JOURNAL OF BACTERIOLOGY, Jan. 2003, p. 391–398 Vol. 185, No. 2
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.2.391–398.2003

CelI, a Noncellulosomal Family 9 Enzyme from *Clostridium thermocellum*, Is a Processive Endoglucanase That Degrades Crystalline Cellulose

Rachel Gilad,[1] Larisa Rabinovich,[1] Sima Yaron,[1] Edward A. Bayer,[2] Raphael Lamed,[3] Harry J. Gilbert,[4] and Yuval Shoham[1]*

Department of Food Engineering and Biotechnology and Institute of Catalysis Science and Technology, Technion—Israel Institute of Technology, Haifa 32000,[1] Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100,[2] and Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv 69978,[3] Israel, and Department of Biological and Nutritional Sciences, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne NE1 7RU, United Kingdom[4]

Received 5 August 2002/Accepted 21 October 2002

The family 9 cellulase gene *celI* of *Clostridium thermocellum*, was previously cloned, expressed, and characterized (G. P. Hazlewood, K. Davidson, J. I. Laurie, N. S. Huskisson, and H. J. Gilbert, J. Gen. Microbiol. 139:307–316, 1993). We have recloned and sequenced the entire *celI* gene and found that the published sequence contained a 53-bp deletion that generated a frameshift mutation, resulting in a truncated and modified C-terminal segment of the protein. The enzymatic properties of the wild-type protein were characterized and found to conform to those of other family 9 glycoside hydrolases with a so-called theme B architecture, where the catalytic module is fused to a family 3c carbohydrate-binding module (CBM3c); CelI also contains a C-terminal CBM3b. The intact recombinant CelI exhibited high levels of activity on all cellulosic substrates tested, with pH and temperature optima of 5.5 and 70°C, respectively, using carboxymethylcellulose as a substrate. Native CelI was capable of solubilizing filter paper, and the distribution of reducing sugar between the soluble and insoluble fractions suggests that the enzyme acts as a processive cellulase. A truncated form of the enzyme, lacking the C terminal CBM3b, failed to bind to crystalline cellulose and displayed reduced activity toward insoluble substrates. A truncated form of the enzyme, in which both the cellulose-binding CBM3b and the fused CBM3c were removed, failed to exhibit significant levels of activity on any of the substrates examined. This study underscores the general nature of this type of enzymatic theme, whereby the fused CBM3c plays a critical accessory role for the family 9 catalytic domain and changes its character to facilitate processive cleavage of recalcitrant cellulose substrates.

Clostridium thermocellum is an anaerobic thermophilic bacterium, known for its efficient degradation of cellulose and other plant cell wall polysaccharides (6). To this end, the cellulase system of this bacterium includes a remarkable variety of enzymes, some existing in the free state but most associated with a multienzyme complex known as the cellulosome (3, 5, 7, 11, 23, 29, 33).

Cellulases are modular enzymes, which possess a catalytic domain that hydrolyzes the β-1,4-glucosidic bond of the cellulose chain (9, 13, 14, 16, 35). The enzymes are categorized into 13 of the 85 sequence-based glycoside hydrolase families (17–19). Glycoside hydrolases that act on high-molecular-weight polymers are usually multimodular enzymes that contain one or more noncatalytic domains that potentiate or modify the primary hydrolytic action of the enzyme.

Thus far, more than 20 different cellulosome-associated enzymes from *C. thermocellum* have been sequenced, including cellulases, xylanases, mannanases, a chitinase, and a lichenase from 10 different glycoside hydrolase families. All of the cellulosomal enzymes contain a specialized accessory module, the dockerin domain, which effects incorporation of the parent enzyme into the complex. The enzymes are attached to the cellulosome via interaction of the dockerin with a complementary cohesin domain, borne in multiple copies by a scaffolding protein. The latter protein also contains a family 3a carbohydrate-binding module (CBM), which binds cellulose and delivers the cellulosome and its complement of enzymes to the insoluble substrate. Three different types of noncellulosomal plant cell wall-degrading enzymes have been described: those that have their own cellulose-binding CBM that delivers the enzyme to its substrate, those that carry an S-layer homology domain that anchors the enzyme onto the bacterial cell surface, and simple enzymes that are free of accessory domains.

Previously, a noncellulosomal *C. thermocellum* enzyme, CelI, and its respective gene, *celI*, were characterized (15). The recombinant enzyme showed endoglucanase activity but displayed no activity against crystalline cellulose. The sequence indicated that CelI contains a family 9 catalytic domain at its N terminus, followed by two very similar family 3 CBMs. At the time, the role of the C terminal CBMs was unclear, although it was acknowledged that they exhibited sequence similarity to modules that bound cellulose. Subsequent studies have shown that family 9 cellulases can be fused to CBM belonging to a

* Corresponding author. Mailing address: Department of Food Engineering and Biotechnology, Technion—Israel Institute of Technology, Haifa 32000, Israel. Phone: (972) 4-829-3072. Fax: (972) 4-832-0742. E-mail: yshoham@tx.chnion.ac.il.

subfamily of family 3, designated CBM3c. The crystal structure of an example of a family 9 cellulase, the endoglucanase E4 from *Thermobifida fusca* (28), revealed that the CBM3c is fused to the catalytic domain, suggesting a close functional interrelationship between the two domains. Instead of binding directly to crystalline cellulose, CBM3c was proposed to play a helper role by feeding individual carbohydrate chains of the substrate into the catalytic site (2, 4, 28). The CBM at the C terminus was of the conventional variety (i.e., from subfamily 3b), designed to bind the enzyme to the surface of the cellulose substrate (33, 36).

Subsequent biochemical evidence for endoglucanase E4 (20, 34) and analogous enzymes (CelG) from *Clostridium cellulolyticum* (12), CenB from *Cellulomonas fimi* (25, 26), and CelZ from *Clostridium stercorarium* (27) indicated that such enzymes can indeed efficiently hydrolyze crystalline cellulose. These results, however, were in direct contradiction to the original data for CelI of *C. thermocellum*, which displayed no activity against crystalline cellulose. The question remained of why the catalytic domain of CelI should be fused to CBM3c but fail to hydrolyze crystalline cellulose. Inspection of the original sequence of CelI revealed 19 residues in the C terminal CBM that were different in character from all of the other family 3 CBMs, suggesting that the *celI* sequence may be incorrect.

