

Two new cellulosome components encoded downstream of *cell* in the genome of *Clostridium thermocellum*: the non-processive endoglucanase CelN and the possibly structural protein CseP

Vladimir V. Zverlov,¹ Galina A. Velikodvorskaya¹ and Wolfgang H. Schwarz²

Correspondence

Wolfgang H. Schwarz

schwarz@mikro.biologie.tu-muenchen.de

¹Institute of Molecular Genetics, Russian Academy of Science, Kurchatov Sq., 123182 Moscow, Russia

²Research Group Microbial Biotechnology, Technische Universität München, Am Hochanger 4, D-85350 Freising-Weihenstephan, Germany

Clostridium thermocellum produces a great number of extracellular cellulases which are free or cellulosome-bound. The nucleotide sequence of a gene cluster containing the genes *cell*, *celN* and *cseP* was determined from *C. thermocellum* strain F7. Gene products Cel9I and Cel9N are structurally related enzymes having a glycosyl hydrolase family 9 and a carbohydrate-binding module (CBM3c), but show characteristic differences: Cel9I is a non-cellulosomal protein with an additional CBM (CBM3b), whereas Cel9N contains a cellulosomal dockerin module and no additional CBM. Although Cel9I is a processive endoglucanase, Cel9N is non-processive. Both enzymes hydrolyse phosphoric acid swollen cellulose, but the products of hydrolysis are different. The CseP protein encoded in the gene cluster is the first component attached to the cellulosomal scaffoldin for which no catalytic activity could be detected. It was shown to be present in the cellulosome. Its sequence is homologous to the spore-coat assembly protein CotH of *Bacillus subtilis*, suggesting a structural role of CseP in the cellulosome.

Received 23 August 2002

Revised 18 October 2002

Accepted 28 October 2002

INTRODUCTION

The enzymic saccharification of native, crystalline cellulose is potentially of great economic importance, but it is not yet cost-effective. This process is difficult because it involves a set of co-operative enzymes which have to hydrolyse a chemically homogeneous substrate with a heterogeneous structure: cellulose is partially crystalline and thus there are a great number of different topologies in the arrangement of the single substrate molecules (Baker *et al.*, 2000). Only by the synchronous co-operation of enzymes with different modes of activity in a synergistic arrangement can this recalcitrant substrate be effectively and completely decomposed (Boisset *et al.*, 1998).

To cope with these problems cellulolytic anaerobic bacteria have developed a defined arrangement of enzymes along a non-catalytic scaffolding protein, the cellulosome (Schwarz, 2001). This large multienzyme complex contains various cellulase components which differ in their modes of action (Lamed *et al.*, 1983; Bayer *et al.*, 1998). The cellulosome

components are aligned on the scaffoldin (e.g. the CipA protein of *Clostridium thermocellum*) of the cellulosome with the aid of their dockerin modules which specifically recognize the CipA-cohesin modules (Béguin *et al.*, 1998). Due to the specificity of the cohesin-dockerin interaction, the presence of a dockerin module sequence in a protein is indicative of its localization in the cellulosome (Mechaly *et al.*, 2000; Salamitou *et al.*, 1994).

The bacterial cellulase systems investigated so far include a variety of non-processive endo- β -1,4-glucanases, producing new ends at random within a polysaccharide chain, and processive cellulases (exo- β -1,4-glucanases) which remain attached to one end of the substrate and split off cellobiose (cellobiohydrolases) or multimers of cellobiose (processive endocellulases) (Reverbel-Leroy *et al.*, 1997; Teeri, 1997). β -1,4-Glucanases with a different mode of action work synergistically to effectively degrade the crystalline substrate (Barr *et al.*, 1996). The catalytic domains of cellulases have been assigned on the basis of sequence comparisons and hydrophobic cluster analysis to glycosyl hydrolase families (GHFs) (Henrissat *et al.*, 1998; CAZY server, <http://afmb.cnrs-mrs.fr/~pedro/CAZY>). In many cases the catalytic modules of processive and non-processive enzymes are closely related and only differ in the depth of their active site pocket or in additional 'helper

Abbreviations: CBM, carbohydrate-binding module; GHF, glycosyl hydrolase family; PASC, phosphoric acid swollen cellulose; p NP-, p -nitrophenyl-.

The GenBank accession number for the sequence reported in this paper is AJ275974.

modules' (Bayer *et al.*, 2000). The direction of processivity, i.e. activity on the reducing or the non-reducing end of the substrate, seems to be defined by the method of substrate binding and release (Parsieglä *et al.*, 2000).

Besides the presence of diverse catalytic modules, activity on crystalline cellulose also depends on the presence of non-catalytic modules which contribute tight substrate binding (Tomme *et al.*, 1998). These carbohydrate-binding modules (CBMs) have a stimulating effect by increasing enzyme–substrate proximity, helping the enzyme to overcome the liquid–solid interface and enhancing the accessibility of the substrate surface by interacting with the crystal (Bolam *et al.*, 1998; Nutt *et al.*, 1998; Pagès *et al.*, 1997). They are indispensable in catalytic systems where the size relationship between substrate and enzyme(s) seriously limits the molecular dynamics of enzyme action.

