JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 308–314
0021-9193/91/010308-07$02.00/0

Vol. 173, No. 1

Unusual Sequence Organization in CenB, an Inverting Endoglucanase from *Cellulomonas fimi*

A. MEINKE,[1,2] C. BRAUN,[3] N. R. GILKES,[1] D. G. KILBURN,[1] R. C. MILLER, JR.,[1] AND R. A. J. WARREN[1]*

Departments of Microbiology[1] and Chemistry,[3] University of British Columbia, 300-6174 University Boulevard, Vancouver, British Columbia, Canada V6T 1W5, and Institut für Biologie III der Universität Freiburg, D-7800 Freiburg im Breisgau, Federal Republic of Germany[2]

Received 2 July 1990/Accepted 22 October 1990

The nucleotide sequence of the *cenB* gene was determined and used to deduce the amino acid sequence of endoglucanase B (CenB) of *Cellulomonas fimi*. CenB comprises 1,012 amino acids and has a molecular weight of 105,905. The polypeptide is divided by so-called linker sequences rich in proline and hydroxyamino acids into five domains: a catalytic domain of 607 amino acids at the N terminus, followed by three repeats of 98 amino acids each which are >60% identical, and a C-terminal domain of 101 amino acids which is 50% identical to the cellulose-binding domains of *C. fimi* cellulases Cex and CenA. A deletion mutant of the *cenB* gene encodes a polypeptide lacking the C-terminal 333 amino acids of CenB. The truncated polypeptide is catalytically active and, like intact CenB, binds to cellulose, suggesting that CenB has a second cellulose-binding site. The sequence of amino acids 1 to 461 of CenB is 35% identical, with a further 15% similarity, to that of a cellulase from avocado, which places CenB in cellulase family E. CenB releases mostly cellobiose and cellotetraose from cellohexaose. Like CenA, CenB hydrolyzes the β-1,4-glucosidic bond with inversion of the anomeric configuration. The pH optimum for CenB is 8.5, and that for CenA is 7.5.

Many cellulases appear to be composed of discrete domains (1a, 2, 17, 33). The functions of the domains are known for some of them (9, 11, 27, 30, 32) and inferred for others on the basis of sequence similarities to domains of known function (1a, 2). A frequent motif is a cellulose-binding domain (CBD) connected to a catalytic domain by a so-called linker sequence rich in proline and/or hydroxyamino acids. The CBD can be at the N or C terminus of a particular enzyme. Sequence similarities between catalytic domains allow the cellulases to be grouped into families (1a, 2, 15). A particular microorganism may produce cellulases belonging to several families, and some families include enzymes from bacteria, fungi, and plants. Sequence similarities are also seen in the CBDs.

The gram-positive bacterium *Cellulomonas fimi* uses cellulose as a carbon source. Several of its cellulase genes were cloned and expressed in *Escherichia coli* (8). Two of the genes, *cex* and *cenA*, encoding exoglucanase Cex and endoglucanase CenA, respectively, were sequenced (23, 35), and the proteins they encode were characterized in some detail (9–11, 18, 19). Both enzymes comprise a CBD of about 110 amino acids joined to a catalytic domain of 280 to 300 amino acids by a linker of about 20 amino acids containing only proline and threonine, the Pro-Thr box. The Pro-Thr boxes are virtually identical and the CBDs are about 50% identical, but the catalytic domains share no identity and belong to different families (15, 23, 33, 35). The CBD is at the N terminus of CenA but the C terminus of Cex. The CBDs and catalytic domains retain their functions after separation by proteolysis (9, 11, 18).

A third *C. fimi* cellulase gene, *cenB*, was also cloned (8), and the enzyme it encodes was identified as endoglucanase CenB (10). Like Cex and CenA, CenB binds to cellulose (24). CenB produced by *E. coli* is more than twice the size of Cex and CenA: the M_rs of Cex, CenA, and CenB are 47, 49, and 110 kDa, respectively. The relationship, if any, of CenB to Cex and CenA was not established. This report presents the nucleotide sequence of the *cenB* gene; an analysis of the deduced amino acid sequence of CenB and its relationship to those of Cex, CenA, and other cellulases; and further analysis of the catalytic activities of CenB.

MATERIALS AND METHODS

Bacterial strains, plasmids, bacteriophage, and media. *E. coli* JM83 and JM101 were used as hosts for plasmids pJB301 and pJB303, which were described previously (24), in the cloning and sequencing experiments. Strain RR1 was used for enzyme production. *E. coli* BD1527 *supE supF* was used as the host for propagation of bacteriophage λ::Tn*5*seq1 *b*221 *c*I857 *O*am29 *P*am80 (22). Cultures were grown at 30°C in LB medium (20) supplemented with 50 μg of ampicillin and/or kanamycin ml^{-1} when appropriate. Solid medium contained 1.5% agar (Difco). The bacteriophage stocks were prepared by the plate lysis method by using overlays containing 0.7% agar.

