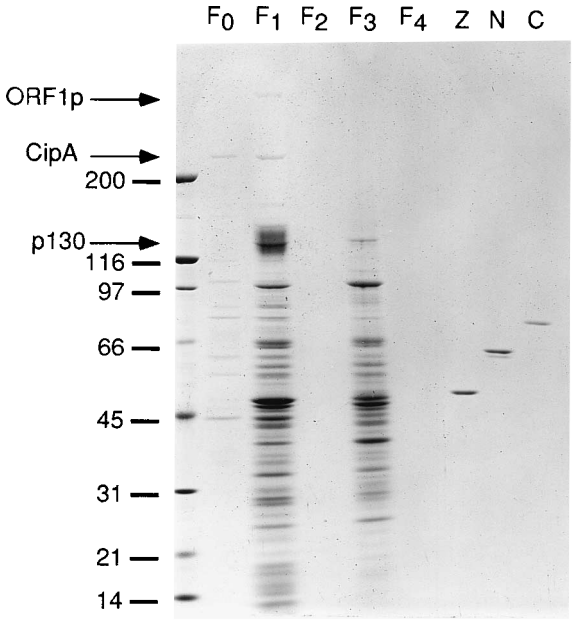

FIG. 4. Coomassie blue-stained 5 to 15% polyacrylamide-SDS gel (2) displaying the culture fractions and the purified chimeric proteins. Samples (10 μl) of fractions F_0 through F_4 (corresponding to 0.1 ml of original culture at OD_{600} = 1.3) and 0.5 μg of each purified chimeric protein were analyzed. Lanes: Z, MalE-LacZα; N, MalE-ORF1p-N; C, MalE-ORF1p-C. The positions of ORF1p, CipA, and the S-layer protein p130 are indicated by arrows. The leftmost lane shows molecular size standards: myosin (200 kDa), β-galactosidase (116 kDa), phosphorylase *b* (97 kDa), albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), trypsin inhibitor (21 kDa), lysozyme (14 kDa). Bands presumably corresponding to CipA and ORF1p are indicated by arrows.

highly similar repeats of 156 residues each. The central region is extremely repetitive and contains 73 copies of the hexapeptide TPSDEP. The TPSDEP repeats are separated from the C-terminal SLH domains by a G/P/T/S-rich linker. Figures 1 and 2 show the structure of the plasmids expressing the first N-terminal repeat and the SLH domains of ORF1p. The first N-terminal repeat of ORF1p was grafted onto MalE to yield MalE-ORF1p-N. In the same way, the SLH domains of ORF1p, flanked by the G/P/T/S-rich linker present in the original protein, were fused to MalE to yield MalE-ORF1p-C. Electrophoresis of purified proteins is shown in Fig. 4 (lanes Z, N, C). The electrophoretic migration of MalE-ORF1p-N, MalE-ORF1p-C and MalE-LacZα, corresponding to molecular masses of 63, 76, and 52 kDa, respectively, was consistent with the predicted masses of the polypeptides (60,719, 72,105, and 50,103 Da, respectively).

Fractionation of *C. thermocellum* cultures. Figure 4 shows the monitoring of the fractionation. Fraction F_0, corresponding to the culture medium, exhibited a set of bands matching with those of the cellulosome (19). A prominent 235-kDa band most likely corresponded to CipA. *C. thermocellum* cells were treated with SDS at 100°C in order to prevent extensive proteolysis of ORF1p. The treatment solubilized many proteins ranging from ca. 340 to 5 kDa (fraction F_1). Such a treatment removed proteins that were noncovalently associated to the cell surface. Thus, several proteins belonging to the *C. thermocellum* cell envelope could be identified in fraction F_1, such as the cellulosomal scaffolding protein CipA, the surface layer protein (130 kDa; unpublished results), and the cell envelope protein OlpA (31), which could be detected in F_1 by Western blotting (data not shown). These proteins were not identified in the other fractions. This suggests that proteins present only in F_1 were located in the cell envelope. F_1 also contained a significant amount of cytoplasmic proteins, which leaked out of the cells because of the boiling SDS treatment. In contrast to cell surface proteins, however, further mechanical disruption of the cells was required to release cytoplasmic proteins entirely (fraction F_3).

Electron micrographs showed that most of the cells treated with hot SDS retained their shape and did not appear to lose much of their cytoplasmic content (Fig. 5). The cell wall appeared as a thin layer which might consist of peptidoglycan. However, some cells were damaged, releasing their cytoplasm into the medium.

No protein was detected after Coomassie blue staining of material extracted by a second treatment with hot SDS (fraction F_2). However, a set of proteins, ranging from 131 to 10 kDa, was released after sonication of SDS-treated cells (fraction F_3). These proteins came from cells which had resisted the hot SDS treatment and were presumably representative of the cytoplasmic fraction of *C. thermocellum* proteins. Pellet C_3 corresponded to envelopes of SDS-washed cells. Washing of C_3 with hot SDS (fraction F_4) and treatment of C_4 (material corresponding to 0.3 ml of original culture) with lysozyme (data not shown) failed to release any component that could be detected upon SDS-PAGE followed by Coomassie blue staining.