During the course of our studies of the CBMs from *C. thermocellum* YS, we have recloned and resequenced the *celI* gene. We found that the published sequence (from strain *C. thermocellum* NCIMB 10682) (15) contains a deletion, which accounts for the loss of sequence similarity between the C terminal CBM of CelI and the other family 3 CBMs. In the present work, the intact, amended CelI was overexpressed, purified, and characterized. The enzyme was found to hydrolyze crystalline cellulose substrates (Avicel and filter paper) efficiently as well as phosphoric acid-swollen cellulose (PASC). These results imply that the defective region of the original *celI* clone affects the binding of the enzyme to crystalline cellulose and/or interferes with the function of CBM3c in guiding a specified cellulose chain of the substrate to the active-site residues of the catalytic domain.

MATERIALS AND METHODS

Bacterial strains and vectors. *C. thermocellum* YS was isolated at the General Electric CRD Center (Schenectady, N.Y.) from soil samples at the hot springs of Yellowstone National Park (1, 24). *Escherichia coli* strains were XL1 blue (Stratgene, La Jolla, Calif.) for general cloning and BL21(λDE3) (Novagen, Madison, Wis.) for expression via the T-7 RNA polymerase system. pET9d (Novagen) was the T-7 expression vector.

DNA manipulation. Chromosomal DNA from *C. thermocellum* YS was prepared by the method of Marmur as described by Johnson (22). About 30 mg of DNA was obtained from 5 g (wet weight) of cells. The *celI* gene and its truncated forms were cloned via PCR with primers that allow its insertion into the pET9d expression vector. The primers were designed based on the sequence published by Hazelwood et al.(15). For the intact *celI* gene the 5′ primer was 5′-ATTGA **CCATGG**AAAATAAAACGGC-′3 (the *Nco*I site is in bold) and the 3′ primer was 3′-CATGCGTTGTACCGATTCTAT**CCTAGG**ACGTTA-′5 (the *Bam*HI site is in bold). For the two truncated forms of the gene, the C terminal primers were 3′-TACATGTTCGAAATACCG**CTCGAG**ATCGAGCATGCC**CCTAGG** CGC-′5 (the *SacI* and *Bam*HI sites are in bold) for *celI*-t1 and 3′-CCCCACGA CCATAAACCTAT**CCTAGG**CCCGAG-′5 (the *Bam*HI site is in bold) for *celI*-t2.

Enzymatic assays. Reactions were carried out at 55°C in 50 mM succinate buffer (pH 5.6)–10 mM $CaCl_2$. To determine the effect of pH, succinate or cacodylate buffers were used. The release of reducing power was monitored by the bicinchoninate (BCA) assay using cellobiose as a standard (10). The cellulolytic substrates were phosphoric acid-swollen cellulose (PASC) carboxymethylcellulose (CMC), microcrystalline cellulose (Avicel), and filter paper (Whatman no. 1). The concentrations in the reactions of PASC, CMC, and Avicel were 2.5, 11, and 0.1 mg/ml, respectively. The reaction mixture with filter paper contained 20 mg of paper in a 2-ml reaction volume. Enzyme concentrations were 3.8, 9.7, and 250 μg/ml for CelI, CelI-t1, and CelI-t2, respectively. One unit of activity was defined as the amount of enzyme that released 1 μmol of cellobiose reducing equivalent per min.

Viscosity. Viscosity measurements were performed in an Oswald 100 viscosimeter. The reaction under standard conditions (in a mixture containing 50 mM succinate buffer [pH 5.6] and 10 mM CaCl with 1.1% CMC as the substrate) was carried out in the viscosimeter with an enzyme (CelI) concentration of 76 μg/ml. The change in viscosity and the increase in reducing sugar with time were determined. For the reducing-sugar assay, small aliquots (300 μl) were removed and placed on ice before the BCA assay was performed.

TLC. Thin-layer chromatography (TLC) was performed using silica gel 60 F254 (0.25 mm) (Merck) with *N*-butanol–ethanol–water (3:2:2) as the running solvent. The sugars were detected by charring with a yellow solution containing 120 g of $(NH_4)_2MoO_4 \cdot 4H_2O$ and 5 g of ceric ammonium nitrate in 800 ml of 10% H_2SO_4.

Phylogenetic analysis. Phylogenetic trees were generated using ClustalW (http://www2.ebi.ac.uk/clustalw/), based on the GenBank sequences (accession codes in parentheses): Family 9 glycoside hydrolases (GH9), theme A1 enzymes (plant enzymes): cellulases from *Prunus persica* (X96853), *Populus alba* (D32166), *Citrus sinensis* (AF000135), *Persea americana* (M17634), *Pinus radiata* (X96853), *Arabidopsis thaliana* (X98543), *Phaseolus vulgaris* (M57400), *Capsium annuum* (X97189), and *Lycopersicon esculentum* (U20590); theme A2 enzymes: CipV scaffoldin from *A. cellulolyticus* (AF155197); endoglucanase L from *Clostridium cellulovorans* (AF132735); and cellulase Cel9-M from *Clostridium cellulolyticum* (AF316823); theme B enzymes: Clotm CelI, endoglucanase I from *Clostridium thermocellum* (L04735 and this work); Thefu E4, endo/exoglucanase E4 from *Thermobifida fusca* (M73322); Cloce CelG, endoglucanase G from *Clostridium cellulolyticum* (M87018); Clotm CelF, endoglucanase F from *Clostridium thermocellum* (X60545); Closr CelZ, exoglucanse Z from *Clostridium stercorarium* (X55299); Clocl EngH, endoglucanase H from *Clostridium cellulovorans* (U34793); and Celfi CelB, endoglucanase B from *Cellulomonas fimi* (M64644); theme C enzymes: cellulase J from *Clostridium thermocellum* (D83704); endoglucanase D from *Clostridium thermocellum* (X04584); endoglucanase C from *Butyrivibrio fibrisolvens* (X55732); and endoglucanase Y from *Clostridium cellulovorans* (AF105330); theme D enzymes: Clotm CbhA, cellobiohydrolase A from *Clostridium thermocellum* (X80993); Psefl CelA, endoglucanase A from *Pseudomonas fluorescens* (X12570); Celfi CelC, endoglucanase C from *Cellulomonas fimi* (X57858); Strre CelI, endoglucanase I from *Streptomyces reticuli* (X65616); and Thefu E1, endoglucanase E1 from *Thermobifida fusca* (L20094); scaffoldin-borne CBM3 sequences: Clotm CipA from *C. thermocellum* (L08665); Cloce CipC from *Clostridium cellulolyticum* (CCU40345); Clocl CbpA from *Clostridium cellulovorans* (M73817); Clojo CipJ, from *Clostridium josui* (AB004845); Bacce CipBc from *Bacteroides cellulosolvens* (AF224509); and Acece CipV from *Acetivibrio cellulolyticus* (AF155197); additional enzyme-borne CBM3 sequences: Erwca CelVI from *Erwinia carotovora* (X79241); Bacla CelA from *Bacillus lautus* (P29719); Bacsu CelA from *Bacillus subtilis* (Q45532); and Closr CelY from *Clostridium stercorarium* (P50900).