Transposition of Tn*5*seq1. *E. coli* JM83(pJB301) was infected with λ::Tn*5*seq1 at a multiplicity of infection of 5 to 10. Plasmid DNA was extracted from the resulting population of kanamycin- and ampicillin-resistant cells and used to transform JM83 to kanamycin and ampicillin resistance. The resulting clones were screened for carboxymethyl (CM)-cellulase activity on CM-cellulose agar plates (28), and the plasmids they contained were analyzed by digestion with appropriate restriction endonucleases to determine the sites of insertion of Tn*5*seq1 in pJB301. An appropriate set of insertions was chosen for sequencing.

DNA sequencing. Plasmid DNA was sequenced on both strands by the double-strand modification of the chain termination sequencing method by using the primers appropriate to Tn*5*seq1 (22). The primers for filling in any gaps were synthesized on an Applied Biosystems 380A automated DNA synthesizer. The M13 universal primer was used to

* Corresponding author.

sequence the deletion in pJB303. The enzyme used was T7 DNA polymerase, with 7-deaza dGTP instead of dGTP in nucleotide mixes adjusted for the high G+C content of *C. fimi* DNA. Plasmid DNA was isolated by the alkaline lysis procedure (4) and purified by dye-buoyant density centrifugation (20).

Sequence comparisons. Amino acid sequences were aligned with the PCGENE program from Intelligenetics. The words "identity" and "identical" are used in preference to "homology" and "homologous" to indicate the relatedness of the sequences (6).

Enzyme assays. Cell extracts were prepared by rupturing the cells in a French pressure cell and then removing the debris by centrifugation. CenB was purified by affinity chromatography on cellulose as described previously (24). CenB activity was quantified with CM-cellulose as the substrate, either colorimetrically or viscometrically, as described previously (10). Cellooligosaccharides were prepared by partial acid hydrolysis of cellulose, followed by chromatographic fractionation on a charcoal-celite-stearic acid column (21). Reaction mixtures contained 6 to 7 U of enzyme and substrate at final concentrations of 0.5 to 10.0 mM in a total volume of 200 μl of 50 mM potassium phosphate buffer, ph 7.0. Hydrolysis of oligosaccharides was determined qualitatively by thin-layer chromatography of the reaction products on silica gel (5). The anomeric configuration of the immediate product of hydrolysis of cellohexaose was determined by ^1H nuclear magnetic resonance as described previously (34).

Binding to Avicel. The enzyme was incubated on ice for 1 h with 10 mg of Avicel in 100 μl of 50 mM phosphate buffer (pH 7.0) in a 1.5-ml Eppendorf tube. The tube was shaken periodically. The Avicel was washed by centrifugation and suspension, once with 0.5 M NaCl and three times with 10 mM Tris hydrochloride (pH 8.0). Finally, the bound enzyme was eluted with water and quantified with CM-cellulose as the substrate.

Enzymes, chemicals, etc. All restriction endonucleases and T7 DNA polymerase were purchased from Bethesda Research Laboratories, Burlington, Ontario, Canada, or Pharmacia, Baie d' Urfe, Quebec, Canada. Radioactive deoxyribonucleoside triphosphates were from New England Nuclear Research Products, Boston, Mass. Low-viscosity CM-cellulose was from Sigma Chemical Co., St. Louis, Mo. The degree of polymerization was 400; the degree of substitution was 0.7.

RESULTS

Nucleotide sequence of *cenB*. The first 165 nucleotides of the coding sequence and the 274 nucleotides immediately upstream of the translational start codon of the *cenB* gene were sequenced previously (24). The M_r of mature CenB, 110 kDa, indicated that the entire coding sequence of the gene was approximately 3 kb. In the present work, Tn*5*seq1 insertions were used to sequence the remainder of the gene and the downstream region.

The sequence obtained contained an open reading frame of 1,045 codons, starting with the ATG identified previously as the translational start codon (Fig. 1). The calculated molecular weight of the encoded polypeptide, less the leader peptide of 33 amino acids, was 105,905, consistent with the 110-kDa M_r of mature CenB. The promoters and ribosome-binding site of *cenB* were identified previously (24). There was a 16-bp inverted repeat 33 nucleotides downstream from the translational stop codon (Fig. 1). This could be a transcriptional stop signal.

Codon usage in *cenB*. Codon usage in *cenB* was very similar to that in *cex* and *cenA* and reflected the 71 mol% G+C in *C. fimi* DNA (Table 1). Twenty-six codons were not used in *cenB*, and 21 of these were also not used in *cex* and *cenA*. The only noticeable differences between the genes were greater usage of the arginine codon CGG and the glycine codon GGG in *cenB*.

The hydroxyamino acid content was 22%. Many enzymes, including cellulases, which hydrolyze polysaccharides have hydroxyamino acid contents in the range of 20 to 30%.

Predicted amino acid sequence of CenB. The predicted amino acid sequence of CenB had several striking features. The sequence of the 100 amino acids at the C terminus of the polypeptide was more than 50% similar to the CBDs of Cex and CenA (Fig. 2). Although its role, if any, in the binding of CenB to cellulose was not established, the sequence was designated CBD_{CenB}. Similar sequences occur at the N or C termini of other bacterial cellulases and a xylanase (1, 3, 13, 14, 36). The consensus for these sequences is given in Fig. 2.