Substrate-binding modules from bacteria have been divided into 26 families by sequence comparison [Carbohydrate-Active enZymes and associated MODular Organization server (CAZyModO website), <http://afmb.cnrs-mrs.fr/~pedro/DB/db.html>; Tomme *et al.*, 1998]. Different binding specificities promote the cellulolysis of different sites on the crystalline substrate (Carrard *et al.*, 2000). They have a size of 40–180 aa. Family 3 plays an eminent role in bacterial cellulases. Three subforms are recognized: CBM3a, CBM3b and CBM3c, with CBM3a and CBM3b binding to crystalline or amorphous cellulose; CBM3c is connected exclusively to the catalytic module of GHF9 cellulases and as well as stabilizing it, may modulate the activity mode of an endoglucanase into a processive endoglucanase by binding a single cellulose molecule (Tormo *et al.*, 1996; Irwin *et al.*, 1998; Bayer *et al.*, 2000).

In this paper we report the detection and characterization of two new cellulosome components which form a gene cluster together with cellulase Cel9I on the genome of *C. thermocellum* F7. We show that Cel9N is an endoglucanase and is, in contrast to its close relative Cel9I, not processive.

CseP is not hydrolytically active, but seems to be a structural component of the cellulosome.

METHODS

Bacterial strains and plasmids. Clone pCU108 from *Clostridium thermocellum* F7 containing an endoglucanase was described by Bumazkin *et al.* (1990). *Escherichia coli* strain XL-1 Blue was used for cloning and strain M15, containing plasmid pREP-4, was used for overexpression of cloned genes (Qiagen). Cultivation of recombinant cells, media and overexpression were as recommended in the manufacturer's handbook. Plasmids pQE30 and pQE31 (Qiagen) were used for cloning PCR products.

Molecular biological methods. The DNA sequence of clone pCU108 was determined from supercoiled double-stranded plasmid DNA on both strands (ThermoSequenase Cycle Sequencing Kit; Amersham) with biotinylated oligonucleotide primers. DNA fragments were detected with a GATC 1500 Direct-Blotting Electrophoresis apparatus (GATC) using streptavidin-conjugated alkaline phosphatase and the chromogenic substrate NBT-BCIP (Promega). PCR was carried out using oligonucleotide primer pairs with chromosomal DNA as template and the Expand High Fidelity PCR System (Boehringer). The oligonucleotide primers were designed to truncate the gene at the N terminus for the leader peptide and to introduce a hexa-histidyl tag sequence for affinity chromatography. All primers introduced a restriction enzyme recognition site at a suitable position to ligate the PCR fragment into the vector in a predetermined way. The oligonucleotide primer sequences used are listed in Table 1. Restriction digests of DNA were done with restriction endonucleases (MBI Fermentas) as recommended by the manufacturer.

Purification of recombinant proteins. Recombinant proteins were purified from 400 ml *E. coli* cultures according to the procedures described in the Qiagen handbook with 3 ml Ni-NTA superflow columns (Qiagen). The purity of the proteins was verified by SDS-PAGE and staining with Coomassie brilliant blue G-250 dye (Serva).

Preparation of cellulosomes. Cellulosomes were purified from 0.5 l cultures of *C. thermocellum* F7 in GS-2 medium (Johnson *et al.*, 1981) containing Avicel with the affinity digestion method (Morag *et al.*, 1992) as described by Zverlov *et al.* (1999).

Enzyme assays. Enzyme aliquots in standard assays were incubated in MES buffer (50 mM) containing 5 mM CaCl₂ at the optimum pH

Table 1. Oligonucleotide primers used in this study

Construct	Terminus	Primer name	Sequence (5'–3')
hCell	N	pCell-Bam	CACCGGATCCGTTGTTTGCAAAGCTATGGG
hCell	C	pCell-Sal3	CAAACAAGTCGACAGCCTGTGTTGTAA
hCellId1	C	pCell-Sal2	TTTGTCTGTCGACTGAGATTCAGTTTTTC
hCellId2	C	pCell-Sal1	ATTGTTTGTGCGACGCATTCACACCGGCTTC
hCelN	N	pCelN-Sph	TGCTGCATGCGTCTTCTTTTGAAGC
hCelN	C	pCelN-Pst3	AACACCTGCAGAAAATTCACATAAGTTAT
hCelNd1	C	pCelN-Pst2	ATCGACTGCAGGCGGTTCTTCAG
hCelNd2	C	pCelN-PstI	TAAATTCTGCAGCCACGGCAGAAAGTATTC
hCseP	N	pCelP-Sph	AAGTTTGGCAGTCATCCAGACCTGAGGGCT
hCseP	C	pCelP-Pstc	AACACCTGCAGAAAATTCACATAAGTTATATAGGG
hCsePd1	C	pCelP-Pst3	AAGGGTGCTGCAGTTCGTTTCAGGAATAGGTGTAGA

and temperature. The concentration of substrates was 1 % for soluble and 2 % (w/v) for insoluble polysaccharides. Reducing sugars released from polymeric substrates were detected by the 3,5-dinitrosalicylic acid method (Wood & Bhat, 1988), assuming that 1 U enzyme liberates 1 μmol glucose equivalent min^{-1} and $(\text{mg protein})^{-1}$. Specific activities were determined in the linear range of the reaction. Protein concentration was determined with Coomassie brilliant blue (Sedmak & Grossberg, 1977). *p*-Nitrophenol liberated from *p*NP-glycosides was measured by its absorption in alkaline solution (0.6 M Na_2CO_3) at 395 nm. 1 U of activity is defined as the amount of enzyme producing 1 μmol *p*-nitrophenol min^{-1} ($0.013 \Delta\text{OD}_{395} = 1 \text{ nmol}$). All determinations were performed in triplicate.

The optimum pH was determined by measuring the specific activity of the enzyme at a given pH (MES buffer). The optimum temperature was the temperature at which the highest activity of the enzyme occurred during incubation for a given time.