RESULTS

In view of the lack of sequence similarity of the C terminal CBM of Cell to other family 3 CBMs, the complete *celI* gene was recloned and sequenced (36). The sequence of the C terminal region did not match the published sequence, and the gene appeared to contain an additional insert of 53 bp. The insert was flanked by a short direct repeat of 5 bp (TTCAA), and it is therefore likely that the original clone contained a deletion due to recombination within this direct repeat (Fig. 1). An identical sequence for the *celI* gene was recently also obtained in a different *C. thermocellum* strain (V. V. Zverlov and W. H. Schwarz, GeneBank accession number AJ275974 [2000]). Based on sequence homology, CelI comprises three elements, a family 9 catalytic domain at its N terminus, a family

A.

```
..TATTCTTTCAA//ACAAGCTTGTTTACGGCAAAGAACCTTGATTTATCTGTACGCAACATGGCTAAGATAG
..Y  S  F   K   Q  A  C  L  R  Q  R  T  L  I  Y  L  Y  A  T  W  L  R  *
```

B.

```
..TATTCTTTCAATCCGTCTGCTTCCGATTATACGGATTGGAACAGGGTGACGTTGTATATTTCAAACAAGCTTGTTTACGGCAAAGAACCTTGA
..Y  S  F  N  P  S  A  S  D  Y  T  D  W  N  R  V  T  L  Y  I  S  N  K  L  V  Y  G  K  E  P  *
```

C.

```
Clotm-CelI*   ..YSFKQACLRQRTLIYLYATWLR*
Clotm-CelI    ..YSYQGL---SAGTVVKSEYIPVYDAGVLVFGREP*
Clocl-EngH    ..FSYEGITKTPGETPVKVTNIPVYDNGVKVFGNEP*
Cloce-CelG    ..FSYDGLPTTS--TVNTVTNIPVYDNGVKVFGNEP*
Calsa-CelA    ..YSFQDIKGVSSGSVVKTKYIPLYDEDIKVWGEEP*
Closr-CelZ    ..YSFRDIKGVTSGNTVKTVYIPVYDDGVLVFGVEP*
Clotm-CelF    ..FSFQGL--EQGFTSKKTEYIPLYDGNVRVWGKVP*
Celfi-CelB    ..PSYTGL---TQTALAKASAITLYDGSTLVWGKEP*
Thefu-E4      ..WSFQGI----GNELAPAPYIVLYDDGVPVWGTAP*
```

FIG. 1. C-terminal sequence of CelI. (A) Originally published sequence of the *celI** gene (15). The direct repeat (TTCAA) is underlined; slashes (//) indicate the start point of the deletion. (B) Sequence of *celI* gene cloned in this work. (C) Sequence alignment of the two contrasting sequences of the C terminus of the CelI CBM3b with the CBMs of representative members of subfamily 3b. Identical and similar residues are shown in bold. Notation for other enzymes: Clocl-EngH, endoglucanase H from *Clostridium cellulovorans*; Cloce-CelG, endoglucanase G from *Clostridium cellulolyticum*; Calsa-CelA, cellulase A from *Caldicellulosiruptor saccharolyticus*; Closr-CelZ, exoglucanse Z from *Clostridium stercorarium*; Clotm-CelF, endoglucanase F from *Clostridium thermocellum*; Celfi-CelB, endoglucanase B from *Cellulomonas fimi*; Thefu-E4, endo/exoglucanase E4 from *Thermobifida fusca*.

3c CBM domain (fused to the catalytic domain), and a family 3b CBM at the C terminus. The assignment of these modules into family 3b and 3c is as follows. The C terminal CBM was very similar to family 3 CBMs that bind to cellulose, but the CBM adjacent to the catalytic domain lacked most of the residues that were proposed to interact directly with the cellulose surface, typical of CBM3c proteins that exhibit no direct affinity for the polysaccharides. The deletion resulted in a frameshift occurring within the long loop between β-strands 7 and 8 of CBM3b, producing a truncated molecule with 19 extraneous amino acid residues. Since the original publication of CelI suggested that the enzyme has no activity on crystalline cellulose, the finding of the potential deletion prompted us to revisit the characteristics of the intact protein as well as those of its truncated forms lacking one or both of its CBMs.