CBD_{CenB} was preceded by three contiguous repeats of a sequence of 98 amino acids which were >60% identical (Fig. 3). The 17 or 18 amino acids at the N terminus of each repeat were very rich in proline and threonine residues and were considered to be analogous to the Pro-Thr boxes of Cex and CenA (Fig. 1). There was also a sequence of 13 amino acids rich in proline and hydroxyamino acids between the third repeat and CBD_{CenB} (Fig. 1). These Pro-Thr boxlike sequences will be referred to as linkers.

A tetrapeptide was repeated with variations at five uniformly spaced locations within the polypeptide: FTTD at amino acids 505 to 508, TTTD at 611 to 614, FTTD at 705 to 708, FTTA at 802 to 805, and FTTL at 902 to 905 (Fig. 1). Four of these repeats occurred at the N termini of the linkers, but that at amino acids 505 to 508 was not part of an obvious linker.

The sequence of amino acids 1 to 461 of CenB was 35% identical to that of a cellulase from avocado, *Persea americana*, which comprises 469 amino acids (31). A further 15% of the amino acids were conservative changes, giving an overall similarity of 50% (Fig. 4). The similarity was greatest for amino acids 1 to 245 of CenB.

A deletion mutant of *cenB*. The nucleotide sequence of a *cenB* deletion mutant described previously (24) showed that it encoded a polypeptide lacking the 333 amino acids at the C terminus of wild-type CenB (Fig. 2). The *cenB* sequence was joined out of frame to the LacZ-encoding sequence so that amino acid 679 of CenB was followed by 14 heterologous amino acids and a stop codon. The deletion removed CBD_{CenB} and two of the adjacent repeats. The calculated molecular weight of the mutant polypeptide was 74,447, in good agreement with the M_r of 72 kDa observed previously (24). The truncated polypeptide could still bind to cellulose. The specific activity of a cell extract of RR1(pJB301), which encodes wild-type CenB, was 136 U mg^{-1}. About 70% of the total activity and 74% of the activity in the periplasm could be bound to Avicel. The corresponding figures for RR1 (pJB303), which encodes the truncated polypeptide, were 118, 23, and 50%. This suggests that the truncated polypeptide has a lower affinity for Avicel than does the native polypeptide.

Catalytic mechanism of CenB. Previous work showed that CenB behaves like an endoglucanase with CM-cellulose as the substrate (10). CenB gave several products from cellotetraose and predominantly cellobiose plus cellotetraose from cellohexaose. The enzyme was inhibited by high concentra-