The components of fraction C_4 were analyzed after acid hydrolysis (Table 2). Major hydrolysis products were glucosamine, alanine, glutamate, and diaminopimelate. The composition was in good agreement with the peptidoglycan composition of numerous clostridia (34). No muramate was detected, indicating that it had been fully converted to glucosamine in the course of acid hydrolysis. Likewise, peaks that could have resulted from the decomposition of glucosamine were absent, indicating that destruction of the latter was minimal. The molar ratio between glucosamine and diaminopimelate was about 2. Assuming that 50% of the glucosamine originated from the hydrolysis of muramic acid, this ratio suggests that peptidoglycan was the predominant component of fraction C_4. However, the presence of minor quantities of secondary cell wall polymers that would be covalently linked to the peptidoglycan cannot be ruled out. The amount of most non-peptidoglycan amino acid residues did not exceed 0.05 mol per mol of glucosamine, i.e., 0.1 mol per mol of original glucosamine. Hence, fraction C_4 appeared to contain little or no covalently associated protein.

Identification and distribution of ORF1p in different fractions of *C. thermocellum*. In order to discriminate between genuine ORF1p and proteins bearing immunologically related determinants, two antisera were raised against different regions of the polypeptide. Preadsorbed, affinity-purified antibodies AS-ORF1p-N and AS-ORF1p-C failed to recognize MalE-LacZα (Fig. 6A and 6B, lane Z). No band was observed with preimmune serum or with antibodies saturated with the corresponding original antigen (MalE-ORF1p-N or MalE-ORF1p-C; data not shown).

A major band, present exclusively in fraction F_1, was revealed by both antibodies suggesting that ORF1p was associated with the cell envelope (Fig. 6A and 6B). By extrapolation, it was estimated that the labeled species migrated like a 340-kDa polypeptide. The predicted mass of ORF1p is 178,219 Da. However, proteins with a high percentage of proline are known to display anomalous migration in SDS-PAGE (12). In the case of ORF1p, proline is the most abundant residue, and the central domain has a quite atypical, very hydrophilic composition, which probably affects binding of SDS and electro-

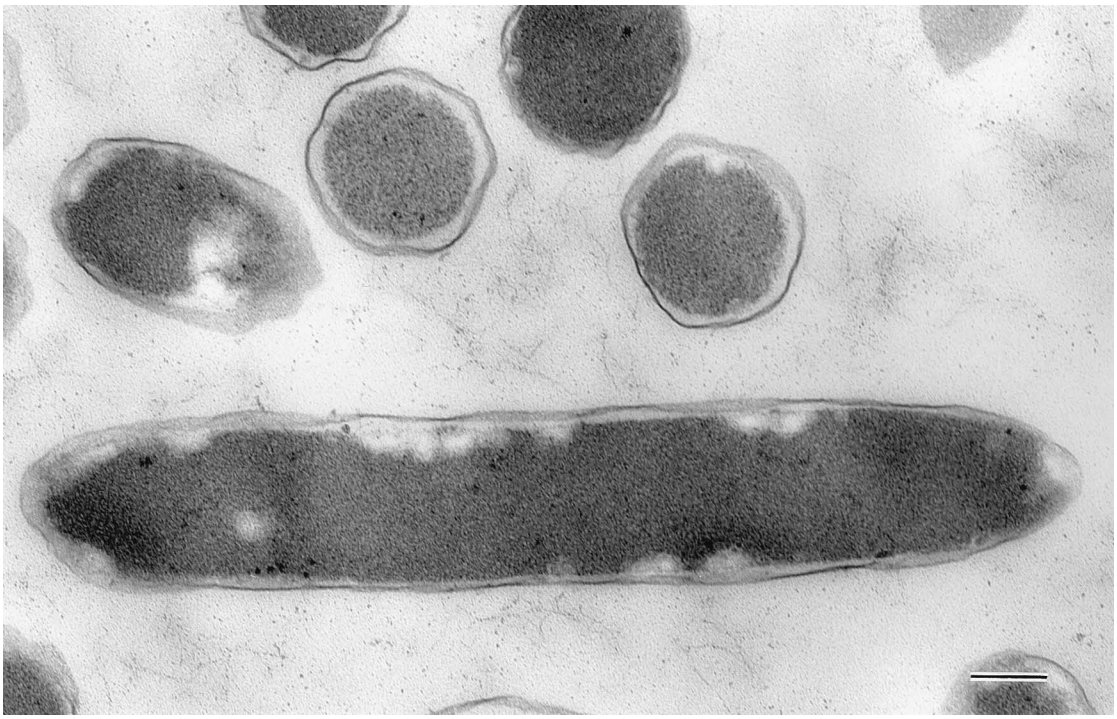

FIG. 5. Ultrastructural morphology of *C. thermocellum* cells after treatment with boiling SDS (see Materials and Methods). Bar, 200 nm.