The release of reducing sugars in the soluble and the insoluble fraction of cellulose was performed with a preparation of phosphoric acid swollen cellulose (PASC) (see below) that pellets easily on centrifugation. Enzyme reactions were set up as described above and incubated. Probes were taken at different time points and centrifuged. The pellet was washed two times with MES buffer and resuspended in 1 vol. buffer. Reducing sugars were determined in the supernatant of the enzyme reaction and the washed pellet separately as described. The pellet probes were centrifuged again after boiling and the optical density of the clear supernatants was determined.

Denaturing gel electrophoresis (SDS-PAGE) and Western blot. SDS-PAGE was performed in 10 % polyacrylamide slab gels in the presence of 0.1 % SDS after boiling the proteins in the presence of 1.25 % (v/v) mercaptoethanol (Laemmli, 1970).

To elicit polyvalent antibodies against CseP, 0.25 mg purified hCsePd1 protein in Freund's adjuvant was used to immunize white rabbits. The antibodies were purified using a serum IgG purification column. Western blotting from SDS-PAGE gel slabs on nitrocellulose membranes was done by standard techniques using donkey anti-rabbit serum conjugated to horseradish peroxidase (Amersham) and 4-chloro-1-naphthol as substrate.

Sequence analysis. Sequence data were analysed with the DNASIS/PROSIS software package (Hitachi Software Engineering). Nucleotide and protein sequence databases were screened using BLAST software at the EBI server at EMBL (<http://www.ebi.ac.uk>). The CAZY server (<http://afmb.cnrs-mrs.fr/~pedro/CAZY>) was used for determining the GHF or the binding domain family [Carbohydrate-Active enZYmes and associated MODular Organization server (CAZyModO website), <http://afmb.cnrs-mrs.fr/~pedro/DB/db.html>].

TLC. Polymeric and oligomeric substrates were hydrolysed to completion under the conditions described above. Hydrolysis products were separated on 0.2 mm aluminium sheet silica gel 60 plates (Merck) with acetonitrile/water as eluent (80:20, v/v). Sugars were detected by spraying the plates with a freshly prepared mixture of 10 ml stock solution with 1 ml *o*-phosphoric acid, followed by heating the plates at 120 °C until colour developed. The stock solution consisted of 1 g diphenylamine and 1 ml aniline dissolved in 100 ml acetone.

Substrates. Birchwood xylan, Avicel CF1, carboxymethyl-cellulose (CM-cellulose, low viscosity) and *p*NP-glycosides were obtained from Sigma-Aldrich, cellodextrins from Merck, barley β -glucan from Megazym and pustulan from Roth. PASC was prepared from Avicel CF1 according to Wood (1988).

RESULTS

A clone, pCU108, expressing endoglucanase activity was isolated previously from a genomic library of thermophilic *C. thermocellum* strain F7 (Bumazkin *et al.*, 1990). The nucleotide sequence of the DNA insert was determined (7583 bp). The upstream 3685 bp were nearly identical with sequences L04735 and L04736 (GenBank accession nos) of *C. thermocellum* strain NCIB 10682, which have been reported to express endoglucanase Cell (Hazlewood *et al.*, 1993). The only difference was a deletion in *cell* (see below). The region further downstream of *cell* was cloned by selecting a fragment from genomic DNA partially digested with restriction endonuclease *Sau*3A, which hybridized against the 3' part of the pCU108 insert. A total of 10 782 bp was sequenced containing six ORFs (GenBank accession no. AJ275974; Fig. 1).

The 540 nt upstream of the *cell* gene encode the 180 aa C terminus of a potential protein (Orf1) with 40 % sequence identity to the C-terminal part of a bacterial methyl-accepting protein, a putative membrane-associated chemotaxis component from *Clostridium acetobutylicum* (GenBank accession no. AE007524). It was also 100 % identical to the fragment sequenced from *C. thermocellum* NCIB 11682 (Hazlewood *et al.*, 1993). Other reading frames in the sequenced region, with no obvious relationship to cellulases or cellulosome components, were *orf3*, encoding a putative GTP-binding protein, and *orf6*, which is homologous to the N terminus of an autolysin protein from *Paenibacillus polymyxa* (Ishikawa *et al.*, 1999; Fig. 1, Table 2). Putative mRNA loop structures were detected downstream of *orf1* (564–597, $\Delta G = -15.2 \text{ kcal}$) and *orf4* (*celN*) (7959–7997, $\Delta G = -20.4 \text{ kcal}$), both containing downstream runs of T residues in both directions, indicating possible factor-independent transcription terminators. ORFs 1–4 and 6 are encoded on the same DNA strand, *orf5* (*cseP*) is on the opposite strand.

Comparison of Cel9I and Cel9N

The gene product of *cell*, Cel9I, is a modular protein consisting of a leader peptide, a catalytic module of GHF9, CBM3c (formerly called C' domain) and a complete CBM3b module (Fig. 2). The gene product of *celN*, Cel9N, has a similar structure, but instead of the CBM3b module it

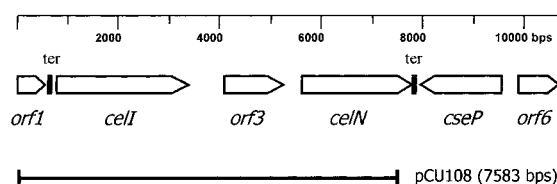


Fig. 1. Physical map of the *cell-celN-cseP* region in the *C. thermocellum* chromosome. The extension of the initial clone, pCU108, is indicated. ter, palindromic structure, a potential transcription terminator.

Table 2. ORFs in the *C. thermocellum cell* region: homology to reported genes of other bacteria

Numbers in parentheses designate the first base in the *C. thermocellum* F7 sequence (GenBank AJ275974).