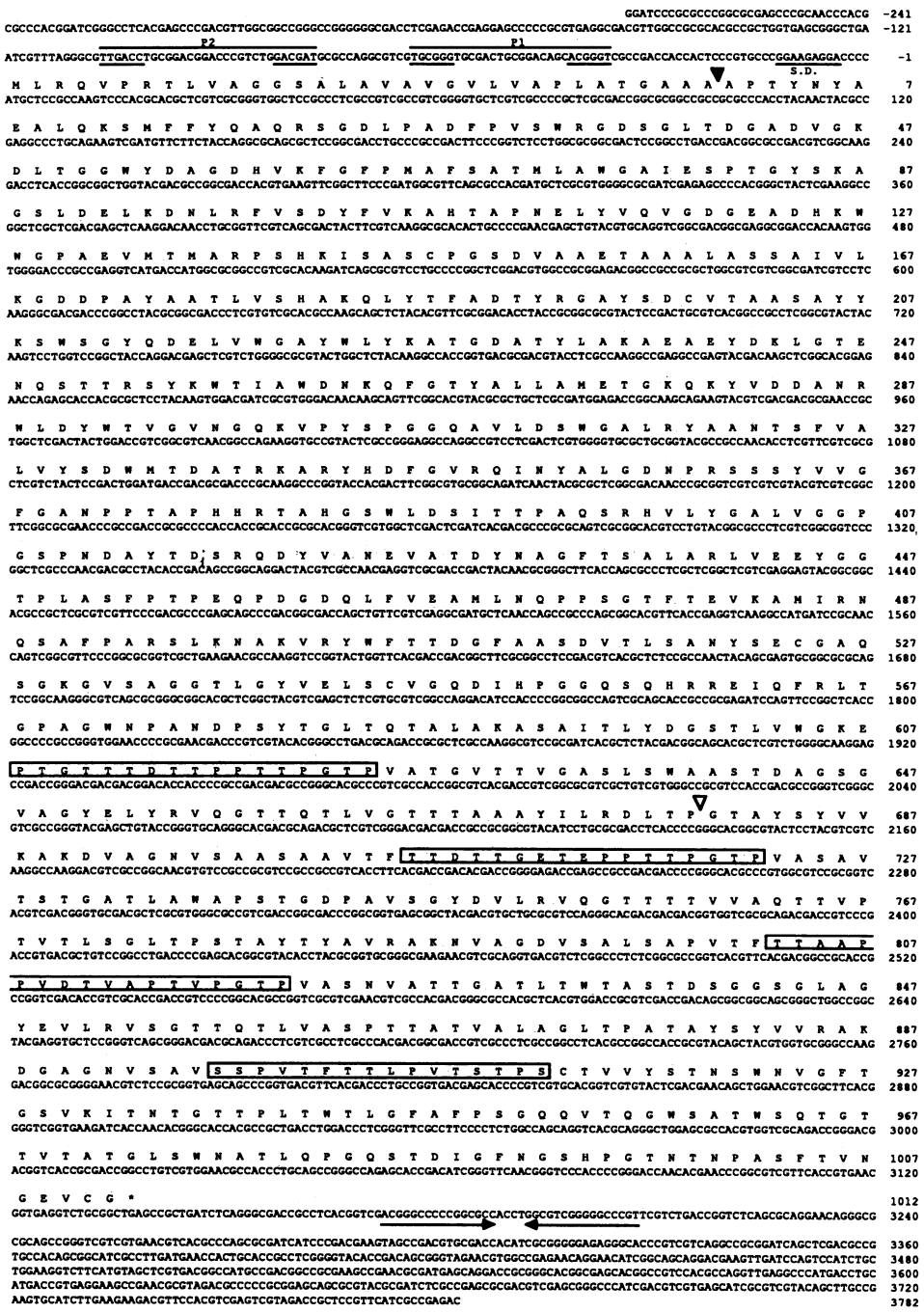

FIG. 1. Nucleotide sequence of *cenB* and its flanking regions and deduced amino acid sequence of CenB. The nucleotide sequence is numbered with +1 as the A of the translation initiation codon of *cenB*. The amino acid sequence is numbered with +1 as the N-terminal amino acid of mature CenB. S.D., Putative ribosome-binding site. P_1 and P_2, previously identified promoters (12). Symbols: ▼, leader peptide-processing site; ▽, endpoint of the C-terminal deletion of *cenB* in pJB303. The inverted repeats downstream of the TGA translational stop codon are underlined. Linkers in the amino acid sequences are boxed.

tions of cellotetraose (data not shown). Direct ^1H nuclear magnetic resonance observation of the chemical shift and coupling constant of the anomeric proton at the new hemiacetal center (34) showed that CenB hydrolyzed cellohexaose with inversion of the configuration. The theoretical α-OH:β-OH ratio at C_1 before hydrolysis of cellohexaose was 1:2, and the measured ratio was 1:2.4. The theoretical α-OH:β-OH ratios at C_1 after hydrolysis of cellohexaose were 2:1 with inversion and 1:4.5 with retention, and the measured ratio was 1.5:1.

pH optimum of CenB. The optimum pH for hydrolysis of CM-cellulose by CenB was 8.5 to 9.0. This contrasted with the optimum of 7.0 to 7.5 observed for CenA (data not shown).

TABLE 1. Codon usage in *cenA*, *cex*, and *cenB*

Amino acid encoded	Codon	No. of occurrences in:		
		cenA	*cex*	*cenB*
Ala	GCT	1	2	1
	GCC	21	28	63
	GCA	2	2	3
	GCG	33	34	73
Arg	CGT	0	1	0
	CGC	14	12	14
	CGA	0	0	0
	CGG	2	6	16
	AGA	1	0	0
	AGG	1	1	0
Asn	AAT	0	0	0
	AAC	22	23	32
Asp	GAT	0	0	0
	GAC	19	28	54
Cys	TGT	0	0	0
	TGC	6	6	6
Gln	CAA	0	0	1
	CAG	21	22	35
Glu	GAA	1	0	0
	GAG	10	14	27
Gly	GGT	4	1	7
	GGC	48	38	79
	GGA	0	0	2
	GGG	3	3	17
His	CAT	0	0	0
	CAC	3	7	13
Ile	ATT	0	0	0
	ATC	12	9	13
	ATA	0	0	0
Leu	TTA	0	0	0
	TTG	0	0	0
	CTT	0	0	0
	CTC	17	16	44
	CTA	0	0	0
	CTG	12	9	23
Lys	AAA	0	0	0
	AAG	13	19	31
Met	ATG	3	6	10
Phe	TTT	0	0	0
	TTC	9	22	29
Pro	CCT	0	1	0
	CCC	10	10	21
	CCA	0	0	1
	CCG	22	22	40
Ser	TCT	0	0	1
	TCC	10	8	22
	TCA	0	0	0
	TCG	15	13	41
	AGT	0	0	0
	AGC	5	9	23
Thr	ACT	0	0	1
	ACC	29	22	61
	ACA	0	0	0
	ACG	21	27	76
Trp	TGG	16	12	26
Tyr	TAT	0	0	0
	TAC	13	10	50
Val	GTT	0	0	0
	GTC	24	27	65
	GTA	0	0	0
	GTG	6	14	24
Stop	TAA	0	0	0
	TGA	1	1	1
	TAG	0	0	0

