Characterization of EngF from \textit{Clostridium cellulovorans} and Identification of a Novel Cellulose Binding Domain

AKIHIKO ICHIISHI,† SALAH SHEWEITA,‡ AND ROY H. DOI*

Section of Molecular and Cellular Biology, University of California, Davis, California 95616

Received 28 July 1997/Accepted 19 December 1997

The physical and enzymatic properties of noncellulosomal endoglucanase F (EngF) from \textit{Clostridium cellulovorans} were studied. Binding studies revealed that the \(K_d\) and the maximum amount of protein bound for acid-swollen cellulose were 1.8 \(\mu\)M and 7.1 \(\mu\)mol/g of cellulose, respectively. The presence of cellobiase but not glucose or maltose could dissociate EngF from cellulose. N- and C-terminally truncated enzymes showed that binding activity was located at some site between amino acid residues 356 and 557 and that enzyme activity was still present when 20 amino acids but not 45 amino acids were removed from the N terminus and when 32 amino acids were removed from the C terminus; when 57 amino acids were removed from the C terminus, all activity was lost. EngF showed low endoglucanase activity and could hydrolyze cellotetraose and cellopentaose but not cellobiose. Activity studies suggested that EngF plays a role as an endoglucanase during cellulose degradation. Comparative sequence analyses indicated strongly that the cellulose binding domain (CBD) is different from previously reported CBDS.

The degradation of cellulose by \textit{Clostridium cellulovorans} is carried out by the action of cellulosomal and noncellulosomal enzymes (2). The isolation and the sequence of the gene for endoglucanase F (EngF), a noncellulosomal enzyme, have been reported (19). The derived sequence for this enzyme lacked the duplicated sequence that typically occurs in cellulosomal enzymes and did not show the typical cellulose binding domain (CBD) sequence reported for a number of endoglucanases (22).

Further analysis of EngF was carried out to determine its biochemical properties and its putative CBD. This was done by use of deletion mutants that could define the sequence required for activity and for binding to the substrate. These studies revealed not only the sequence required for catalytic activity but also a most likely domain for binding to the substrate. These results will be useful in future studies of the possible synergistic action of cellulases produced by \textit{C. cellulovorans}.

MATERIALS AND METHODS

Bacterial strains and plasmids. Escherichia coli XL1-Blue (Stratagene, La Jolla, Calif.) was the host strain for all cloning experiments. \textit{E. coli} TB1 was used as the host strain for the production of recombinant proteins, and \textit{E. coli} C2386 was used as the host strain for the production of uracil-containing single-stranded DNAs, which were used for site-directed mutagenesis.

pMAL-EngB was used for the expression of fusion proteins between maltose binding protein (MBP) and EngB and was constructed from pMAL-c2 (New England Biolabs Inc., Beverly, Mass.) and the engB gene from \textit{C. cellulovorans} (1). pEO52V was used for the expression of EngD (12), and pEF4 was used for the expression of the fusion protein between MBP and EngF (19).

pEF4 deletion mutants were constructed from pEF4 by site-directed mutagenesis (13). For C-terminal deletion mutants, pEF4 was mutagenized with an oligonucleotide which contained a stop codon and an XhoI restriction site (Table 1). To construct N-terminal deletion mutants, pEF4 was constructed from pEF4 by use of oligonucleotide AI 005 to make a restriction site (XmnI) between MBP and EngF. pEF4 was mutagenized with an oligonucleotide that included an Smal restriction site (Table 1), and then the N-terminal part (XmnI-Smal fragments) were removed by use of XmnI and Smal. pEF55 was constructed from pEF32 with oligonucleotides AI 105 and AI 013.

General DNA procedures. DNA was manipulated by standard procedures (1, 18). Enzymatic treatments of DNA were carried out as recommended by the manufacturers. DNA fragments were recovered after electrophoresis with the QIAquick gel extraction kit (Qiagen Inc., Chatsworth, Calif.).

Protein purification. MBP fusion proteins were purified from \textit{E. coli} with an amylase column (New England BioLabs). \textit{E. coli} harboring a plasmid which encoded a MBP fusion protein was grown in LB medium (18) supplemented with 100 \(\mu\)g of ampicillin per ml, and the culture was induced to produce an MBP fusion protein by adding isopropyl-\(\beta\)-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The cells were broken by sonication, and the crude cell extracts were poured into the amylase column. After the column was washed with TES buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA [pH 7.4]), the fusion protein was eluted with TES buffer containing 10 mM maltose. To purify EngB and EngF, purified fusion proteins were cleaved with factor Xa. After dialysis with sodium phosphate buffer (pH 8.0), the cleavage mixture was passed over a Q Sepharose FF column (Pharmacia Biotech Inc., Piscataway, N.J.). EngB and EngF were then eluted with phosphate buffer (pH 8) and an NaCl gradient from 0 to 500 mM. EngD was purified from the periplasmic fraction of \textit{E. coli} TB1 harboring pEO52V according to the method of Hamamoto et al. (11).