Cloning and expression of the *celI* gene and its truncated forms. The complete *celI* gene and its truncated forms were cloned via PCR, using *C. thermocellum* chromosomal DNA as the template. The PCR primers were designed to include *Nco*I and *Bam*HI sites that allowed cloning of the resulting amplification products into the T-7 expression vector pET9d, resulting in plasmids pET9d-*celI*, containing the intact gene, pET9d-*celI*-t1, containing the truncated form of the *celI* gene without the C terminal CBM3b, and pET9d-*celI*-t2, containing the *celI* gene lacking both CBMs (Fig. 2). The *celI*, *celI*-t1, and *celI*-t2 genes were 2,505, 1,884, and 1,398 bp, encoding proteins with molecular masses of 92,879, 70,491, and 52,758 Da, respectively. Once the genes were cloned into pET9d, their integrity was verified by sequencing and then they were transformed into *E. coli* BL21(DE3) cells. The genes were overexpressed by growing the cells overnight on Terrific Broth without induction, collecting the cells by centrifugation, and breaking them by two passages through a French press. Following centrifugation, the soluble fraction was heat treated (55°C for 30 min), centrifuged again to remove most of the host cell proteins, and applied to a gel filtration column. All of the proteins were over 85% pure as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Figure 3 shows the SDS-PAGE results for the proteins following purification. Interestingly, the intact CelI protein was not stable during storage at 4°C and appeared to break down with time. However, the degradation products were thermolabile and could be readily removed by an additional heat treatment (Fig. 3, lane E). Enzyme activity was unaffected even after heat treatment at 50°C for 20 h.

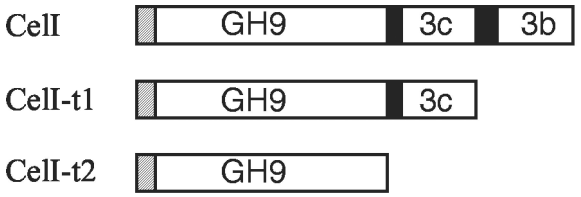

FIG. 2. Schematic diagram illustrating the modular architecture of the recombinant constructs used in this study. The 835-residue CelI represents the intact enzyme, containing an N-terminal signal sequence (hatched), followed by a fused GH9, a CBM3c, and a CBM3b. The last three modules are separated by characteristic linker segments. CelI-t1 is a 628-residue truncated form in which the cellulose-binding C-terminal CBM3b was deleted. CelI-t2 represents the 466-residue catalytic module resulting from the deletion of both CBMs.

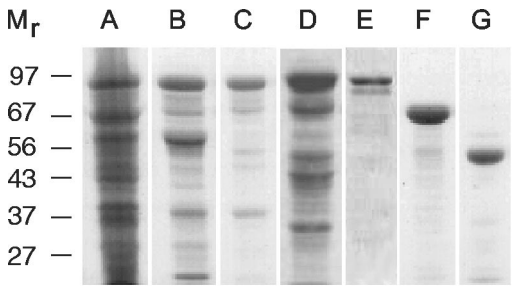

FIG. 3. Expression, purification, and maintenance of intact and truncated forms of recombinant CelI. Lanes: A, extract of *E. coli* BL21(DE3) containing pET9d(CelI); B, soluble fraction after heat treatment (55°C for 30 min); C, purified CelI; D, purified CelI after storage at 4°C; E, the same sample as in lane D, subjected to an additional heat treatment (50 to 55°C for ~20 h); F, purified CelI-t1; G, purified CelI-t2.

Biochemical characterization of CelI. The intact recombinant enzyme, CelI, exhibited typical enzyme kinetics on the soluble cellulose derivative CMC. The amount of reducing sugar liberated from the substrate was linear with time and was proportional to enzyme concentration. Similar levels of activity were obtained with CelI-t1; however, no detectable activity was observed with CelI-t2. The activity of the intact CelI exhibited a pH optimum of 5.5 at 60°C and a temperature optimum of 70°C in a 10-min assay.

The two truncated derivatives of CelI, lacking the C terminal CBM3, would not be expected to bind to crystalline cellulose. To examine whether the latter CBM is indeed required for binding to the substrate, CelI and its truncated forms were mixed with crystalline cellulose (Avicel) and the washed insoluble fraction was subjected to SDS-PAGE (Fig. 4). As can be seen from the figure, the intact CelI bound strongly to cellulose, but neither of the truncated forms were capable of binding to the substrate.

The activities of the intact CelI and its truncated forms on various cellulosic substrates, including CMC, PASC, microcrystalline cellulose (Avicel), and filter paper, were examined. For each of these substrates, the activities were determined

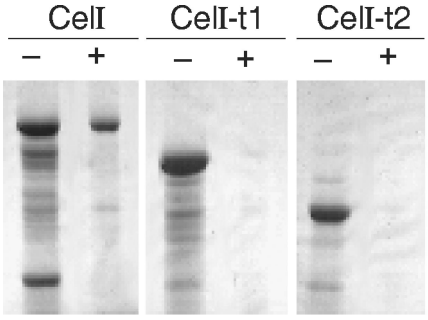

FIG. 4. Cellulose-binding properties of intact and truncated forms of CelI. Samples of CelI, CelI-t1, and CelI-t2 were subjected to interaction with microcrystalline cellulose and washed, and the retained fraction was released with SDS-containing sample buffer and subjected to SDS-PAGE. Lanes designated (−) represent the samples prior to interaction with cellulose, and those designated (+) represent the sequestered proteins. Only the intact CelI was retained on the cellulose, indicating the requirement of CBM3b for binding.

TABLE 1. Activities of CelI and its truncated forms on various cellulosic substrates

Protein	CMC		PASC		Avicel		Filter paper	
	Sp act[a]	%[b]	Sp act	%	Sp act	%	Sp act	%
CelI	1.2×10^3	100	93	100	17	100	1.0	100
CelI-t1	1.7×10^3	140	27	29	4.5	26	0.12	12
CelI-t2	ND[c]	0	ND[d]	0	0.13	0.76	ND[e]	0

[a] Specific activity is defined as the generation of cellobiose (micromoles per minute) per unit protein (in micromoles).
[b] Percentage of specific activity of intact CelI.
[c] No activity detected (<5.2 μmol of cellobiose min^{-1} μmol of $protein^{-1}$).
[d] No activity detected (<0.76 μmol of cellobiose min^{-1} μmol of $protein^{-1}$).
[e] No activity detected (<6.4×10^{-3} μmol of cellobiose min^{-1} μmol of $protein^{-1}$).

within the initial linear range of the reaction. Intact CelI exhibited relatively high levels of activity on all cellulosic substrates tested (Table 1). Under the conditions examined, the highest activity was obtained on CMC, followed by PASC, Avicel, and filter paper. TLC analysis of the soluble sugars released from Avicel indicated that the main product was cellobiose with small amounts of cellotriose and glucose (Fig. 5). CelI-t1 exhibited even higher endoglucanse activity (on CMC), but the lack of CBM3b appeared to cause highly reduced activities on the insoluble forms of the substrate. CelI-t2 (lacking both CBMs) displayed no detectable activity on CMC, PASC, and filter paper. Surprisingly, marginal but measurable activity was detected for CelI-t2 on Avicel.