ORF	Size (aa)	Start codon	Stop codon	Site	Homologous gene	Bacterium	Protein sequence identity	Gene product	Reference or accession no.
<i>orf1</i>	178 (incomplete)		TAG (537)		<i>orf1</i>	<i>Clostridium thermocellum</i> NCIB 10682	178/178 (100%)	Putative sensory transducer protein	Hazlewood <i>et al.</i> (1993)
<i>celI</i>	887	ATG (783)	TGA (3444)	AGGAGG (772)	<i>celI</i>	<i>Clostridium thermocellum</i> NCIB 10682	860/860 (100%)*	Endo- β -1,4-glucanase	Hazlewood <i>et al.</i> (1993)
<i>orf3</i>	400	ATG (4108)	TAG (5308)	AGGAGGT (4090)	CAC1651	<i>Clostridium acetobutylicum</i>	230/396 (58%)	Putative GTPase	Nolling <i>et al.</i> (2001)
<i>celN</i>	742	ATG (5699)	TAA (7925)	GGGGGT (5685)	<i>celI</i>	<i>Clostridium thermocellum</i>	294/625 (47%)	Endo- β -1,4-glucanase	Hazlewood <i>et al.</i> (1993); this study
<i>cseP</i>	533	ATG (9652)	TAA (8053)	GGAGGT (9664)	<i>cotH</i>	<i>Bacillus subtilis</i>	97/396 (24%)	Inner spore coat protein	Zilhao <i>et al.</i> (1999)
<i>orf6</i>		259 (incomplete)	ATG (10005)	GGAGG (9992)	<i>cwlU</i>	<i>Paenibacillus polymyxa</i>	41/143 (28%)	Autolysin protein	Ishikawa <i>et al.</i> (1999)

*Deletion and frame shift in *C. thermocellum* NCIB 10682.

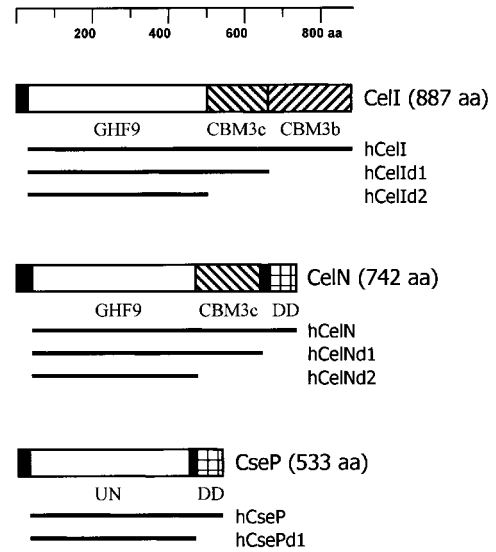


Fig. 2. Structure of the gene products Cell, CelN and CseP. LP, leader peptide; DD, dockerin domain; UN, module with unknown function. The extent of the amino acid sequences expressed by different deletion clones is indicated.

contains a spacer peptide (PT-box) and a dockerin module for attachment to the cellulosome (Fig. 2). The structural formula of Ct_{F7}-Cel9I is GH9/CBM3c/CBM3b, and that of Ct_{F7}-Cel9N is GH9/CBM3c/PT/DD, according to the recent nomenclature (Henrissat *et al.*, 1998), with GH designating the catalytic domain, PT the PT-box and DD the dockerin domain. Despite their similar structure Cel9I thus seems to be a free enzyme, whereas Cel9N might be integrated into the cellulosome.

The GHF9-CBM3c parts, common to both proteins, share 47% sequence identity. To compare their biochemical traits, *celI* and *celN* were recombinantly expressed and the His-tagged proteins purified (hCellI and hCelN). hCellI had a temperature optimum of 60 °C on PASC, 10 °C lower than that of hCelN; the optimal pH was 6.0 and 5.4, respectively, for hCellI and hCelN (Table 3). All activity determinations were performed in buffer containing calcium, because the exclusion of calcium reduced the activity of hCellI on PASC to one-tenth. hCelN was much less influenced.

Both enzymes were active on soluble β -1,4- and mixed-linkage glucans, but not on β -1,3-glucan or cellobiose (Table 3). The approximately 100:1 preference for soluble mixed-linkage glucans versus CM-cellulose was the same as described for CellI of *C. thermocellum* NCIB 10682 (Hazlewood *et al.*, 1993). The activity of both enzymes on partially regenerated cellulose (PASC) was much lower, as was generally found with endo-active cellulases. Both enzymes have almost undetectable activity on micro-crystalline cellulose (Avicel), but traces of cellobiose could be observed. However, these data do not support the

Table 3. Substrate specificity of Cel9I and Cel9N

Recombinant proteins hCellI and hCelN were incubated with the substrate and reducing sugars or *p*-nitrophenol released were determined. pH_{opt} and T_{opt} were determined with PASC. Specific activity was calculated in $U\ mg^{-1}$ and $U\ \mu mol^{-1}$, considering the molecular mass of hCellI (97 046 Da) and hCelN (80 699 Da). ND, activity not detectable (minimum detection value in parentheses).