TABLE 2. Components of fraction I (peptidoglycan-bound MalE-ORF1p-C), fraction C_4, and MalE-ORF1p-C liberated by acid hydrolysis

Component	Amt		
	Fraction I[a]	Fraction C_4[a]	MalE-ORF1-C[b]
Glucosamine[c]	1.000	1.000	ND[d]
Diaminopimelate	0.540	0.493	ND
D + N	1.261	0.081	1.215
E + Q	1.593	0.589	1.035
C	0.005	0.003	ND
T	0.646	0.040	0.597
S	0.707	0.121	0.569
G	1.181	0.122	1.017
A	1.942	0.780	1.182
V	0.479	0.049	0.505
M	0.025	ND	0.021
I	0.637	0.050	0.574
L	0.877	0.064	0.825
Y	0.437	ND	0.422
F	0.517	ND	0.462
H	0.162	0.029	0.129
K	0.900	0.061	0.861
R	0.315	ND	0.280
Total[e]	11.684	1.989	9.694

[a] Mol/mol of glucosamine.
[b] Values normalized relative to an average amount of 1.59×10^{-2} nmol of MalE-ORF1p-C.
[c] Including muramic acid degraded to glucosamine by acid hydrolysis.
[d] ND, not detected.
[e] Excluding glucosamine and diaminopimelate.

phoretic migration. Further deviation from regular electrophoretic behavior may be expected if ORF1p is glycosylated, as are many bacterial cell surface proteins (35). No change in electrophoretic migration was observed upon heating the sample in 2% SDS + 10 mM EDTA prior to electrophoresis (data not shown). Formation of a Ca^{2+}-stabilized, SDS-stable dimer similar to the trimer observed for the S-layer of *Thermus thermophilus* (5) is therefore unlikely.

The ORF1p band was accompanied by minor bands of lower M_r which could result either from proteolytic events or from immunological cross-reaction. Because of the highly repetitive structure of the TPSDEP-containing segment, proteolytic cleavage of ORF1p would be expected to generate a more complex pattern than the three minor bands of 61, 88, and 120 kDa detected by AS-ORF1p-C (Fig. 6B). However, the three bands might correspond to cross-reacting polypeptides of *C. thermocellum* containing SLH domains.

Localization of ORF1p by electron microscopy. ORF1p was visualized on the surface of whole cells using AS-ORF1p-N as a probe. No labeling was detected when the antiserum was omitted (data not shown). The protein was associated with the outermost layer surrounding the cells (Fig. 7a). This layer appears similar to the protuberance-forming, ferritin-binding layer previously described in references 3 and 31. Labeling of whole cells with AS-ORF1p-C antibodies was much less intense than with AS-ORF1p-N (Fig. 7b). The difference suggests that the N-terminal epitope may be more accessible than the C terminus of the protein, which may be buried within the cell envelope. The less intense labeling observed with AS-ORF1p-C was not due to a lesser efficiency of the antibody fraction. Indeed, when labeling was performed on thin sections, the C-terminal epitope was exposed and was labeled at least as efficiently as the N-terminal epitope (Fig. 7c and 7d).

Interactions of the SLH domains of ORF1p with other cellular components. The interaction of the SLH domains of

