Recognition Specificity of the Duplicated Segments Present in *Clostridium thermocellum* Endoglucanase CellD and in the Cellulosome-Integrating Protein CipA

SYLVIE SALAMITOU,1 ODETTÉ RAYNAUD,1 MARC LEMAIRE,1 MICHAEL COUGHLAN,2† PIERRE BÉGUIN,1,* AND JEAN-PAUL AUBERT1

Unité de Physiologie Cellulaire and URA 1300 Centre National de la Recherche Scientifique, Département des Biotechnologies, Institut Pasteur, 75724 Paris Cedex 15, France,1 and Department of Biochemistry, University College, Galway, Ireland2

Received 14 December 1993/Accepted 28 February 1994

*Clostridium thermocellum* is one of the most efficient microorganisms with respect to crystalline cellulose degradation. This property is conferred on the bacteria by a high-molecular-mass complex (2 × 10^9 to 4 × 10^9 Da) termed the cellulosome (12, 13).

This complex, which is very difficult to dissociate into its individual components, contains at least 14 different polypeptides, as visualized by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Most of the components exhibited endoglucanase activity when tested by a carboxymethyl cellulose gel overlay method (3, 12). In addition, a 250-kDa glycoprotein termed CipA was shown to bind to cellulose (14, 17) and to act as a scaffolding protein of the cellulosome. Catalytic subunits of the cellulosome contain a conserved duplicated segment which serves as a docking sequence. This docking sequence interacts with a complementary receptor whose sequence is reiterated ninefold within the sequence of CipA (6, 8, 21). CipA also possesses a cellulose-binding domain (16) and a COOH-terminal duplicated segment similar to that present in the catalytic subunits (6, 8).

Downstream of CipA, another gene carrying an open reading frame termed ORF3 encodes a protein, ORF3p, which can also bind to the duplicated segment of the catalytic subunits (6). The sequence of ORF3p comprises an N-terminal domain of 140 amino acids which is similar to the nine receptors for the duplicated segment carried by CipA. The COOH terminus of ORF3p is composed of three repeats of about 65 amino acids each; these show similarity to the surface layer proteins of several bacteria (7).

On the basis of these features, it was proposed that ORF3p could serve to anchor the cellulosome to the cell surface. The receptor domain of ORF3p was surmised to bind to the duplicated segment of CipA, while surface layer-like determinants would hold ORF3p on the cell surface. Both the duplicated segment of CipA and the receptor domain of ORF3p are the most divergent forms of either type of domain. It was suggested that these divergences could be the basis of a specific interaction between the duplicated segment of CipA and the receptor domain of ORF3p (7).

In this study, we tested whether the receptor domain of ORF3p was able to interact with the duplicated segment of CipA, rather than with the duplicated segment of the catalytic subunits. Chimeric proteins were constructed by grafting either the duplicated segment of endoglucanase CellD or that of CipA onto the C terminus of endoglucanase CelC, which does not possess any duplicated segment of its own. The proteins thus obtained were termed CelC-DSCelD and CelC-DSCipA, respectively. Similarly, the receptor domain of ORF3p and a representative receptor domain of CipA were each fused to the *Escherichia coli* maltose-binding protein MalE to yield MalE-RDORF3p and MalE-RDORF3p, respectively. The interaction between each receptor domain and duplicated segment was investigated (Fig. 1). In addition, CelC-DSCipA was labeled with ^125^I and used as a probe to detect *C. thermocellum* proteins interacting with the duplicated segment of CipA.

**MATERIALS AND METHODS**

**Bacterial strains, plasmids, and culture conditions.** Bacterial strains and plasmids that were used in this study are listed in Table 1. *E. coli* TG1 carrying pCT1450, pCT1701, pCT330,
or pCT332 (Table 1) was grown in Luria-Bertani medium supplemented with carbenicillin (100 μg/ml) (1). *C. thermocellum* NCIB 10682 was grown anaerobically with gentle stirring at 60°C in complete CM3-3 medium (18) containing either 10 g of cellulose MN300 (Macherey, Nagel and Co., Düren, Federal Republic of Germany) or 5 g of cellulose (Fluka AG, Buchs, Switzerland) per liter.

**DNA manipulations.** Restriction enzymes were used as recommended by the suppliers. Unless otherwise stated, DNA manipulations were done as described previously (2).

Oligonucleotides were synthesized with a Cyclone Plus synthesizer (Milligen/Millipore Corp.). Hybridization of oligonucleotides (20-mers) was performed with 1 μg of each oligonucleotide in 20 μl of 100 mM Tris-HCl buffer (pH 7.5) containing 50 mM MgCl₂. Denaturation was performed at 70°C (5 min), and hybridization proceeded in a heating block that was allowed to cool from 42°C to room temperature.