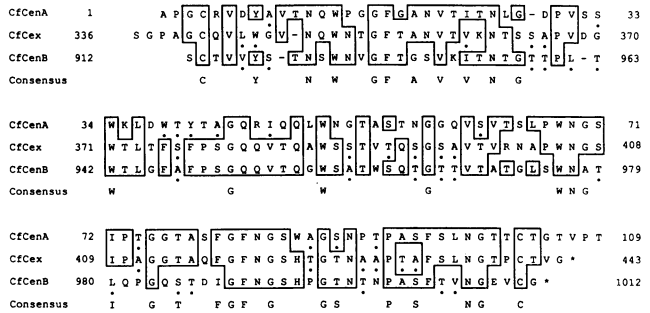

FIG. 2. Amino acid sequences of the CBDs. CfCenA, Endoglucanase CenA of *C. fimi*; CfCex, exoglucanase Cex of *C. fimi*; CfCenB, putative CBD of endoglucanase CenB of *C. fimi*; Consensus, consensus sequence for the CBDs of nine bacterial cellulases, including CenA, CenB, and Cex. The boxes enclose identical residues. The dots indicate conservative changes. The numbers refer to the positions of the amino acids in the sequences of the mature polypeptides. The dashes indicate gaps left to improve the alignment. The asterisks indicate stop codons.

DISCUSSION

CenB is unusual in several respects. It has three repeats of a sequence of 98 amino acids, it has a bacterial-type CBD, and it is among the longest cellulases sequenced to date. Repeated sequences and CBDs occur in other cellulases (1a, 2). However, except for cellobiohydrolase CbhII of the fungus *Trichoderma reesei*, which has an N-terminal fungal-type CBD followed by a duplication of 20 amino acids (29), the cellulases described to date have either repeated sequences or a CBD. The longest repeated sequence reported previously is of 61 amino acids, in endoglucanase CelB from alkalophilic *Bacillus* sp. strain N4 (7). The mature forms of CenB and CelB from *Caldocellum saccharolyticum* (25) and the Egl proenzyme from *Pseudomonas fluorescens* subsp. *cellulosa* (13) comprise 1,012, 1,011, and 962 amino acids, respectively. There are other cellulases of similar or greater size, but until their sequences are published, their structural organizations cannot be compared with those of CenB, CelB, and Egl.

A partially cloned and sequenced open reading frame, *clfX*, from *Cellulomonas flavigena* appears to encode a polypeptide with some close similarities to CenB. The C terminus of this polypeptide, ClfX, comprises a linker sequence of 37 amino acids containing only proline, aspartate, glutamate, and threonine; this is followed by a sequence of 97 amino acids which is 50% identical to the copy of the repeated sequence immediately preceding CBD_{CenB} in CenB

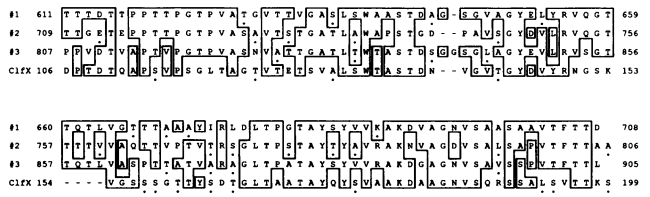

FIG. 3. Repeated amino acid sequences in CenB. The boxes enclose identical residues. The dots indicate conservative changes, and the dashes indicate gaps left to improve the alignment. The numbers refer to the positions of the amino acids in the sequence of mature CenB. ClfX is part of the amino acid sequence predicted from an open reading frame on a fragment of *Cellulomonas flavigena* DNA that encodes a cellulase.

```
CfCenB    1  APTYNYAEALQKSMFFYQAQRSGDLPADFPVSWRGDSGLTDGADVGKDLTGGWYDAGDHVKFGFPMAFSATMLAWGAIESPTGYSKAGSLDELKDNLRFV  100
PaCel     1  -SDLH sd  E  iL fEG    K  tNQRlt       s  sSYHV  V  y     Nl   L    tt       I  --F CLMPEQveNArAA  wS   97

CfCenB  101  SDYFVKAHTA-PNELYVQVGDGEADHKWWGPAEVMTMARPSHKISASCPGSDVAAETAAALASSAIVLKGDDPAYAATLVSHAKQLYTFADTYRGAYSDC  199
PaCel    98  t  Ll  S  TS S      ePN   rC ERP D DTP NVY v tQN              aas  FGDS Ss stK lHT VKvfE   Q   s   S  197

CfCenB  200  V-TAASAYYKSWSGYQDELVWGAYWLYKATGDATYLAKAEAEYDKLGTENQSTTRSYKWTIAWDNKQFGTYA--LLAMETGKQKYVDDANRWLDYWTVGV  296
PaCel   198  lGsVVCPf C y   n   l   S  Hr sQN s mtYIQsNGHT  adDDDY-- fS DDKRVGTKVLLSKGF QDRIEEL L KVHtDNyiCSLIP T  295

CfCenB  297  NGQKVPYSPGGQAVLDSWGALRYAANTSFVALVYSDWM----------TDATRKARYHDFGVRQINYALGDNPRSSSYVVGFGANPPTAPHHRTAHGSWL  386
PaCel   296  SSFQAQ t   LLYKG ASN Q VtS a lL T aNylNSSGGHASCG TtVTAKNLISLAKK vD I  Q  AKM  m    ERY QHV   GsSLPSv  395

CfCenB  387  DSITTPAQSRHVLYGALVGGPGSPNDAYTDSRQDYVANEVATDYNAGFTSALARLVEEYGGTPLASFPTPEQPDGD                        462
PaCel   396  Q-vHPNsIPCNAGfQY YSS PN  ILVGAILGGPDNRdSFs DRNNyQQsEPATYINAPLVGALa FaANPVTE*                         469
```

FIG. 4. Similarity between the amino acid sequences of CenB (CfCenB) and an avocado endoglucanase (PaCel). Identical amino acids in the second sequence are left blank. Conservative changes are indicated by lowercase letters. The numbers refer to the positions of the amino acids in the sequences of the mature polypeptides. The dashes indicate gaps left to improve the alignment.