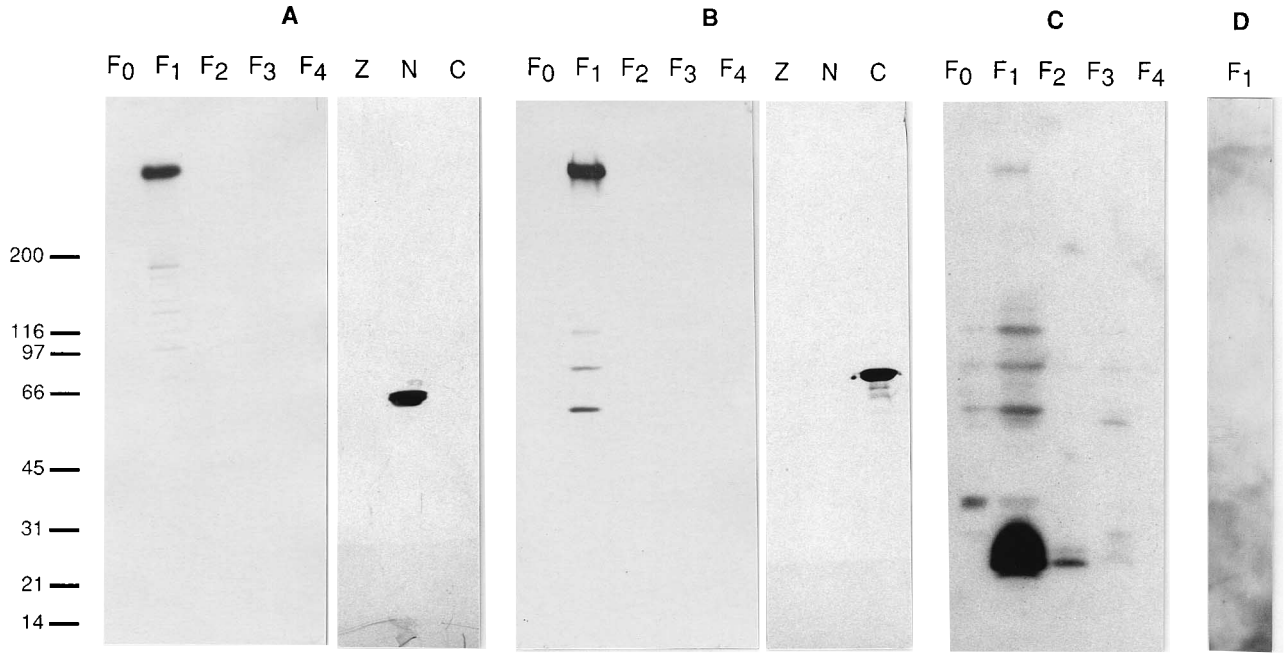

FIG. 6. Autoradiograms from Western blotting carried out on fractions of *C. thermocellum* (lanes F_0 to F_4) or on purified chimeric proteins (lanes Z, N, and C). Electrophoresis was performed as in Fig. 4 prior to blotting. Blots were probed either with AS-ORF1p-N (panel A), AS-ORF1p-C (panel B), ^{125}I-labeled MalE-ORF1p-C (panel C), or ^{125}I-MalE-LacZα (panel D). Samples (10 μl) of fractions F_0 through F_4 (corresponding to 0.1 ml of original culture at OD_{600} = 1.3) and 0.5 μg of each purified chimeric protein were analyzed. Lanes: Z, MalE-LacZα; N, MalE-ORF1p-N; C, MalE-ORF1p-C. The migration of molecular size markers (in kilodaltons) is indicated on the left.

MalE-ORF1p-C with *C. thermocellum* peptidoglycan was tested by incubating MalE-ORF1p-C with peptidoglycan (fraction C_4). Under the conditions of the assay, about 50% of the protein cosedimented with the peptidoglycan fraction (Fig. 8B, lane I). Resuspending the pellet in phosphate buffer failed to release any detectable MalE-ORF1p-C (Fig. 8B, lane W). Under the same conditions, MalE-ORF1p-N and MalE-LacZα did not cosediment with peptidoglycan (Fig. 8C and 8D, lane I). In the same way, MalE-ORF1p-C did not precipitate when incubated without the peptidoglycan fraction (Fig. 8A). As mentioned above, fraction C_4 consisted of peptidoglycan devoid of associated proteins. Treating fraction C_4 with proteases, such as trypsin or V8 protease, did not affect its ability to bind MalE-ORF1p-C (data not shown). Thus, these data suggest that the SLH domains of ORF1p interact directly with the *C. thermocellum* peptidoglycan (or whatever secondary cell wall polymer might be associated with it in fraction C_4). The difference in amino acid content relative to glucosamine between fraction I (peptidoglycan + MalE-ORF1p-C) and fraction C_4 (peptidoglycan alone) amounted to 9.7 mol/mol of glucosamine, corresponding to 1.59×10^{-2} mol of MalE-ORF1p-C bound per mol of glucosamine (Table 2).

Proteins interacting with the SLH domains of ORF1p were detected by Western blotting of fractions F_0 through F_4. A set of polypeptides present mostly in F_1 was revealed when blots were probed with ^{125}I-labeled MalE-ORF1p-C (Fig. 6C) but not with ^{125}I-labeled MalE-ORF1p-N (data not shown) or MalE-LacZα (Fig. 6D). The most strongly labeled species, migrating towards the bottom of the gel (Fig. 6C, F_1), was resolved into a 26- and a 28-kDa band upon shorter exposure. The proportion of the 26-kDa band relative to the 28-kDa band increased upon storage at −20°C (data not shown), suggesting that the 26-kDa species may derive from the 28-kDa species by proteolysis. In addition to the 26- to 28-kDa species and a minor 38-kDa species, four bands of 340, 120, 88, and 61 kDa, respectively, were detected in F_1 and, to a lesser extent, in the culture medium. These bands comigrated with the bands labeled by the anti-SLH AS-ORF1p-C antibodies. Therefore, we checked whether SLH domains would promote binding of MalE-ORF1p-C to proteins known to contain similar domains. Purified MalE-LacZα, MalE-ORF1p-C, OlpA (31), and MalE-RDOlpA, a fusion protein carrying the N-terminal domain of OlpA (32), were transferred onto nitrocellulose. When the blot was probed with ^{125}I-labeled MalE-ORF1p-C, MalE-ORF1p-C and OlpA, which contain SLH domains, were labeled, whereas MalE-LacZα and MalE-RDOlpA, which are devoid of such domains, were not (Fig. 9B). Moreover, nondenaturing electrophoresis indicated the presence of oligomeric forms of MalE-ORF1p-C, while MalE appeared monomeric (data not shown). These data indicate that MalE-ORF1p-C can participate in intermolecular SLH-SLH interactions. Indeed, the 61-, 88-, and 120-kDa bands revealed by AS-ORF1p-C and ^{125}I-labeled MalE-ORF1p-C might correspond to polypeptides of *C. thermocellum* known to possess SLH domains. Thus, immunoblotting with anti-OlpA revealed a polypeptide comigrating with the 61-kDa band (data not shown). The 88- and 120-kDa bands might represent the polypeptides encoded by ORF2 (11) and the xylanase gene *xynX* (GenBank accession no. M67438) of *C. thermocellum* (predicted molecular masses of 74,867 and 120,374, respectively).