PCR amplification was performed with 6 pmol of each primer, 200 μM each deoxynucleoside triphosphate, and 10 ng of cloned *cipA* DNA (6) as a template. Taq polymerase (2.5 U; New England Biolabs) was used in a total volume of 100 μl of buffer supplied by the manufacturer. PCR was carried out for 25 cycles; the parameters were as follows: annealing, 1 min at 55°C; elongation, 2 min at 72°C; and denaturation, 1 min at 95°C. The PCR product was electrophoresed from a polyacrylamide gel as described previously (2).

The modified or amplified sequences were confirmed by double-stranded DNA sequencing with a Sequenase kit (U.S. Biochemicals).

**Construction of pCT1701.** An *SphI* site and a termination codon were inserted into the pMal-cRI vector by cloning of an appropriate oligonucleotide duplex (Fig. 2) between the *SalI* and the HindIII sites of the polylinker, yielding pMal-*SphI*. The 698-bp *EcoRI-SphI* fragment encoding the receptor domain of ORF3p (nucleotides 8109 to 8761 of the sequence given by Fujino et al. [7]) was subcloned into pMal-*SphI*. The reading frame of the insert was restored by filling in of the *SalI* site with the Klenow fragment of DNA polymerase, yielding pCT1701.

**Construction of pCT1450.** The 495-bp *Sau3A* fragment encoding the seventh receptor of *cipA* (nucleotides 6232 to 6727 of the sequence given by Gerngross et al. [8]) was cloned at the *BamHI* site of pTrc99C. A termination codon was created downstream from the insert by filling in of the *XbaI* site of the polylinker with the Klenow fragment of DNA polymerase. The fragment was cloned in pMal-cRI by use of the *EcoRI* and *SalI* sites, which bracket the insert and the stop codon. The reading frame of the insert was restored by deletion of the *EcoRI* site with mucn bean nuclease (Fig. 2), yielding pCT1450.

**Construction of pCT332.** A 356-bp fragment encoding the C-terminal duplicated segment of CipA and carrying artificial *XhoI* and HindIII sites was synthesized by PCR (Fig. 3). The forward primer was 5′-CAG GCT CCA ACT CGA GTG TGG GTA GGA-3′, and the reverse primer was 5′-CCC CCT TCA AAA GCT TAT GTA TAG CAC-3′. The fragment was inserted between the *XhoI* and HindIII sites of pCT329 to yield pCT331. The *DraI-SmaI* and *SmaI-HindIII* fragments of the *celC-DScipA* gene of pCT331 were cloned between the *SmaI* and HindIII sites of pUC18, yielding pCT332 (Fig. 3). Hence, in this plasmid, the fifth codon of *celC* is fused in frame with the *lacZ* gene.

**Protein purification.** Unless otherwise stated, all operations were performed at 4°C. Protein concentrations were measured.

---

**TABLE 1.** Bacterial strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant feature</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>TG1</td>
<td>9</td>
</tr>
<tr>
<td><em>Clostridium thermocellum</em></td>
<td>NCIB 10682</td>
<td></td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC18</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>pTrc99C</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>pMal-cRI</td>
<td></td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>pCT1701</td>
<td>pMal-cRI derivative that encodes the chimeric protein MalE-RDORF3p</td>
<td>This work</td>
</tr>
<tr>
<td>pCT1450</td>
<td>pMal-cRI derivative that encodes the chimeric protein MalE-RDCipA</td>
<td>This work</td>
</tr>
<tr>
<td>pCT306</td>
<td>pUC9 derivative carrying the wild-type <em>celC</em> gene</td>
<td>15</td>
</tr>
<tr>
<td>pCT329</td>
<td>pCT306 derivative in which the stop codon of <em>celC</em> was mutated to an <em>XhoI</em> site</td>
<td>19</td>
</tr>
<tr>
<td>pCT330</td>
<td>pCT329 derivative that encodes the chimeric protein CelC-DScipA</td>
<td>19</td>
</tr>
<tr>
<td>pCT332</td>
<td>pCT329 derivative that encodes the chimeric protein CelC-DScipA</td>
<td>This work</td>
</tr>
</tbody>
</table>
The chimeric protein MalE-RDORF3p was produced from an E. coli TG1 clone carrying pCT1701. Cells (500 ml) were grown at 37°C to an optical density at 600 nm (OD<sub>600</sub>) of 0.4, and isopropyl-β-D-thiogalactoside was added to a final concentration of 0.3 mM. Cells were cultivated for an additional 2.5 h at 37°C, harvested by centrifugation, and resuspended in 50 mM Tris-HCl (pH 7.5)–150 mM NaCl–1 mM phenylmethylsulfonyl fluoride–0.01% (wt/vol) sodium azide (buffer 1). Cells were disrupted by sonication and centrifuged, yielding 250 mg of protein in 25 ml of supernatant. Aliquots of the supernatant (120 mg of protein in 12 ml) were diluted in buffer 1 to a concentration of 3 mg/ml and applied to an amylose affinity column (5 by 2.5 cm; New England Biolabs). Elution was performed with 10 mM maltose in buffer 1. The pool of MalE-RDORF3p (14 mg/500 ml of culture) was concentrated by ammonium sulfate precipitation (80% saturation) and dialyzed against 50 mM Tris-HCl (pH 7.5).