Is CelI a processive endoglucanase? To examine experimentally whether CelI is a processive enzyme, we determined the specific viscosity as a function of reducing-sugar production when acting on CMC as a substrate. In addition, the ratio of reducing sugar in the soluble and insoluble fractions was assessed, using filter paper as a substrate. The hydrolysis of CMC was accompanied by a sharp decrease in the specific viscosity of the solution and a relatively small increase in the level of reducing sugar (Fig. 6). These results suggest that CelI exhibits a typical endo type of activity toward CMC. It is likely that the carboxymethyl substituent on cellulose prevents the enzyme from acting in a processive manner (21). However, when we compared the distribution of the reducing-sugar ends remain-

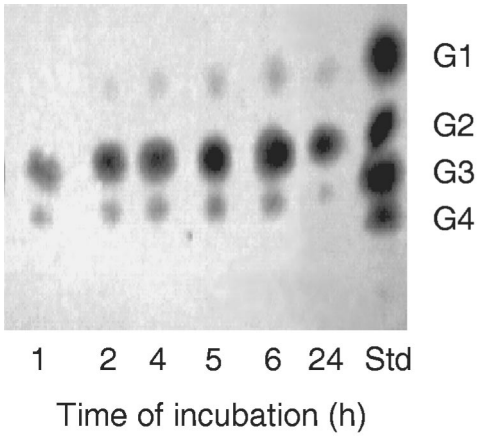

FIG. 5. TLC analysis of soluble sugars following degradation of crystalline cellulose by CelI.

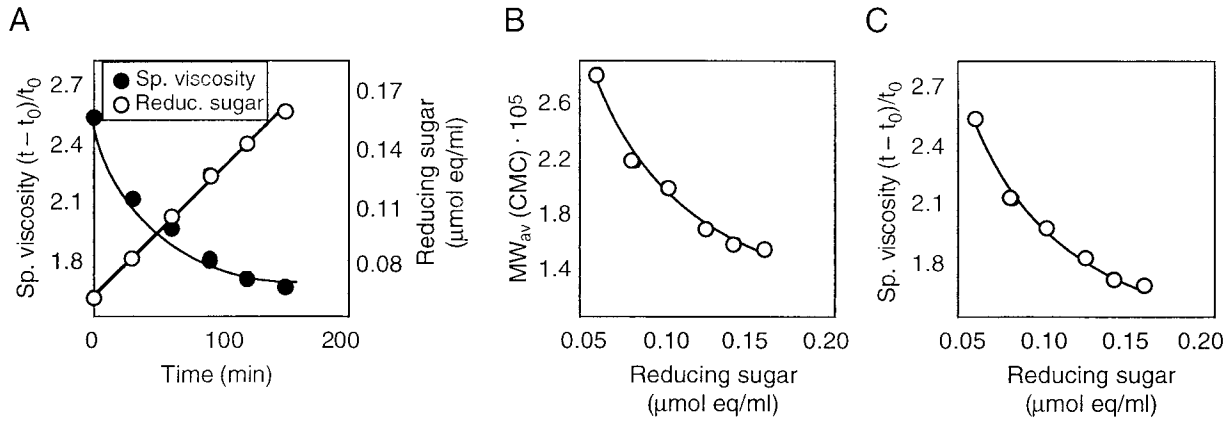

FIG. 6. Hydrolysis of CMC by CelI. (A) Release of reducing sugar and concomitant decrease in viscosity with time. (B and C) Dependence of the decrease in the molecular weight of CMC on the release of reducing sugar. Specific viscosity: $V_{sp} = (t - t_0)/t_0$, where t_0 is time of flow of the buffer and t is time of flow of the CMC mixture. The reaction was performed at 55°C, and the reaction mixture comprised 1.1% CMC in 50 mM succinate buffer (pH 5.6) containing 10 mM $CaCl_2$. The molecular weight of CMC was calculated from the equation $V_{sp} = K(M_w)^{\alpha}$, where $\alpha = 0.66$ for a random coil. K was determined according to the molecular weight of CMC prior to hydrolysis.

ing on filter paper (insoluble) with that of reducing sugar released into the soluble fraction, we found that there was only a nominal change in the reducing power of the insoluble fraction compared with the sharp increase in that of the soluble fraction. The increase in the amount of soluble reducing sugar observed during a 10-min reaction period was >50-fold. Since the increase in the amount of soluble reducing sugar was not accompanied by a similar increase associated with the insoluble substrate, in this case CelI exhibited an exo type of activity. Taken together, these results strongly suggest that CelI acts as a processive endoglucanase on insoluble substrates, exhibiting both endo (acting on CMC) and exo types of activities.

CelI alone can solublize filter paper. One of the features of cellulose degradation is the synergistic effect displayed by the combined action of endo- and exoglucanases. This property is explained by the need to generate reducing ends by endoglucanases for subsequent degradation by the exoglucanase. Since processive endoglucanases can combine the two types of activities, it is possible that such enzymes can degrade crystalline cellulose substrates on their own. To examine whether CelI can hydrolyze completely recalcitrant cellulosic substrates, we incubated the enzyme in the presence of filter paper. Under the conditions tested (1:10 enzyme-to-substrate ratio), the resulting mixture became turbid after several days of incubation (Fig. 7A). Reducing-sugar analysis indicated that over 40% of the theoretical amount of reducing sugar was in the soluble fraction. TLC analysis revealed that most of the soluble product was in the form of glucose and cellobiose (Fig. 7B).