Property	hCellI		hCelN	
pH_{opt}	6.0		5.4	
T_{opt} (°C)	60		70	
Substrate specificity:	($U\ mg^{-1}$)	($U\ \mu mol^{-1}$)	($U\ mg^{-1}$)	($U\ \mu mol^{-1}$)
Barley β -glucan	300	29 100	360	29 160
CM-cellulose	2.3	223	3.4	275
PASC	2.4	233	2.6	211
Avicel	ND (<0.01)	ND	ND (<0.01)	ND
Laminarin	ND (<0.01)	ND	ND (<0.01)	ND
<i>p</i> NP-G1	ND (<0.002)	ND	ND (<0.002)	ND
<i>p</i> NP-G2	ND (<0.002)	ND	ND (<0.002)	ND
<i>p</i> NP-G3	ND (<0.002)	ND	ND (<0.002)	ND
<i>p</i> NP-G4	0.016	1.6	0.013	1.1

assumption of an 'Avicelase' activity. hCellI showed no activity and hCelN showed low activity on cellotriose (G3), whereas higher cellodextrins were hydrolysed, with lower activity of hCellI on cellotetraose (G4). Cellopentaose (G5) was readily hydrolysed by both enzymes. *p*-Nitrophenol was released from the G5-analogue *p*NP-G4, whereas shorter aryl-cellodextrins were not or very slowly degraded. hCelN thus hydrolysed shorter cellodextrins more readily than hCellI.

Hydrolytic mode

To gather more information on the hydrolytic mode of action, the degradation products were analysed by TLC (Fig. 3). hCellI degraded G4 slowly to G3 and glucose (G1), less to cellobiose (G2), whereas hCelN yielded mainly G2 and much less G3 and G1. G5 was completely degraded by hCellI and hCelN to either G4 + G1 or G3 + G2, and hCelN split G4 almost completely to G2.

PASC was hydrolysed by hCellI to G4 which subsequently was degraded to G3 + G1 and less to G2 (Fig. 4); no larger cellodextrins could be detected even early in the reaction. This suggests a processive action mode of CellI. In contrast, hCelN produced larger cellodextrins first from PASC, which were later on degraded to G4 (and subsequently to G2) and less to G3 and G1. This indicates a non-processive endohydrolytic action mode for CelN.

To investigate the endo-mode of Cel9N further and to compare it with that of Cel9I, the hydrolytic mode of both enzymes was investigated on PASC. An adaptation of the assay by Irwin *et al.* (1993) was used to measure the occurrence of reducing ends in the soluble and the insoluble fraction on incubation with the enzymes. An endoglucanase should – at least initially – produce new reducing ends only in the insoluble fraction, whereas a processive glucanase (exoglucanase) would immediately produce short cellodextrins which appear in the soluble

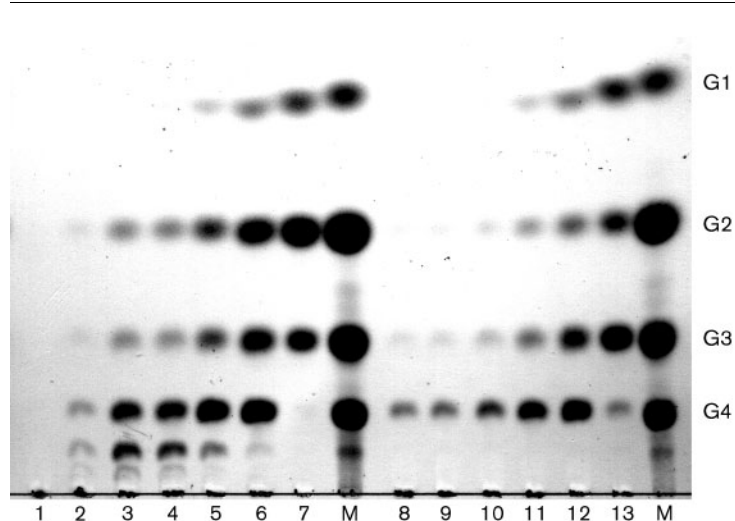


Fig. 3. Hydrolysis of PASC. TLC of the products from a digestion with hCelN (lanes 1–7) and hCellI (lanes 8–13). Incubation times increase from left to right: 0.5, 2, 5, 15, 45, 120 min and overnight.

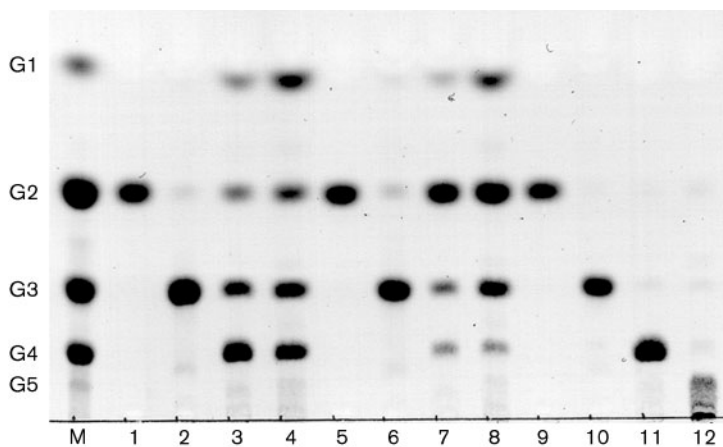


Fig. 4. TLC of enzymic degradation products produced by hCell (lanes 1–4) and hCelN (lanes 5–8). Substrates were cellobiose (G2; lanes 1, 5), cellotriose (G3; lanes 2, 6), cellotetraose (G4; lanes 3, 7) and cellopentaose (G5; lanes 4, 8). Markers are G2 (lane 9), G3 (lane 10), G4 (lanes 11) and G5 (lane 12).

fraction. hCell produced reducing residues in the soluble as well as in the insoluble phase, whereas with hCelN additional reducing ends were initially found only in the insoluble phase (Fig. 5). Cel9N thus behaves as expected for a non-processive endoglucanase. The simultaneous appearance of reducing power in the soluble and insoluble phase indicates that hCell has to be regarded as a processive endoglucanase. Both enzymes thus differ dramatically in the hydrolytic mode, despite the fact that they are closely related and possess the same GHF9-CBM3c arrangement.