MalE-RDCCipA, produced from E. coli TG1 carrying pCT1450, was purified according to the same protocol, except that, after the concentration step, the preparation was chromatographed twice on an Ultragel AAc 44 gel filtration column (LKB) equilibrated with 50 mM Tris-HCl (pH 7.5). The protein was concentrated and dialyzed as described above. A 1-liter culture yielded 3.8 mg of MalE-RDCCipA.

CelC-DSCipA was purified from inclusion bodies produced in E. coli TG1 (pCT330) as described by Tokatlidis et al. (19). CelC-DSCipA produced from E. coli TG1 (pCT332) also formed inclusion bodies and was purified as described previously for 68-kDa CelD (19). The pool of CelC-DSCipA was concentrated by ammonium sulfate precipitation (70% saturation) and then dialyzed against 50 mM Tris-HCl (pH 7.5). A 1-liter culture yielded 5.1 mg of CelC-DSCipA.

Aliquots of the proteins were kept frozen at -80°C until used.

**Enzyme assays.** The activities of CelC, CelC-DSCipA, and...
p-nitrophenol from 176, VOL. 1994

Tokatlidis was constant amount of concentrations platform. 2.5 MalE-RDORF3p (7 after the in the to the to the the exponential phase of growth (OD600 = 1.1) (step 1). Cellulose-
grown cultures were inoculated with a 1/50 volume of an exponentially growing starter culture (OD600 = 2) in cellubiose medium. Cells were harvested after 20 h while still bound to cellulose by centrifugation of the culture (step 1). Proteins from culture supernatants were concentrated by the slow addition of 2 volumes of cold acetone. The precipitates were solubilized in 50 mM Tris-HCl (pH 7.5) (buffer 2) and dialyzed overnight against 1 liter of the same buffer. The volume of the dialyzed culture supernatant fractions was then adjusted to 1/10 of the original culture volume with buffer 2 (step 2). Pellets resulting from step 1 were washed with 0.154 M NaCl, suspended in buffer 2 (1 to 1/5 of the original culture volume), and solubilized in cellobiose-grown cells (respectively), and disrupted by sonication. The suspensions were centrifuged twice for 5 min each time at 3,000 × g (step 3). The pellet was discarded in the case of cellobiose-grown cells, whereas in the case of cellulose-grown cells it was washed with 1 M NaCl and suspended in buffer 2 to 1/10 of the original culture volume (step 6); it constituted the sonicated cellulose-bound fraction. The supernatant resulting from step 3 was centrifuged for 20 min at 40,000 × g (step 4). The 40,000 × g supernatant constituted the supernatant of sonicated cells (supernatant sonication in Fig. 4). The pellet was washed with 1 M NaCl and resuspended in buffer 1 to 1/100 of the original culture volume (step 5). The suspension constituted the cell envelope fraction.

RESULTS

Engineering and purification of chimeric proteins. Figure 2 shows the structures of the plasmids expressing the receptor domains of ORF3p and CipA. Both domains were fused to the COOH terminus of the MalE protein to facilitate purification. The CipA receptor corresponded to the seventh internally repeated element of CipA (8). The ORF3p receptor was preceded by 7 amino acids belonging to the putative signal peptide. Both receptors were flanked on the COOH-terminal side by G/P/T/S-rich linker segments present in the original proteins. MalE-RDORF3p and MalE-RDCipA were purified as 68- and 64-kDa species, respectively (Fig. 5).