DISCUSSION

Electron microscopy supported by biochemical investigations demonstrated the localization of ORF1p on the cell surface of *C. thermocellum*. ORF1p appeared to be located in the same protuberance-forming outer layer as OlpA (31). This

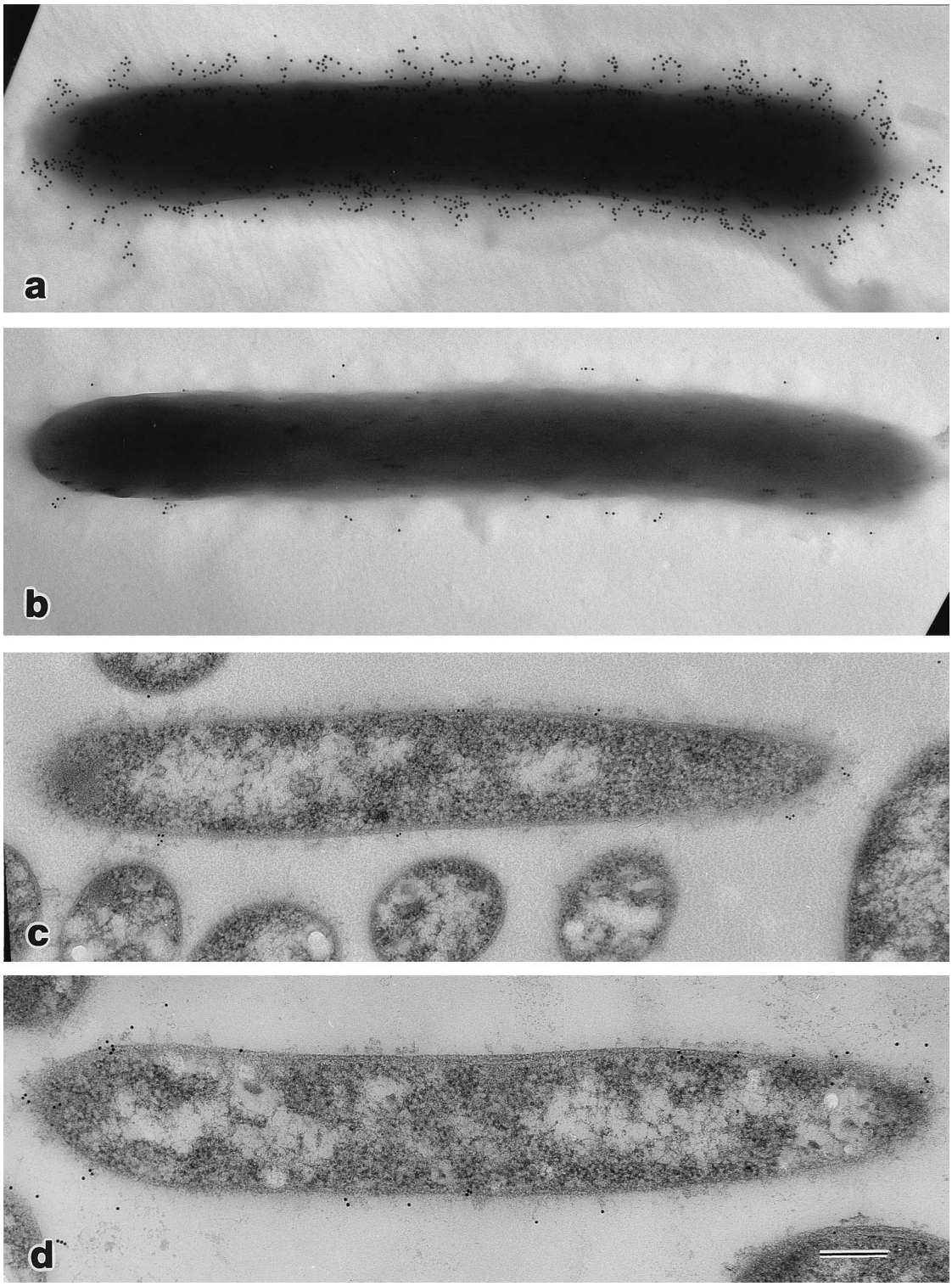

FIG. 7. Whole cells (panels a and b) and thin sections (panels c and d) of *C. thermocellum*. Cells in panels a and c were labeled with AS-ORF1p-N antibodies. Cells in panels b and d were labeled with AS-ORF1p-C antibodies. Bar, 200 nm.

layer consists of amorphous, anionic material reacting with ferritin (3, 31). By analogy with OlpA, we propose to name *olpB* (for outer layer protein B) the gene carrying ORF1 and its product OlpB.