Cellulose binding assay. For quantitative analysis, the cellulose binding assay mixture contained 1 mg of Avicel PH101, acid-swollen cellulose (ASC), or xylan and an appropriate amount of enzyme (50 \(\mu\)g/ml to 1 mg/ml [total protein]) in a final volume of 1 ml of 50 mM sodium phosphate buffer (pH 8.0). The ASC was prepared from Avicel as described previously (24). The mixture was incubated at 4°C for 1 h with slow vertical rotation. After 1 h of incubation, cellulose was removed by centrifugation and the free protein concentration ([\(P\], micromolar units) and [\(PC\)]max (micromoles per gram of cellulose) are determined by subtracting [\(P\)] from the total protein concentration. All assays were done in triplicate. Adsorption parameters were obtained by use of the equation of Sakoda and Hiromi (17), \[\frac{[P]}{[P][PC]_{max}+K_d}\], where \(K_d\) (micromolar units) and [\(PC\)]max (micromoles per gram of cellulose) are the equilibrium dissociation constant and the maximum amount of protein bound, respectively.

For qualitative analysis, assay tubes contained 0.5 mg of ASC and 25 \(\mu\)g of enzyme in 0.5 ml of PSM buffer (50 mM sodium phosphate [pH 8.0], 50 mM NaCl, 10 mM maltose). After incubation at 4°C for 1 h with vertical rotation, assay tubes were spun in a microcentrifuge to sediment the cellulose. After the PSM buffer was removed, the pellets were washed three times with 1 ml of PSM buffer. The pellets were then resuspended in 50 \(\mu\)l of sodium dodecyl sulfate (SDS) sample buffer and analyzed by SDS–12.5% polyacrylamide gel electrophoresis (PAGE).

Enzyme assay. Cellulase activity on Avicel PH101, ASC, and carboxymethyl cellulose (CMC, low viscosity, 50 to 200 cps; Sigma) was assayed by measuring liberated reducing sugars as \(\alpha\)-glucoside equivalents by the bicinchoninic acid method (9). Activity on Avicel was assayed after incubation (16 h at 37°C) of 1.8 \(\mu\)g of EngB or EngD or 6.1 \(\mu\)g of EngF with 5 mg of Avicel in 1 ml of 50 mM

Copyright © 1998, American Society for Microbiology
acetic acid (5% H2SO4 in methanol) and baked for 10 min at 100°C. For detection of
the products, the plates were sprayed with staining reagent (TLC) plates (Merck)
with a solvent system containing 1-butanol–ethanol–water (5:5:2). After 15 min of
preequilibration, 80 mg of a 0.4% CMC solution was weighed in a chamber for
small sample volumes and used for the viscometric assay. Viscometric activity was
measured by use of a Brookfield DV-III rheometer according to the method of
Garcia et al. (9). Eight milliliters of enzyme was added to the remaining CMC
solution. The spindle was removed and 80 µl of buffer containing no maltose (lanes
2, 4, and 6). The ASC pellets were suspended in SDS sample buffer, and bound
proteins were analyzed by SDS–12.5% PAGE. Lane M, molecular weight standards
(in thousands). (B) Determination of the CBD and catalytic domain of EngF.
Since EngF did bind to ASC, we decided to determine the location of the CBD of
EngF by constructing a series of EngF deletion mutants. EngF deletion mutants were
made from total protein concentration. Each data point is the mean for three replicate
samples.

RESULTS AND DISCUSSION

Adsonption assay. We had shown previously that EngF of C. cellulovorans
would bind very weakly to Avicel (19). In order to quantify its binding ability, we
measured the adsorption of recombinant EngF to Avicel, ASC, and xylan. A time
course experiment showed that 1 h was sufficient for equilibrium binding of
the enzyme to the cellulose preparations (data not shown). Furthermore, the presence of 10 mg of Avicel did not result in greater binding of the enzyme (data not shown). Figure 1 shows a typical equilibrium adsorption isotherm for the adsorption of EngF to ASC. The Kd and [PC]max for ASC were estimated to be 1.8 µM and 7.1 µmol/g of cellulose, respectively. On the other hand, an insignificant amount of EngF bound to Avicel and xylan. These results indicated that the CBD of the noncellulosomal enzyme EngF had a high affinity for amorphous cellulose. On the other hand, the CBD of C. cellulovorans CbpA, the scaffolding protein of the cellulose, showed a high affinity for crystalline cellulose (10).