C. thermocellum endoglucanase CelC was chosen as a support protein for constructs carrying the duplicated segments of CelD and CipA (Fig. 3). In both constructs, the third codon preceding the first codon of the duplicated segment was fused to the XhoI site created by mutagenizing the stop codon of celC. Both proteins could be isolated as intact forms from inclusion bodies, which provided protection against E. coli proteases (19, 20). The molecular mass of CelC-DSCipA was 49 kDa (Fig. 5). The same value was found previously for CelC-DSCelD (19). The specific activities of CelC-DSCelD

FIG. 4. Scheme of the fractionation procedure. 1, centrifugation at 3,000 × g for 5 min; 2, acetone precipitation and solubilization of the precipitate; 3, wash and suspension, sonication, and centrifugation at 3,000 × g for 5 min; 4, centrifugation at 40,000 × g; 5, wash and suspension; 6, wash and suspension (cellulose-grown cultures only).

FIG. 5. Coomassie blue-stained SDS-polyacrylamide gel displaying the different chimeric proteins that were constructed. Lanes: 1, MalERDORF3p (4 µg); 2, MalERDCipA (2 µg); 3, CelC-DSCipA (2 µg). The leftmost lane shows markers, in kilodaltons.
and CelC-DSCipA for p-NPC were 30 U/mg (19) and 14 U/mg, respectively; that of unmodified CelC was 33 U/mg (19).

**Affinity measurements.** Figure 6 shows Klotz plots of the data obtained for the binding of 

\[ 1/\alpha_i = 1/\alpha_i + K_i \]

for unmodified CelC. The association constant \( K_i \) is deduced from the equation \( 1/\alpha_i = 1/\alpha_0 + K_i \).

**Detection of proteins interacting with the duplicated segment of CipA.** A specific set of bands was revealed when 

\[ 125\text{I}-\text{labeled CelC-DSCipA} \]

was used to probe Western blots of proteins originating from various fractions of \( C.\ thermocellum \) cultures (Fig. 7A). Three major species, migrating as 170–, 116–, and 60-kDa polypeptides in SDS-PAGE, were labeled. They were termed p170, p116, and p60, respectively.

**DISCUSSION**

This study was designed to determine whether the ORF3p receptor had a higher affinity for the duplicated segment of CipA than for the duplicated segment borne by catalytic subunits of the cellulosome, such as that of CelD. Actually, it was shown that, contrary to the model proposed by Fujino et al. (7), ORF3p did not bind to the duplicated segment of CipA. However, the binding of CelC-DSCipA to a set of hitherto-unidentified receptor proteins indicated that the duplicated segment of CipA grafted onto CelC was functional. At present, it is not known whether p170, p116, and p60 are encoded by different genes or arise from posttranslational processing of the same gene product. Characterization of these proteins and cloning and sequencing of the corresponding gene(s) will answer this question and should reveal new perspectives on the architecture of the cellulosome and its attachment to the cell surface.

These data also contradict the hypothesis that the duplicated segment of CipA interacts with receptors borne by CipA molecules, thus integrating them into a multimeric structure, as was proposed by Gerngross et al. (8).
with the ORF3p receptor. Indeed, the sequence of the chosen CipA receptor appears closer to the consensus sequences of CipA receptors than the sequence of the ORF3p receptor and, similarly, the sequence of the duplicated segment of CeID is close to the consensus sequences defined for different duplicated segments. However, it was not possible to compare individual receptors in their natural context, since CipA contains nine of them. Although both types of receptors were inserted within the same context (MalE), it is difficult to assess whether their binding properties are exactly the same as those in the original proteins. Taking into account these considerations, the two binding constants appear rather similar. Both constants are on the order of $10^7$ M$^{-1}$, a reasonably high value for protein-protein interactions, particularly since the choice of partners and binding conditions (37°C, since 60°C, the physiological growth temperature of C. thermocellum, was not possible with MalE constructs) might not have been optimal.

In the absence of ORF3 mutants, it is difficult to assess the role of ORF3p in C. thermocellum. However, the binding properties of MalE-RDORF3 suggest that ORF3p probably binds to individual carbohydrates by means of their duplicated segments. Another potential clue to the function of ORF3p is its localization. It was surmised, on the basis of nucleotide sequence analysis of the ORF3 gene, that ORF3p should be located on the cell surface. This hypothesis was investigated in the work reported in the accompanying paper (16a), in which the location of ORF3p on the cell surface was studied by biochemical methods and by electron microscopy.

ACKNOWLEDGMENTS

S.S. was the recipient of a fellowship from Agence de l’Environnement et de la Maîtrise de l’Énergie under contract with the Lyonnaise des Eaux Company. 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 University of Paris 7.

REFERENCES