As can be judged from the intensity of the OlpB band after staining with Coomassie blue (Fig. 4), OlpB is present in the F_1 fraction in lower amounts than CipA. Except for the SLH domains, OlpB shows no structural similarity with proteins having a known function. Thus, its role remains elusive.

Sequence analysis suggests that SLH domains may serve to

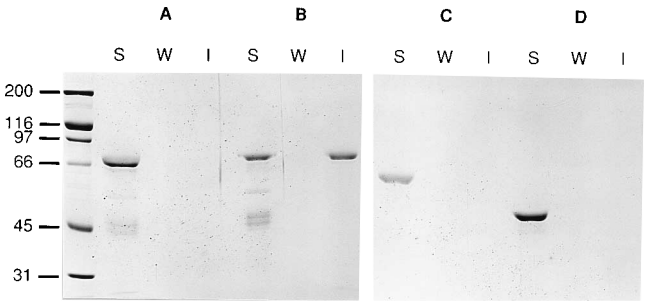

FIG. 8. Interaction of MalE-ORF1p-C with the peptidoglycan of *C. thermocellum*. MalE-ORF1p-C was incubated either in the absence (panel A) or in the presence (panel B) of the peptidoglycan. Any insoluble material was pelleted and washed as described in Materials and Methods. Panels C and D show similar experiments performed with MalE-ORF1p-N and MalE-LacZα, respectively. Lanes: S, soluble fraction; W, sodium phosphate wash; I, insoluble fraction. The leftmost lane shows molecular size markers (in kilodaltons).

anchor proteins to the cell surface of bacteria. SLH domains occur in proteins which are (or can be reasonably surmised to be) located in the cell envelope and which share otherwise little sequence similarity. Several observations led Lupas et al. (24) to propose a putative role of SLH domains in binding to peptidoglycan. Thus, the Ompα protein of *T. maritima* appears as a fibrous protein connecting the outer membrane to the cell body (8), and it was proposed that the N terminus of Ompα may anchor the protein to the peptidoglycan layer (8, 24). Treatment with lysozyme facilitates the extraction of the S-layer protein (P_{100}) of *T. thermophilus* with Triton X-100 (9), suggesting direct or indirect interaction with the peptidoglycan. In the case of *B. sphaericus*, the intact S-layer protein was compared with a proteolytically truncated form missing an 18-kDa fragment which may correspond to the N-terminal SLH domain. Both proteins were able to form a crystalline lattice, but the truncated form lost the ability to bind to the peptidoglycan layer (15). In agreement with these arguments, our data provide direct evidence that SLH domains enhance the capability of proteins to bind to peptidoglycan.

It has also been hypothesized that SLH domains could participate in homologous interactions. According to secondary structure predictions, SLH domains should consist of two α-helices flanking a β-strand. Since a single β-strand would be structurally meaningless, it was proposed that the latter may participate in intermolecular interactions involving SLH domains (24). Indeed, blotting experiments showed that SLH domains conferred on ^{125}I-labeled MalE-ORF1p-C the capacity to bind to other proteins carrying similar domains.

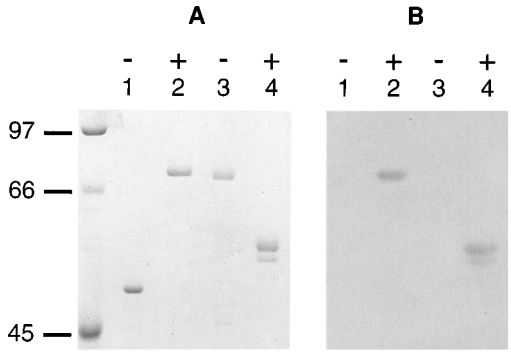

FIG. 9. Interaction of ^{125}I-labeled MalE-ORF1p-C with SLH domains of purified proteins. Proteins were loaded on 10% polyacrylamide-SDS gels and were either stained with Coomassie blue (panel A) or blotted onto nitrocellulose filter and probed with ^{125}I-labeled MalE-ORF1p-C (panel B). A 0.5-μg sample of each purified protein was loaded. The presence or absence of SLH domains in each protein is indicated by + or −, respectively. Lanes: 1, MalE-LacZα; 2, MalE-ORF1p-C; 3, MalE-RDOlpA; 4, OlpA. The leftmost lane shows molecular size markers (in kilodaltons).