DISCUSSION

The family 9 glycoside hydrolases (GH9) comprise one of the most extensive and versatile types of cellulases known in nature. The GH9 enzymes are widely distributed among the phyla, occurring in numerous aerobic and anaerobic bacteria, anaerobic fungi, slime molds, and various plant and animal species (http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html) (8). Some species contain numerous different genes which encode different GH9 enzymes. In *C. thermocellum*, no fewer than nine GH9 enzymes have been sequenced by conventional methods, and the impending genome sequence will undoubtedly add to this tally.

Enzymes that contain the GH9 catalytic domain can have at least four different molecular architectures (4). In this context,

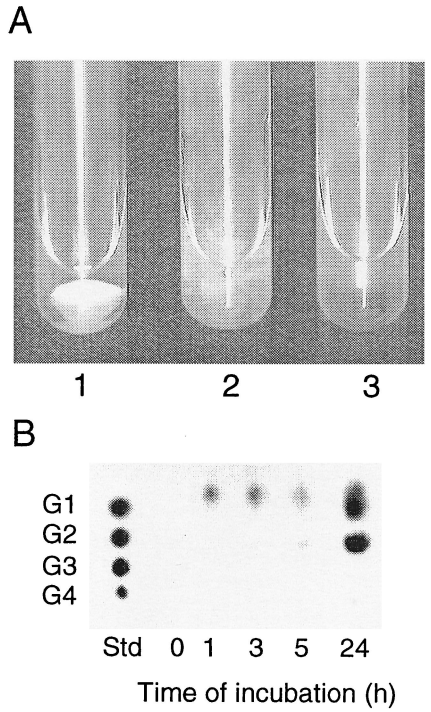

FIG. 7. Solubilization of crystalline cellulose by intact recombinant CelI. (A) Degradation of filter paper by CelI: 1, filter paper without enzyme; 2, filter paper with enzyme; 3, enzyme without paper. Approximately 40% soluble sugar was obtained after 8 days of incubation of the substrate in the presence of the enzyme. (B) TLC analysis of soluble sugars on hydrolysis of filter paper by CelI. Approximately 12% soluble sugar (cellobiose and glucose) was obtained after 24 h of incubation.

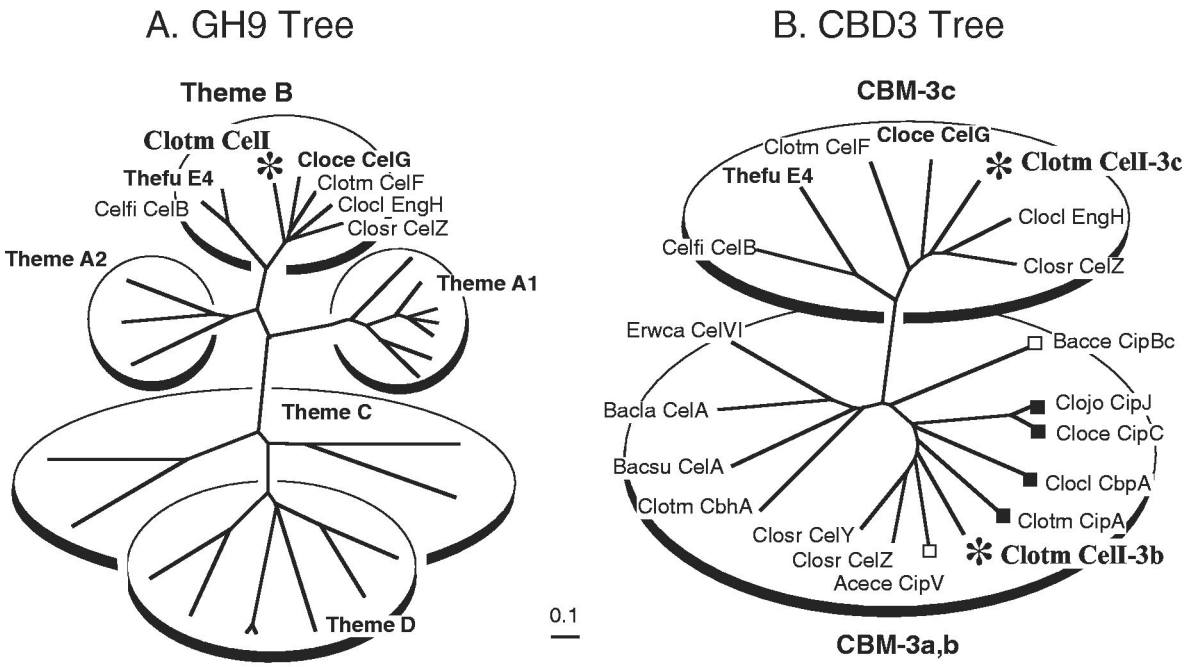

FIG. 8. Phylogenetic analysis of the three component modules of CelI, denoted by asterisks. (A) Relationship of the family 9 catalytic module of CelI with other GH9s. (B) Relationship of the two CBM3s of CelI to other CBM3s. Scaffoldin-based CBMs are indicated by squares; those from family 3a are indicated by solid symbols, and those from family 3b are indicated by open symbols. Phylogenetic trees were generated using ClustalW (http://www2.ebi.ac.uk/clustalw/?). For a precise list of the proteins and their accession numbers, consult Materials and Methods. Scale bar indicates the percentage (0.1) of amino acid substitutions.

the theme A enzymes lack an accompanying accessory module, the theme B enzymes possess a CBM3c fused to the C terminus of the GH9 catalytic domain, the theme C enzymes lack such a module but contain an N-terminal immunoglobulin domain, and the theme D enzymes contain an immunoglobulin domain but also contain at least one additional family 4 CBM.

The nature of the accessory module(s) and the molecular architecture of these enzymes are reflected in the phylogenetic relationship of the amino acid sequences of the GH9 catalytic domain itself (Fig. 8A). The theme A and C enzymes, e.g., *C. cellulolyticum* CelM and *C. thermocellum* CelD, are potent endoglucanases. In contrast, the theme B and D enzymes are known for their rather efficient degradation of insoluble cellulose substrates. It thus appears that the modular arrangement of the GH9 enzymes—namely, the addition of a fused CBM3c or of a CBM4—serves to modulate the basic catalytic properties of the GH9 catalytic domain.