Influence of the C-terminal modules

The influence of the C-terminal modules on enzymic activity was investigated by constructing deletion proteins of Cel9I and Cel9N (Fig. 2). The purified truncated proteins

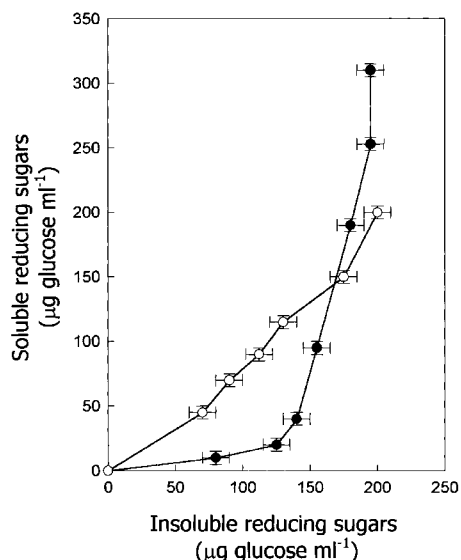


Fig. 5. Estimation of the increase in reducing power (glucose equivalents) in PASC by hCell (open circles) and hCelN (closed circles). Soluble and insoluble fractions were separated by centrifugation and reducing equivalents were estimated by using the dinitrosalicylic acid method.

had the expected molecular mass in denaturing SDS-PAGE (data not shown). hCelId1 and hCelNd1 have a homologous structure containing the GHF9 and the CBM3c modules; the d2 derivatives contained only the catalytic modules. The pH optimum of hydrolysis did not change with the deletion of the C-terminal modules. However, the temperature optimum of enzymic activity dropped by 15 °C for both proteins when the CBM3c module was removed. If assayed at 60 and 70 °C, respectively, both proteins were inactive. The specific activity at the temperature optimum was increased in the d1 deletions roughly in accordance with the decrease in molecular mass. Despite the drop in the temperature used for the estimation and the further decrease in molecular mass of the d2 deletion proteins, the specific activity on barley β -glucan dropped only by about 7 % in hCelNd2 and even increased slightly in hCelId2. The ratio of the CM-cellulose to β -glucan activity was reduced from 7.7 to 6.1 % in the deletion mutant hCelId2. No significant change in that ratio (9.5 vs 9.6 %) was seen with deletions of Cel9N (data not shown).

CseP: a non-catalytic cellulosome component

The fifth reading frame in the sequenced genomic fragment was called *cseP* (cellulosomal element P) due to the presence of a dockerin module type I in its C terminus. *cseP* is separated from the preceding gene, *celN*, by a potential bidirectional transcription terminator and is transcribed from the opposite DNA strand. The complete gene was expressed in *E. coli* (hCseP) and the gene product was purified by His-tag affinity chromatography. All cellulosome components containing a dockerin type I module investigated so far had hydrolytic activity. The sequence of CseP was compared to the databases, but not even a weak sequence similarity to any known hydrolytic enzyme could be detected. Nevertheless, hCseP was incubated with 10 polysaccharides, 14 arylglycosides and *p*NP-acetate (Table 4). No hydrolytic activity could be detected with these substrates, neither by release of the chromophore aglycon nor with reducing sugars. CseP presumably is not hydrolytically active.

However, the amino acid sequence is sufficiently homologous to the complete CotH protein from *Bacillus*

Table 4. Detection of catalytic activity of CseP: substrates used

pNP-glycosides		Polysaccharides	
α-Glycosides	α-L-Arabinofuranoside	β-1,4	Avicel (microcrystalline)
	α-L-Arabinopyranoside		PASC (amorph. cellulose)
	α-L-Rhamnopyranoside		CM-cellulose
β-Glycosides	β-D-Xylopyranoside	β-1,3-1,4	Mannan
	β-D-Galactopyranoside		Xylan (arabinoxylan)
	β-D-Lactoside		Chitin
	β-D-Fucopyranoside		Barley β-glucan
	β-D-Mannopyranoside	β-1,3	Laminarin
	β-D-Triacetylchitotriose		Curdlan
	β-D-Glucopyranoside	β-1,6	Pustulan
	β-D-Cellobioside		
	β-D-Cellotrioside		
	β-D-Cellotetraoside		
Ester	Acetate		

subtilis (Table 2), a structural component of the spore coat (Zilhao *et al.*, 1999). Moreover, a BLAST search with genomic sequences showed significant sequence homology to a number of putative genes of unknown function in the following bacteria (expressed as sequence identity/fragment length): *Clostridium difficile* (29 %/410), *Clostridium perfringens* (26 %/189), *Bacillus cereus* (23 %/411), *Bacillus anthracis* (24 %/321), *Fibrobacter succinogenes* (23 %/276), *Ruminococcus albus* (21 %/293), and two genes in the archaeon *Methanosarcina acetivorans* (25 %/221 and 24 %/256).

To verify the expression of *cseP* in *C. thermocellum* and to prove the presence of CseP in the cellulosome, polyclonal antibodies were raised against hCsePd1. This mutant protein is stripped of the dockerin domain which is present in all cellulosomal components and might cause cross-reactions. The antibodies were used to selectively label the CseP protein in a denaturing SDS-PAGE of cellulosomal proteins and as a control, the purified hCsePd1 protein (Fig. 6). Among the separated cellulosomal proteins only one band of 61 kDa was recognized by the antibody which was in accordance with the deduced molecular mass for native CseP (61 532 Da). This indicated that the antibody was specific and that CseP was present in the cellulosome.