The affinity of SLH domains for the peptidoglycan supports the hypothesis that they may serve as a sorting signal for cell-bound exoproteins. So far, two other determinants are known to mediate sorting of proteins to the cell surface. One is the LPXTG motif located towards the C terminus of streptococcal and staphylococcal surface proteins. This motif appears to attach covalently to the peptidoglycan by means of a transpeptidylation reaction (28). The other determinant is the reiterated motif located at the C terminus of the surface protein A (PspA) of *Streptococcus pneumoniae* (38). The *S. pneumoniae* determinant binds noncovalently to choline residues located in the lipoteichoic acid of *S. pneumoniae.*

It is not quite clear, however, how the binding properties mentioned above fit with the localization of OlpA (31) and OlpB within the envelope of *C. thermocellum*. Immunocytochemical labeling suggested that both components accumulate in the amorphous outer layer of the cell surface, rather than in the close vicinity of the peptidoglycan layer. Moreover, it is doubtful whether SLH domains interact directly with the major protein of the S-layer of *C. thermocellum*. A typical S-layer, with hexagonal symmetry, has been identified in *C. thermocellum* (reference 29 and unpublished data). Its major component appears to be a 130-kDa glycoprotein, first described and characterized by Lamed and Bayer (reference 18 and unpublished data). This protein does not react with AS-ORF1p-C antibodies or with ^{125}I-labeled MalE-ORF1p-C. This suggests that S-layer assembly in *C. thermocellum* does not rely on interactions mediated by SLH domains. Cloning and sequencing of the 130-kDa protein should provide definitive evidence to this point.

However, attachment to the cell surface is not necessarily a permanent feature of cell envelope proteins. Some S-layer proteins are shed off into the culture medium (1, 30). It may be that OlpA and OlpB bind transiently to the peptidoglycan by means of their SLH domains, prior to being released and trapped in the outer layer surrounding the cells, which also contains cellulosomes (3). It is also conceivable that SLH domain-bearing proteins, like OlpA, OlpB, and, most likely, the protein encoded by ORF2, which has not been studied yet, may contribute to the formation of a protein network within the outer surface layer. Such a network could involve homologous interactions between SLH domains, or between SLH domains and the 26- to 28-kDa component.

ACKNOWLEDGMENTS

We thank F. Baleux (Unité de Chimie Organique, Institut Pasteur) for performing the analysis of the peptidoglycan of *C. thermocellum*. We are grateful to J.-P. Aubert and M. Schwartz for their continuing interest and helpful discussions.

M.L. was the recipient of a fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche. This work was supported by contracts CPL C462 and AIR1-CT-0321 from the Commission of the European Communities and by research funds from the University of Paris 7.

REFERENCES

1. **Adachi, T., H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1991. Repression of the cell wall protein gene operon in *Bacillus brevis* 47 by magnesium and calcium ions. J. Bacteriol. **173:**4243–4245.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1990. Current protocols in molecular biology. Greene Publishing and Wiley Interscience, New York.