The effects of soluble carbohydrates on the adsorption of EngF to ASC were examined. The EngF bound to ASC was washed with PS buffer (50 mM phosphate, 50 mM NaCl [pH 8]) that included 10 mM glucose, cellobiose, or maltose. Glucose and maltose did not cause a dissociation of EngF from ASC, but EngF was dissociated from ASC when it was washed with buffer containing 10 mM cellulbiose (Fig. 2A).

Determination of the CBD and catalytic domain of EngF. Since EngF did bind to ASC, we decided to determine the location of the CBD of EngF by constructing a series of EngF deletion mutants. EngF deletion mutants were made from pEF4, which carried genes encoding the MBP-EngF fusion protein.

Fig. 1. Equilibrium adsorption isotherm for the adsorption of EngF to ASC. Adsonption assays were done at 4°C for 1 h. After 1 h of incubation, free protein concentration was measured with the MicroBCA reagent. Bound protein concentration ([P]) was determined by subtracting free protein concentration ([P]) from total protein concentration. Each data point is the mean for three replicate samples.
proteins. All deletion mutants were fused to MBP, located at the N-terminal end (Fig. 3). MBP itself showed very weak binding to ASC. All MBP-EngF deletion mutants also showed weak binding to ASC. We eliminated the weak binding effect of MBP by using a buffer that included maltose and sodium chloride (50 mM sodium phosphate buffer [pH 8.0], 10 mM maltose, 50 mM NaCl). This buffer did not affect the binding of EngF to ASC (Fig. 2). The binding of complete EngF and EngF from deletion mutants (Fig. 2B) indicated that complete EngF and the truncated protein produced by pEF32 (Fig. 3) were able to bind to cellulose in the presence of maltose, whereas the protein produced by pEF7 (Fig. 3) was unable to bind to cellulose in the presence of maltose. These results supported our conclusion that pEF7 lacked the CBD.

A summary of adsorption assays with a series of EngF deletion mutants is shown in Fig. 3. N-terminal deletion mutants were still able to bind to ASC when 355 amino acids had been removed. When 400 amino acids were deleted from the N-terminal end of EngF, the protein produced by this deletion mutant (pEF33) lost its ability to bind to ASC. On the other hand, when we deleted only 32 amino acids from the C terminus (pEF15 and pEF35), the deletion protein could not bind to ASC. The ability of the protein produced by pEF32 to bind to cellulose strongly indicated that the CBD of EngF existed at the C-terminal end of the molecule, between residues 356 and 557.

We also determined the catalytic domain of EngF by using a series of N- and C-terminal deletion mutants and CMC plate assays (Fig. 3). When 20 amino acids were deleted from the N terminus of EngF (pEF21), there was a slight drop in enzyme activity. Deletion of 45 amino acids from the N terminus (pEF22) resulted in a total loss of enzyme activity, indicating that some portion of the enzyme between 57 and 109 amino acids from the C terminus was required for enzyme activity. Therefore, the catalytic site was found between amino acids 50 and 500.

Proteins produced by pEF8 and pEF15 had enzyme activity but did not have binding ability. The protein produced by pEF22 retained binding ability but did not have CMCase activity. These results indicated that EngF contained two domains, the CBD and the catalytic domain. Partially homologous endoglucanases, including CelF from Bacillus sp. strain 1139 and endoglucanase from Bacillus sp. strain KSM-635, could have deletions of up to 404 or 584 amino acids from the C-terminal end, respectively (6, 15, 16), without a loss of activity. These amino acids correspond to amino acid 389 of EngF (Fig. 4).

Comparison of EngF activity to the activities of other C. cellulovorans glucanases. EngF had very weak activity on Avicel, and the activity on ASC of EngF was 100 times lower than that of EngB or EngD. The specific activities of EngF on Avicel, ASC, and CMC were 0.08, 1.7, and 9.6 μmol of glucose/min/μmol of protein, respectively. The CMCase activity of EngF was 150 times lower than that of EngB (3) or 30 times lower than that of EngD (11), but viscometric analyses of CMC hydrolysis indicated that EngF was a true endoglucanase (Fig. 5). EngB and EngD had similar activities on ASC. The activity on Avicel of EngD was higher than that of EngB, but the CMCase activity of EngD was lower than that of EngB. These results suggested that EngF played a secondary role in cellulose degradation.

Analysis of cello-oligosaccharide hydrolysis by TLC. EngF could not hydrolyze cellobiose but did hydrolyze cello-oligosaccharides. When 109 amino acids from the C terminus (pEF13) resulted in a
points (data not shown). EngB and EngD were much more active on cellodextrins than EngF, according to product yields and the total hydrolysis of cellotetraose and cellopentaose. EngD produced glucose from all three substrates, but EngB and EngF did not produce glucose from any of the substrates. This result may have been due to transglycosylation, which has been reported for a homologous enzyme from a Bacillus species (7). Thus, EngF differed from EngB and EngD in specific activities, in substrates degraded, and in the products that were formed.