DISCUSSION

A 10 kbp fragment of the *C. thermocellum* F7 genome was sequenced. It encodes three clustered genes with a potential role in the hydrolysis of cellulose. In addition to the cellulosomal protein CseP, it encodes the two endoglucanases Cel9I and Cel9N which are separated by a putative gene for a GTPase, possibly a component of a sugar transport system. The core protein GHF9-CBM3c of the two endoglucanases have 47 % identical aa residues (in over 300) and are the most similar sequences in a GenBank

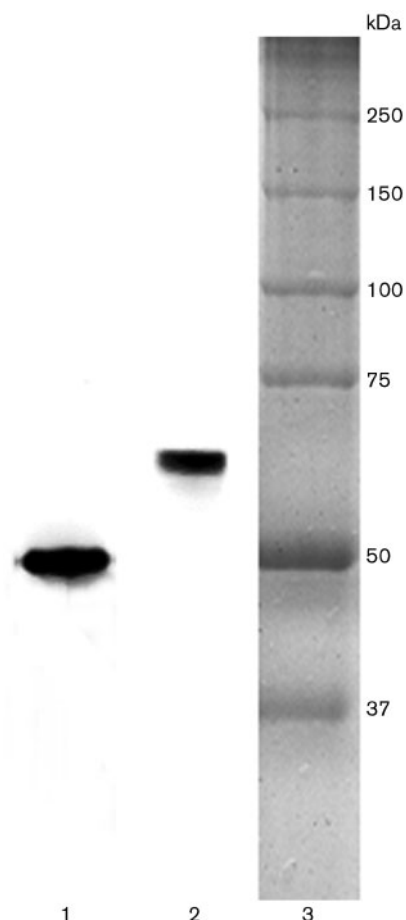


Fig. 6. Detection of CseP in the cellulosome. Western blot of recombinant CseP and the cellulosomal preparation with anti-hCsePd1 antibodies, separated in denaturing SDS-PAGE. Lanes: 1, hCsePd1; 2, cellulosome; 3, molecular mass markers.

search, so that a gene duplication has to be assumed, which might have been followed by module shuffling. The module arrangement of the two cellulases (a catalytic module followed by a CBM3c module, formerly called C' domain) corresponds to the theme B enzymes discussed by Bayer *et al.* (2000). However, the two related proteins have different C-terminal modules.

Ct_{F7}-Cel9I is identical to Ct_{NCIB}-Cell from *C. thermocellum* NCIB 10682 (Hazlewood *et al.*, 1993) except for an accidental deletion of a 53 bp fragment between two 6 bp direct repeats near the 3' end of the gene (TTTCAA; bp 2837–2890 in L04735), possibly due to an error in sequence alignment in the Ct_{NCIB}-Cell sequence. This introduced a frameshift in the C-terminal part of the protein. The high degree of sequence identity between large stretches of DNA derived from different strains of *C. thermocellum* (3625 bp) emphasizes the close relationship between different bacterial isolates on different continents at least in the Northern hemisphere. Such high similarity was also found with other genes when sequenced from two strains, like *celA-chiA*, *celC*, *licA*, *licB* and others (Zverlov *et al.*, 2002; K. P. Fuchs, V. V. Zverlov, G. A. Velikodvorskaya, F. Lottspeich & W. H. Schwarz, unpublished results). Because of the identity with the already described Ct_{ATTC}-Cel9I sequence, Ct_{F7}-Cel9I was investigated only so far as to compare it with the newly identified Ct_{F7}-Cel9N.

Next to themselves, the sequences of Cel9I and Cel9N are most similar to the 1,4- β -glucanases CelA from *Anaerocellum thermophilum* (GenBank accession no. Z86105, 47 % identity), CelA and CelE from *Caldicellulosiruptor* sp. Tok7B.1 (L32742 and AF078042, 46 %), CelF from *C. thermocellum* (X60545, 45 %), a cellulase from *Bacillus* sp. BP-23 (AJ133614, 46 %) and Avicelase I (CelZ) from *Clostridium stercorarium* (S12021, 45 %). All these enzymes have a GH9-CBM3c arrangement of modules and show a comparable substrate specificity, being slightly active on amorphous and crystalline cellulose, but with the highest activity on barley β -glucan. Within the GH9-CBM3c group of cellulases the hydrolytic mode of the E4-glucanase of *Thermobifida fusca* has been investigated in the greatest detail (Irwin *et al.*, 1998). These authors suggested an accessory role for the CBM3c module in holding a single cellulose chain and feeding it into the active site pocket of E4, after the endoglucanolytic activity of the catalytic centre has produced a new nick in the cellulose molecule. This makes the enzyme a processive endoglucanase, a trait not yet tested for the other members of that family.

The processive endoglucanase type of hydrolytic activity could be verified for Ct_{F7}-Cel9I by assaying the release of reducing ends in the soluble versus the insoluble fraction of digested PASC. However, no processivity could be detected with Cel9N. This was surprising in the light of the interpretation of the three-dimensional data from endoglucanase E4 which would suggest a similar function of the CBM3c module for the other enzymes with an equivalent module arrangement. However, the accessory function of

the CBM3c module depends on correct spacing and orientation towards the active site pocket of the GH9 module which could be easily modified by a very limited number of mutations. With respect to substrate specificity, the CBM3c module might have a different or no function in Cel9N.