(Fig. 3); finally, there is a sequence of 104 amino acids which is >60% identical to CBD_{CenB} (1).

Short sequences of amino acids rich in proline and/or hydroxyamino acids appear to link discrete functional domains in many cellulases. In CelB, for example, two linkers divide the enzyme into three domains. The N-terminal domain has exoglucanase activity, and the C-terminal domain has endoglucanase activity (25). CenB and Egl are also divided into discrete regions by linkers, the C-terminal region being a CBD in each case (Fig. 5). This suggests that the repeated sequences in CenB form three discrete domains, the function(s) of which have yet to be determined.

The CBDs of CenA (11), Cex (11), CenB (24), and Egl (13) can be removed without destroying the enzymatic activities of the proteins (Fig. 5). The active fragments from CenA and Cex do not bind to cellulose (11, 18); the active fragment from CenB does, but not as strongly as the intact enzyme. Although the binding of CBD_{CenB} to cellulose has yet to be tested, CenB could have a second site for binding to cellulose. This may be its catalytic domain, perhaps the active site itself, or a second discrete binding sequence. The catalytic domain has no obvious linker sequences within it to delineate a second putative binding domain (Fig. 1 and 6). The sequence of amino acids 463 to 603, which includes the tetrapeptide FTTD at residues 505 to 508 and is not homologous to the avocado cellulase sequence, could contain another cellulose-binding site.

Sequence similarities between and the clustering of hydrophobic amino acids in the catalytic domains of cellulases place endoglucanase CelD from *Clostridium thermocellum* (16), endoglucanase Egl from *P. fluorescens* subsp. *cellulosa* (13), and the endoglucanase from avocado (31) in family E of cellulases and related enzymes (2, 15). Sequence similarities alone, however, place CelD and Egl in subfamily E_1 and the avocado enzyme in subfamily E_2 (2). CenB is similar to the avocado enzyme but not to CelD and Egl, placing it in subfamily E_2 (Table 2). Interestingly, there is a short sequence of amino acids which is conserved in all four enzymes (Fig. 6).

The three *C. fimi* cellulases characterized to date belong to three different catalytic domain families; CenA belongs to family B, CenB belongs to family E, and Cex belongs to family F (2, 15). However, they all have the same type of CBD. Such diversity and conservation within the cellulases of a microorganism may be quite common. *C. thermocellum* produces endoglucanases belonging to families A, D, and E which have the same repeated sequence of 24 amino acids (2), and *T. reesei* produces cellulases belonging to families A, B, and C which have similar CBDs and linkers (2, 15, 17). It is clear that many cellulases arose by shuffling of a limited number of progenitor sequences. Families A, B, and F contain fungal and bacterial enzymes, and family E contains bacterial enzymes and a plant enzyme(s) (2, 15).

Although CenB is more than twice as long as Cex and CenA, the difference between the catalytic domains is less

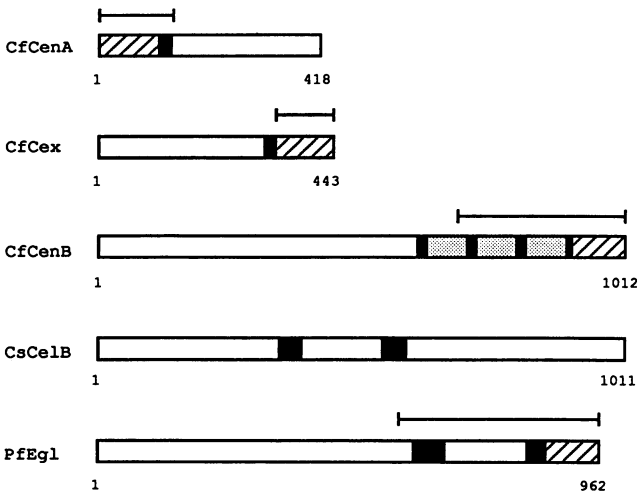

FIG. 5. Overall structures of cellulases. The lengths of the polypeptides are drawn to scale. ▬, Linker sequence; ▨, CBD sequence; ▩, repeated sequence; ⊢⊣, part of the polypeptide removed without loss of catalytic activity. The numbers indicate the first and last amino acids in the mature polypeptides, except for PfEgl, for which the numbers refer to the proenzyme. CfCenA, CfCenB, and CfCex are endoglucanases CenA and CenB and exoglucanase Cex from *C. fimi*, respectively. CsCelB is the bifunctional endoglucanase-exoglucanase from *C. saccharolyticum*. PfEgl is an endoglucanase from *P. fluorescens* subsp. *cellulosa*.

```
CfCenB   41  D G A D V G K D L T G G W Y D A G D H V K F G F P M A F S  69
PaCel    40  D G S S Y H V D L V G G Y Y D A G D N L K F G L P M A F T  68
CtCelD  143  N G Q H T K K D S T K G W H D A G D Y N K Y V V N A G I T  171
PfEgl   186  G T C N Y S L N V T K G W Y D A G D H G K Y V V N G G I S  214
```