The results of the present study demonstrate that *C. thermocellum* CelI is a conventional GH9 theme B cellulase. The activities of the intact recombinant enzyme and its truncated forms (whereby one or both accessory CBMs were deleted) on various soluble and insoluble cellulose substrates (Table 1) were entirely consistent with those reported for analogous enzymes, e.g., *T. fusca* E4 and *C. cellulolyticum* CelG (12, 20, 34). Interestingly, both E4 and CelG produced mainly cellotetraose in the initial hydrolysis of cellulose (12, 20). The final product of CelI was mainly cellobiose, which presumably resulted from subsequent cleavage of cellotetraose. It is thus clear that *C. thermocellum* CelI is a highly active endoglucanase. The endoglucanase activity is dependent on the presence of the fused CBM3c, since in its absence, the GH9 catalytic domain alone (CelI-t2) displays no detectable activity. In the absence of the C-terminal cellulose-binding CBM3b, however, the truncated enzyme (CelI-t1) shows enhanced activity against CMC, indicating that the cellulose-binding activity may interfere with the endoglucanase activity of the GH9-CBM3c system. Nevertheless, both family 3 CBMs are required for full activity on insoluble forms of cellulose, the CBM3b for the initial targeting and attachment to the substrate and the CBM3c as a modulating complement of the GH9 catalytic domain. It is worth noting that in the first work on CelI, the enzyme exhibited very low CMCase activity (15). This low activity could not be explained based only on the deletion in CBM3c since CelI-t1 exhibits enhanced activity against CMC. It is possible, therefore, that in the original work, CelI was damaged in additional domains, for example in CBM3c.

Phylogenetic analysis of the sequences of the family 3 CBMs indicates a clear separation between those of family 3b and 3c (Fig. 8B). The difference reflects the respective cellulose-binding face of CBM3b, which is clearly altered in CBM3c: many of the recognized cellulose-binding residues are not conserved in CBM3c, and the question thus arose whether this type of CBM would indeed act in a cellulose-binding capacity per se. Interestingly, the distribution of sequences within the CBM3c cluster closely mirrors those of the GH9 theme B counterpart (compare the distribution of the catalytic domains in Fig. 8A with that of the respective CBM3c in Fig. 8B), suggestive of the close structural relationship between the two modules.

The precise role of the fused CBM3c was proposed earlier by Sakon et al. (28) on the basis of the crystal structure of the *T. fusca* cellulase E4, which included both the GH9 catalytic domain and the fused CBM3c. The structure revealed an array

of conserved residues on the surface of the CBM3c that aligns precisely with the active site of the catalytic domain. The authors proposed that this array, which includes N470, E478, K480, R557, E559, Q561, and R563, may interact with an extended cellulose chain en route to the active-site cleft, thereby converting the GH9 theme B enzyme from a simple endoglucanase to a processive enzyme. It may also be noted that additional residues, e.g., F476 and Y520, are also uniquely conserved on the same surface of the fused CBM3c. With the possible exception of *C. thermocellum* CelF, in which some of the latter residues have been replaced, the CBM3c-specific residues associated with the other theme B enzymes (shown in Fig. 8) are strictly conserved, the inference being that the proposed function is retained in all of these enzymes.

The original defect of the *celI* clone (15) involved a frameshift which altered a relatively short segment of the molecule but resulted in a deleted portion that presumably cropped the final two β-strands at the C terminal of CBM3b. These two strands contribute to the formation of the shallow groove, located on the opposite side of the cellulose-binding plane of the module, and would not be expected to be involved directly in cellulose binding (33). Nevertheless, the resultant deformity of the molecule would not appear to be limited to local perturbations, since the original recombinant CelI displayed very little catalytic activity (15). The original construct exhibited essentially no detectable activity on insoluble recalcitrant cellulosic substrates and very low levels of activity on CMC. In the present work, correction of the sequence and expression of the amended protein resulted in an enzyme that was 150 times more active on CMC and exhibited high levels of activity on all cellulose preparations used. In fact, its truncated form (CelI-t1), which lacks the entire cellulose-binding CBM3b, was even more active on CMC, suggesting that the distorted CBM3b in the original CelI interfered with the normal function of the neighboring fused CBM3c.

In this work we produced and characterized truncated forms of a GH9 theme B cellulase, CelI from *C. thermocellum*. The catalytic characteristics of this amended enzyme are in complete accord with those previously reported for analogous cellulases from other bacteria. This study thus emphasizes the general nature of this type of enzymatic theme, whereby the fused CBM3c serves an important accessory role for the catalytic domain and changes its character to facilitate processive cleavage of recalcitrant cellulose substrates.

ACKNOWLEDGMENTS

This research was supported by the Israel Science Foundation (grants 771/01, 446/01, and 250/99). Additional support was provided by the Otto Meyerhof Center for Biotechnology, established by the Minerva Foundation, (Munich, Germany), and by funds from the Technion-Niedersachsen Cooperation (Hanover, Germany).