3. **Bayer, E. A., and R. Lamed.** 1986. Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. J. Bacteriol. **167:**828–836.
4. **Béguin, P., and J. P. Aubert.** 1994. The biological degradation of cellulose. FEMS Microbiol. Rev. **13:**25–58.
5. **Berenguer, J., M. L. M. Faraldo, and M. A. de Pedro.** 1988. Ca^{2+}-stabilized oligomeric complexes are major components of the cell envelope of *Thermus thermophilus* HB8. J. Bacteriol. **170:**2441–2447.
6. **Bowditch, R. D., P. Baumann, and A. A. Yousten.** 1989. Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer protein from *Bacillus sphaericus* 2362 and a related cryptic gene. J. Bacteriol. **171:**4178–4188.
7. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
8. **Engel, A. M., Z. Cejka, A. Lupas, F. Lottspeich, and W. Baumeister.** 1992. Isolation and cloning of Ompα, a coiled-coil protein spanning the periplasmic space of the ancestral eubacterium *Thermotoga maritima*. EMBO J. **11:**4369–4378.
9. **Faraldo, M. L., M. A. de Pedros, and J. Berenguer.** 1988. Purification, composition and Ca^{2+}-binding properties of the monomeric protein of the S-layer of *Thermus thermophilus*. FEBS Lett. **235:**117–121.
10. **Fujino, T., P. Béguin, and J.-P. Aubert.** 1992. Cloning of a *Clostridium thermocellum* DNA fragment encoding polypeptides that bind the catalytic components of the cellulosome. FEMS Microbiol. Lett. **94:**165–170.
11. **Fujino, T., P. Béguin, and J.-P. Aubert.** 1993. Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in the attachment of the cellulosome to the cell surface. J. Bacteriol. **175:**1891–1899.
12. **Furthmayr, H., and R. Timpl.** 1971. Characterization of collagen peptides by sodium dodecylsulfate-polyacrylamide electrophoresis. Anal. Biochem. **41:** 510–516.
13. **Gerngross, U. T., M. P. M. Romaniec, N. S. Huskisson, and A. L. Demain.** 1993. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal S_L-protein reveals an unusual degree of internal homology. Mol. Microbiol. **8:**325–334.
14. **Gibson, T. J.** 1984. Studies on the Epstein-Barr virus genome. Ph.D. thesis. University of Cambridge, Cambridge.
15. **Hastie, A. T., and C. C. Brinton, Jr.** 1979. Specific interaction of the tetragonally arrayed protein layer of *Bacillus sphaericus* with its peptidoglycan sacculus. J. Bacteriol. **138:**1010–1021.
16. **Hostalka, F., A. Moultrie, and F. Stutzenberger.** 1992. Influence of carbon source on cell surface topology of *Thermomonospora curvata*. J. Bacteriol. **174:**7048–7052.
17. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
18. **Lamed, R., and E. A. Bayer.** 1988. The cellulosome concept: exocellular/extracellular enzyme reactor centers for efficient binding and cellulolysis, p. 101–116. *In* J.-P. Aubert, P. Béguin, and J. Millet (ed.), FEMS Symposium no. 43. Biochemistry and genetics of cellulose degradation. Academic Press, London.
19. **Lamed, R., and E. A. Bayer.** 1988. The cellulosome of *Clostridium thermocellum*. Adv. Appl. Microbiol. **33:**1–46.
20. **Lamed, R., J. Naimark, E. Morgenstern, and E. A. Bayer.** 1987. Specialized cell surface structures in cellulolytic bacteria. J. Bacteriol. **169:**3792–3800.
21. **Lamed, R., E. Setter, R. Kenig, and E. A. Bayer.** 1983. The cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. Biotechnol. Bioeng. Symp. **13:**163–181.
22. **Lee, Y.-E., S. E. Lowe, and J. G. Zeikus.** 1992. Molecular biology and physiological biochemistry of xylan degradation by thermoanaerobes, p. 275–288. *In* J. Visser, G. Beldman, M. A. Kusters-van Someren, and A. G. J. Voragen (eds.), Progress in biotechnology, vol. 7. Xylans and xylanases. Elsevier, Amsterdam.
23. **Lemaire, M., and P. Béguin.** 1993. Nucleotide sequence of the *celG* gene of *Clostridium thermocellum* and characterization of its product, endoglucanase CelG. J. Bacteriol. **175:**3353–3360.
24. **Lupas, A., H. Engelhardt, J. Peters, U. Santarius, S. Volker, and W. Baumeister.** 1994. Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. J. Bacteriol. **176:**1224–1233.
25. **Matuschek, M., G. Burchhardt, K. Sahm, and H. Bahl.** 1994. Pullulanase of *Thermoanaerobacter thermosulfurigenes* EM1 (*Clostridium thermosulfurogenes*): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface. J. Bacteriol. **176:**3295–3302.
26. **Mead, D. A., E. Szczesna-Skorupa, and B. Kemper.** 1986. Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Eng. **1:**67–74.
27. **Morag, E., E. A. Bayer, and R. Lamed.** 1992. Unorthodox intra-subunit interactions in the cellulosome of *Clostridium thermocellum*: identification of structural transitions induced in the S1 subunit. Appl. Biochem. Biotechnol. **33:**205–217.
28. **Navarre, W. W., and O. Schneewind.** 1994. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in Gram-positive bacteria. Mol. Microbiol. **14:**115–121.
29. **Nolte, A.** 1992. Ultrastrukturelle, enzymkinetische und immunocytochemische Untersuchungen der extrazellulären Cellulase von *Clostridium thermocellum* JW20. Ph.D. thesis. Georg-August-Universität Göttingen, Göttingen, Germany.
30. **Peyret, J.-L., N. Bayan, G. Joliff, T. Gulik-Krzywicki, L. Mathieu, E. Shechter, and G. Leblon.** 1993. Characterization of the *cspB* gene encoding PS2, an ordered surface-layer protein in *Corynebacterium glutamicum*. Mol. Microbiol. **9:**97–109.
31. **Salamitou, S., M. Lemaire, T. Fujino, H. Ohayon, P. Gounon, P. Béguin, and J. P. Aubert.** 1994. Subcellular localization of *Clostridium thermocellum* ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome. J. Bacteriol. **176:**2828–2834.
32. **Salamitou, S., O. Raynaud, M. Lemaire, M. Coughlan, P. Béguin, and J. P. Aubert.** 1994. Recognition specificity of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA. J. Bacteriol. **176:**2822–2827.
33. **Salamitou, S., K. Tokatlidis, P. Béguin, and J. P. Aubert.** 1992. Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. FEBS Lett. **304:**89–92.
34. **Schleifer, K. H., and O. Kandler.** 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol. Rev. **36:**407–477.
35. **Sleytr, U. B., P. Messner, D. Pum, and M. Sara.** 1993. Crystalline bacterial cell surface layers: general principles and application potential. J. Appl. Bacteriol. **74:**21S–32S.
36. **Tailliez, P., H. Girard, J. Millet, and P. Béguin.** 1989. Enhanced cellulose fermentation by an asporogenous and ethanol-tolerant mutant of *Clostridium thermocellum*. Appl. Environ. Microbiol. **55:**207–211.
37. **Tokatlidis, K., S. Salamitou, P. Béguin, P. Dhurjati, and J.-P. Aubert.** 1991. Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. FEBS Lett. **291:**185–188.
38. **Yother, J., and J. M. White.** 1994. Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein PspA. J. Bacteriol. **176:**2976–2985.