FIG. 6. Conserved sequences in the enzymes of cellulase family E. CfCenB, endoglucanase CenB from *C. fimi*; PaCel, an endoglucanase from *P. americana*; CtCelD, endoglucanase D from *C. thermocellum*; PfEgl, an endoglucanase from *P. fluorescens* subsp. *cellulosa*. The numbers refer to the positions of the amino acids in the sequences of the mature polypeptide, except for PfEgl, which is numbered from the N-terminal methionine because the processing site has not been identified. The boxes enclose identical residues. The dots indicate conservative changes.

TABLE 2. Sequence similarities between the enzymes in cellulase family E

Enzyme[a]	% Identity (% conservative changes) with:		
	CtCelD	PfEgl	PaEgl
PfEgl	31 (16)		
PaEgl	20 (11)	20 (13.5)	
CfCenB	22 (12.5)	22 (12.5)	34 (16.5)

[a] CfCenB, Amino acids 1 to 469 of endoglucanase CenB from *C. fimi*; CtCelD, amino acids 102 to 542 of endoglucanase CelD from *C. thermocellum*; PaEgl, amino acids 1 to 469 of an endoglucanase from *P. americana*; PfEgl, amino acids 146 to 607 of an endoglucanase from *P. fluorescens* (numbering starts at Met).

pronounced. The catalytic domains of Cex and CenA are about 300 amino acids long, and that of CenB is probably about 450 amino acids or less, on the basis of the deletion in pJB303, its homology with the avocado enzyme, and the possible presence of a second binding sequence.

CenA (34) and CenB hydrolyze the β-1,4-glucosidic bond with inversion of the anomeric configuration, and Cex hydrolyzes it with retention of the configuration (34). Enzymes that cause inversion must hydrolyze the bond by a mechanism different from that of those that retain the configuration (26). If enzymes within a family are closely related, they should all hydrolyze with inversion or retention of the anomeric configuration and have similar pH optima. These criteria could be as important as sequence similarities in assigning cellulases to families.

ACKNOWLEDGMENTS

We thank D. Berg for providing λ::Tn*5*seq1.