Homology of EngF with other glucanases. The amino acid sequence of EngF showed high homology with those of CelA from Clostridium josui (5), endoglucanase from Bacillus sp. strain 22-28 (14), CelC from Bacillus sp. strain N-4 (8), CelF from Bacillus sp. strain 1139 (7), endoglucanase from Bacillus sp. strain KSM-64 (20), and endoglucanase from Bacillus sp. strain KSM-635 (16). Alignments of the C-terminal regions of these enzymes are shown in Fig. 4. These enzymes are members of family 5, and EngF may be a retaining enzyme, since six enzymes of family 5 screened for their stereochemistry were retaining enzymes (23).

CBDs, which have been found not only in cellulases but also in xylanases, have been grouped into 10 families, families I to X, on the basis of amino acid sequence homology (22). Previously reported enzymes, such as CelC from Bacillus sp. strain 22-28 (14), CelF from Bacillus sp. strain 1139 (7), endoglucanase from Bacillus sp. strain KSM-64 (20), and endoglucanase from Bacillus sp. strain KSM-635 (16), had not been included in the search for homologous sequences was done with a BLAST search of GenBank. Retrieved sequences were aligned with the multiple-alignment program Clustal W (21). Alignment was then edited manually for maximal fit. CC, EngF from C. cellulovorans (GenBank accession no. U37056) (19); CJ, CelA from C. josui (D85526); 22-28, endoglucanase from Bacillus sp. strain 22-28 (D85236); N-4, CelC from Bacillus sp. strain N-4 (M25500) (8); 1139, CelF from Bacillus sp. strain 1139 (D00066 and N00066) (7); KSM-64, endoglucanase from Bacillus sp. strain KSM-64 (M84963) (20); KSM-635, endoglucanase from Bacillus sp. strain KSM-635 (M27420) (16).

**FIG. 4.** Comparison of the amino acid sequence of the CBD region of EngF from C. cellulovorans with putative CBD regions of endoglucanases from other organisms. Identical amino acids are indicated with an asterisk, and similar amino acids are indicated with a plus sign (both above and below sequences); gaps are indicated with a dash. The numbers on the left indicate the amino acid positions in the proteins. Vertical arrows and letters indicate termini of the following proteins: A, C-terminal ends of truncated CelF (Bacillus sp. strain 1139; L-404) and endoglucanase (Bacillus sp. strain KSM-635; V-584); B, N-terminal end (G-401) of pEF33; and C, C-terminal end (I-525) of pEF35. The search for homologous sequences was done with a BLAST search of GenBank. Retrieved sequences were aligned with the multiple-alignment program Clustal W (21). Alignment was then edited manually for maximal fit. CC, EngF from C. cellulovorans (GenBank accession no. U37056) (19); CJ, CelA from C. josui (D85526); 22-28, endoglucanase from Bacillus sp. strain 22-28 (D85236); N-4, CelC from Bacillus sp. strain N-4 (M25500) (8); 1139, CelF from Bacillus sp. strain 1139 (D00066 and N00066) (7); KSM-64, endoglucanase from Bacillus sp. strain KSM-64 (M84963) (20); KSM-635, endoglucanase from Bacillus sp. strain KSM-635 (M27420) (16).

**TABLE 2.** TLC analysis of hydrolysis products

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Product(s) formed on the following substrate&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>EngB</td>
<td>(2G) 2G, 3G</td>
</tr>
<tr>
<td>EngD</td>
<td>(1G), (2G) 1G, 2G, 3G</td>
</tr>
<tr>
<td>EngF</td>
<td>None (2G) 2G, 3G</td>
</tr>
</tbody>
</table>

<sup>a</sup> The reaction mixture contained 0.36 μg of either EngB or EngD or 0.61 μg of EngF and cellotriose, cellotetraose, or cellopentaose in 50 mM sodium acetate buffer (pH 5.0). The reaction was carried out at 37°C for 14 h.

<sup>b</sup> 1G, glucose; 2G, cellobiose; 3G, cellotriose. Products in parentheses showed faint but significant spots on TLC plates.
those families, nor was the existence of CBDs analyzed or reported for these enzymes. We suggest that these enzymes, which have high homology to EngF in their C-terminal regions, contain CBDs and that some or all of the highly conserved C-terminal sequences comprise a new type of CBD family.

ACKNOWLEDGMENTS

We thank David Johnston and Sharon P. Shoemaker for assistance in the viscometric assays, Jae-Soon Park for the pMAL-EngB construct, and Laercio M. Malburg, Jr., for useful discussions and suggestions. We thank Seikagaku America, Inc., for providing us with cellulose.

This work was supported in part by grant 94-37308-0399(1) from the U.S. Department of Agriculture and by grant DE-FG03-92ER20069 from the U.S. Department of Energy.

REFERENCES