REFERENCES

1. **Bayer, E. A., R. Kenig, and R. Lamed.** 1983. Adherence of *Clostridium thermocellum* to cellulose. J. Bacteriol. **156:**818–827.
2. **Bayer, E. A., E. Morag, R. Lamed, S. Yaron, and Y. Shoham.** 1998. Cellulosome structure: four-pronged attack using biochemistry, molecular biology, crystallography and bioinformatics, p. 39–67. *In* M. Claeyssens, W. Nerinckx, and K. Piens (ed.), Carbohydrases from *Trichoderma reesei* and other microorganisms. The Royal Society of Chemistry, London, United Kingdom.
3. **Bayer, E. A., L. J. W. Shimon, R. Lamed, and Y. Shoham.** 1998. Cellulosomes: structure and ultrastructure. J. Struct. Biol. **124:**221–234.
4. **Bayer, E. A., Y. Shoham, and R. Lamed.** 2000. Cellulose-decomposing prokaryotes and their enzyme systems. *In* M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (ed.), The prokaryotes: an evolving electronic resource for the microbiological community, 3rd ed. (Latest update release 3.7, September 2001.) Springer-Verlag, New York, N.Y.
5. **Béguin, P., and M. Lemaire.** 1996. The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. Crit. Rev. Biochem. Mol. Biol. **31:**201–236.
6. **Béguin, P., J. Millet, and J.-P. Aubert.** 1992. Cellulose degradation by *Clostridium thermocellum*: from manure to molecular biology. FEMS Microbiol. Lett. **100:**523–528.
7. **Bhat, K. M., and T. M. Wood.** 1992. The cellulase of the anaerobic bacterium *Clostridium thermocellum*: isolation, dissociation, and reassociation of the cellulosome Carbohydr. Res. **227:**293–300.
8. **Coutinho, P. M., and B. Henrissat.** 1999. Carbohydrate-active enzymes: an integrated database approach, p. 3–12. *In* H. J. Gilbert, G. J. Davies, B. Henrissat, and B. Svensson (ed.), Recent advances in carbohydrate bioengineering. The Royal Society of Chemistry, London, United Kingdom.
9. **Davies, G., and B. Henrissat.** 1995. Structures and mechanisms of glycosyl hydrolases. Structure **3:**853–859.
10. **Doner, L. W., and P. L. Irwin.** 1992. Assay of reducing end-groups in oligosaccharide homolgues with 2,2′-bicinchoninate. Anal. Biochem. **202:**50–53.
11. **Felix, C. R., and L. G. Ljungdahl.** 1993. The cellulosome—the exocellular organelle of *Clostridium*. Annu. Rev. Microbiol. **47:**791–819.
12. **Gal, L., C. Gaudin, A. Belaich, S. Pagès, C. Tardif, and J.-P. Belaich.** 1997. CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose. J. Bacteriol. **179:**6595–6601.
13. **Gilbert, H. J., and G. P. Hazlewood.** 1993. Bacterial cellulases and xylanases. J. Gen. Microbiol. **139:**187–194.
14. **Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. J. Miller, and R. A. J. Warren.** 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55:**303–315.
15. **Hazlewood, G. P., K. Davidson, J. I. Laurie, N. S. Huskisson, and H. J. Gilbert.** 1993. Gene sequence and properties of CelI, a family E endoglucanase from *Clostridium thermocellum*. J. Gen. Microbiol. **139:**307–316.
16. **Henrissat, B., and P. Bork.** 1996. On the classification of modular proteins. Protein Eng. **9:**725–726.
17. **Henrissat, B., and P. M. Coutinho.** 2001. Classification of glycoside hydrolases and glycosyltransferases from hyperthermophiles. Methods Enzymol. **330:**183–201.
18. **Henrissat, B., and G. J. Davies.** 2000. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. Plant Physiol. **124:**1515–1519.
19. **Henrissat, B., and A. Romeu.** 1995. Families, superfamilies and subfamilies of glycosyl hydrolases. Biochem. J. **311:**350–351.
20. **Irwin, D., D.-H. Shin, S. Zhang, B. K. Barr, J. Sakon, P. A. Karplus, and D. B. Wilson.** 1998. Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. J. Bacteriol. **180:**1709–1714.
21. **Irwin, D. C., S. Zhang, and D. B. Wilson.** 2000. Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. Eur J Biochem. **267:**4988–4997.
22. **Johnson, J. L.** 1981. Genetic characterization, p. 243–265. *In* P. Gerhardt, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg, and G. B. Phillips (ed.), Manual of methods for general bacteriology. American Society for Microbiology, Washington, D.C.
23. **Lamed, R., and E. A. Bayer.** 1988. The cellulosome of *Clostridium thermocellum*. Adv. Appl. Microbiol. **33:**1–46.
24. **Lamed, R., E. Setter, and E. A. Bayer.** 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. J. Bacteriol. **156:**828–836.
25. **Meinke, A., C. Braun, N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., and R. A. Warren.** 1991. Unusual sequence organization in CenB, an inverting endoglucanase from *Cellulomonas fimi*. J. Bacteriol. **173:**308–314.
26. **Meinke, A., N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., and R. A. Warren.** 1991. Multiple domains in endoglucanase B (CenB) from *Cellulomonas fimi*: functions and relatedness to domains in other polypeptides. J. Bacteriol. **173:**7126–7135.
27. **Riedel, K., and K. Bornnenmeier.** 1999. Active-site mutations which change the substrate specificity of the *Clostridium stercorarium* cellulase CelZ. Eur. J. Biochem. **262:**218–223.
28. **Sakon, J., D. Irwin, D. B. Wilson, and P. A. Karplus.** 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nat. Struct. Biol. **4:**810–818.
29. **Schwarz, W. H.** 2001. The cellulosome and cellulose degradation by anaerobic bacteria. Appl. Microbiol. Biotechnol. **56:**634–649.
30. **Shimon, L. J., S. Pages, A. Belaich, J. P. Belaich, E. A. Bayer, R. Lamed, Y. Shoham, and F. Frolow.** 2000. Structure of a family IIIa scaffoldin CBD from the cellulosome of *Clostridium cellulolyticum* at 2.2 A resolution. Acta Crystallogr. Ser. D **56:**1560–1568.

31. **Shoham, Y., R. Lamed, and E. A. Bayer.** 1999. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microbiol. **7:**275–281.
32. **Tomme, P., R. A. J. Warren, and N. R. Gilkes.** 1995. Cellulose hydrolysis by bacteria and fungi. Adv. Microb. Physiol. **37:**1–81.
33. **Tormo, J., R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham, and T. A. Steitz.** 1996. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J. **15:** 5739–5751.
34. **Wilson, D. B., D. Irwin, J. Sakon, and P. A. Karplus.** 1998. *Thermomonospora fusca* cellulase E4: A processive endocellulase, p. 133–138. *In* M. Claeyssens, W. Nerinckx, and K. Piens (ed.), Carbohydrases from *Trichoderma reesei* and other microorganisms. The Royal Society of Chemistry, London, United Kingdom.
35. **Wilson, D. B., and D. C. Irwin.** 1999. Genetics and properties of cellulases. Adv. Biochem. Eng. **65:**1–21.
36. **Yaron, S.** 1999. Ph.D. thesis. Technion—Israel Institute of Technology, Haifa, Israel.