A.M. thanks the Land Baden-Württemberg for a scholarship through the Landesgraduiertenförderungsgesetz. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. **Al-Tawheed, A. R.** 1988. M.S. thesis. Trinity College, Dublin.
1a. **Béguin, P.** Annu. Rev. Microbiol., in press.
2. **Béguin, P., J. Millet, S. Chauvaux, E. Yagüe, P. Tomme, and J.-P. Aubert.** 1989. Genetics of bacterial cellulases, p. 57–72. *In* M. P. Coughlan (ed.), Enzyme systems for lignocellulose degradation. Elsevier Applied Science, London.
3. **Berger, E., W. A. Jones, D. T. Jones, and D. R. Woods.** 1989. Cloning and sequencing of an endoglucanase (*end*1) gene from *Butyrivibrio fibrisolvens* H17c. Mol. Gen. Genet. **219**:193–198.
4. **Birnboim, H. G., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
5. **Chiricho, W. J., and R. D. Brown, Jr.** 1985. Separation of [1-^{3}H]cellooligosaccharides by thin-layer chromatography: assay for cellulolytic enzymes. Anal. Biochem. **150**:264–272.
6. **Doolittle, R. F.** 1986. Of URFs and ORFs: a primer on how to analyze derived amino acid sequences, p. vii and 103. University Science Books, Mill Valley, Calif.
7. **Fukumori, F., N. Sashihara, T. Kudo, and K. Horikoshi.** 1986. Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. strain N-4 and their strong homology. J. Bacteriol. **168**:479–485.
8. **Gilkes, N. R., D. G. Kilburn, M. L. Langsford, R. C. Miller, Jr., W. W. Wakarchuk, R. A. J. Warren, D. J. Whittle, and W. R. K. Wong.** 1984. Isolation and characterization of *Escherichia coli* clones expressing cellulase genes from *Cellulomonas fimi*. J. Gen. Microbiol. **130**:1377–1384.
9. **Gilkes, N. R., D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1989. Structural and functional analysis of a bacterial cellulase by proteolysis. J. Biol. Chem. **264**:17802–17808.
10. **Gilkes, N. R., M. L. Langsford, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1984. Mode of action and substrate specificities of cellulases from cloned bacterial genes. J. Biol. Chem. **259**:10455–10459.
11. **Gilkes, N. R., R. A. J. Warren, R. C. Miller, Jr., and D. G. Kilburn.** 1988. Precise excision of the cellulose-binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. **263**:10401–10407.
12. **Greenberg, N. M., R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1987. Regulation and initiation of *cenB* transcripts of *Cellulomonas fimi*. J. Bacteriol. **169**:4674–4677.
13. **Hall, J., and H. J. Gilbert.** 1988. The nucleotide sequence of a carboxymethylcellulase gene from *Pseudomonas fluorescens* subsp. *cellulosa*. Mol. Gen. Genet. **213**:112–117.
14. **Hall, J., G. P. Hazelwood, N. S. Huskisson, A. J. Durrant, and H. J. Gilbert.** 1989. Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subspecies *cellulosa*: internal signal sequence and unusual protein processing. Mol. Microbiol. **3**:1211–1219.
15. **Henrissat, B., M. Claeyssens, P. Tomme, L. Lemesle, and J.-P. Mornon.** 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene **81**:83–95.
16. **Joliff, G., P. Béguin, and J.-P. Aubert.** 1986. Nucleotide sequence of the cellulase gene *cel*D encoding endoglucanase D of *Clostridium thermocellum*. Nucleic Acids Res. **14**:8605–8613.
17. **Knowles, J., P. Lehtovaara, and T. Teeri.** 1987. Cellulase families and their genes. Trends Biotech. **5**:255–261.
18. **Langsford, M. L., N. R. Gilkes, B. Singh, B. Moser, R. C. Miller, Jr., R. A. J. Warren, and D. G. Kilburn.** 1987. Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. FEBS Lett. **225**:163–167.
19. **Langsford, M. L., N. R. Gilkes, W. W. Wakarchuk, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1984. The cellulase system of *Cellulomonas fimi*. J. Gen. Microbiol. **130**: 1367–1376.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Miller, G. L., J. Dean, and R. Blum.** 1960. A study of methods for preparing oligosaccharides from cellulose. Arch. Biochem. Biophys. **91**:21–26.
22. **Nag, D. K., H. V. Huang, and D. E. Berg.** 1988. Bidirectional chain-termination nucleotide sequencing: transposon Tn*5*seq1 as a mobile source of primer sites. Gene **64**:135–145.
23. **O'Neill, G. P., S. H. Goh, R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1986. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. Gene **44**:325–330.
24. **Owolabi, J. B., P. Béguin, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1988. Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. Appl. Environ. Microbiol. **54**:518–523.
25. **Saul, D. J., L. C. Williams, D. R. Love, L. W. Chamley, and P. L. Bergquist.** 1989. Nucleotide sequence of a gene from *Caldocellum saccharolyticum* encoding for exocellulase and endocellulase activity. Nucleic Acids Res. **17**:439.
26. **Sinnott, M. L.** 1987. Glycosyl group transfer, p. 259–297. *In* M. I. Page and A. Williams (ed.), Enzyme mechanisms. Royal Society of Chemistry, London.
27. **Stahlberg, J., G. Johansson, and G. Pettersson.** 1988. A binding-site-deficient, catalytically active, core protein of endoglucanase III from the culture filtrate of *Trichoderma reesei*. Eur. J. Biochem. **173**:179–183.
28. **Teather, R. M., and P. J. Wood.** 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from bovine rumen. Appl. Environ. Microbiol. **43**:777–780.
29. **Teeri, T. T., P. Lehtovaara, S. Kauppinen, I. Salovuori, and J. Knowles.** 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. Gene **51**:43–52.
30. **Tomme, P., H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeyssens.** 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM

9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. **170**:575–581.

31. **Tucker, M. L., M. L. Durbin, M. T. Clegg, and L. N. Lewis.** 1987. Avocado cellulase: nucleotide sequence of a putative full-length cDNA clone and evidence for a small gene family. Plant Mol. Biol. **9**:197–203.
32. **Van Tilbeurgh, H., P. Tomme, M. Claeyssens, R. Bhikhabhai, and G. Pettersson.** 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. Separation of functional domains. FEBS Lett. **204**:223–227.
33. **Warren, R. A. J., C. F. Beck, N. R. Gilkes, D. G. Kilburn, M. L. Langsford, R. C. Miller, Jr., G. P. O'Neill, M. Scheufens, and W. K. R. Wong.** 1986. Sequence conservation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi*. Protein Struct. Funct. Genet. **1**:335–341.
34. **Withers, S. G., D. Dombroski, L. Berven, D. G. Kilburn, R. C. Miller, Jr., R. A. J. Warren, and N. R. Gilkes.** 1986. Direct ^{1}H N.M.R. determination of the stereochemical course of hydrolyses catalyzed by glucanase components of the cellulase complex. Biochem. Biophys. Res. Commun. **139**:487–494.
35. **Wong, W. K. R., B. Gerhard, Z. M. Guo, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Characterization and structure of an endoglucanase gene *cen*A of *Cellulomonas fimi*. Gene **44**:315–324.
36. **Yablonsky, M. D., T. Bartley, K. O. Elliston, S. K. Kahrs, Z. P. Shalita, and D. E. Eveleigh.** 1988. Characterization and cloning of the cellulase complex of *Microbispora bispora*. FEMS Symp. **43**:249–266.