The different hydrolysis mode of Cel9I and Cel9N was corroborated by the analysis of the occurrence of hydrolysis products. By its processive action, Cel9I produces mainly cellotetraose which in a second, non-processive step is slowly hydrolysed further. The products of Cel9N initially are long oligosaccharides which are continuously hydrolysed to smaller ones. On small oligosaccharides, such as cellotetraose, Cel9N obviously has a higher activity compared with Cel9I.

Another hint at the differences in processivity comes from the effect calcium ions have on the hydrolytic activity. Whereas the activity of hCell on PASC was inhibited by removing Ca²⁺ ions, no or little effect was observed on barley β -glucan. In contrast, Ca²⁺ depletion had no effect with hCelN (data not shown). CBM family 3 is known to depend in its binding function on the presence of Ca²⁺ ions, and binding a cellulose strand supposedly is an integral part of Cel9I action on cellulose, but presumably not on mixed-linkage β -1,3-1,4-glucan. Abolishing the binding to cellulose would thereby negatively influence the activity on PASC. The endo-mode of Cel9N may be entirely independent of the activity of a binding module and thus was not influenced by Ca²⁺. Measurements with Avicel have shown that the CBM3c module in hCelN does not bind to crystalline cellulose (data not shown). The so-called 'helper-module', CBM3c, thus may not play a role of in cellulose hydrolysis by Cel9N.

However, according to the three-dimensional model of endoglucanase E4 from *Thermobifida fusca*, the CBM3c module is tightly attached to the catalytic module and seems to stabilize the integrity of the catalytic protein module. The protection from denaturation could be a consequence of this close association, as seen by the thermolability of the deletion proteins hCelId2 and hCelNd2. The function of a binding module as a 'thermostabilizing domain' was also observed in CelZ (*C. stercorarium*; Riedel *et al.*, 1998) in the same enzyme family and in other enzymes like a xylanase from *Thermotoga neapolitana* and *Thermotoga maritima* (Zverlov *et al.*, 1996; Meissner *et al.*, 2000).

It is interesting to note that all bacterial cellulolytic enzyme systems contain, in addition to the indispensable GHF48 enzyme, one or more GHF9 enzymes (Schwarz, 2001). A variety of GHF9-based endoglucanases seem to be necessary to fulfil the needs of effectively attacking a crystalline cellulosic substrate which displays a multitude of topologies. This variety in enzymes is verified by different module combinations, and the greatest diversity is seen in the GHF9 enzyme family. In the cellulosome of *C. thermocellum* so far seven GHF9 enzymes have been identified and investigated: two theme C enzymes (CelD, CelJ), two theme D enzymes

(CbhA, CelK) and three theme B enzymes (CelF, CelQ, CelN), according to the classification of Bayer *et al.* (2000). An additional GHF9 enzyme, Cel9I, is a free cellulolytic component which might assist in the process of attacking different substrate topologies. Cel9I therefore is equipped with a substrate-binding module of its own (CBM3b) which makes it independent of the cellulosomal substrate-binding capacity, whereas Cel9N can rely on the strongly binding CBM3 attached to the CipA component of the cellulosome. The differences between the closely related enzymes Cel9I and Cel9N reflect a potential for an even greater variability in the ability of the GHF9 cellulases to hydrolyse different topologies in their substrate cellulose.

The gene *cseP*, which is also present in the sequenced 10 kb fragment, can be predicted to be transcribed independently of *celI* and *celN*. The protein CseP consists of an N-terminal module without any homology to hydrolytic enzymes in the databases and a dockerin type I module, which is regarded as an indicator for binding to the cohesin modules located on the cellulosomal core protein CipA. The gene product, CseP, was further investigated as a potentially interesting new cellulosome component. Our investigations with 25 substrates representing a wide range of glycans and glycosides did not reveal any hydrolytic activity. However, CseP showed significant homology to CotH from *Bacillus subtilis*, which is a structural component of the spore and is needed for proper formation of the spore coat (Zilhao *et al.*, 1999). In addition, sequence similarity with other reading frames from genomic sequences of spore- or cellulosome-forming bacteria was observed. It was also similar to an uncharacterized protein of a single archaeon, *Methanosarcina acetivorans*, which is unique among the *Archaea* in forming complex multi-cellular structures (Galagan *et al.*, 2002). This implies the involvement of this family of proteins in the building of multi-protein structures.

One could argue that the occurrence of *cseP* was a mishap in the evolution of a cellulosomal protein, created by unsuccessful module shuffling. Such a gene would not be expected to be expressed. However, Western blots with anti-hCsePd1 antibodies verified that *cseP* is expressed and the protein successfully incorporated into the cellulosome. It was present in the cellulosome as a single band in significant amounts. These data indicate a potential role of CseP in the structural integrity of the assembly of the cellulosome and open new perspectives for the investigation of cellulosomal structure and function.

ACKNOWLEDGEMENTS

This work was supported by a grant from RFFI to G. A. V (Ref. 00-04-48197) and grants from the Deutsche Forschungsgemeinschaft DFG (Ref. 436 RUS 17/88/00) and the A-v-Humboldt Foundation to V. V. Z. We are very grateful to W. L. Staudenbauer for many stimulating discussions.

REFERENCES

- Baker, A. A., Helbert, W., Sugiyama, J. & Miles, M. J. (2000). New insight into cellulose structure by atomic force microscopy shows the I_x crystal phase at near-atomic resolution. *Biophys J* **79**, 1139–1145.
- Barr, B., Hsieh, Y. L., Ganem, B. & Wilson, D. B. (1996). Identification of two functionally different classes of exocellulases. *Biochem* **35**, 586–592.
- Bayer, E. A., Morag, E., Lamed, R., Yaron, S. & Shoham, Y. (1998). Cellulosome structure: four-pronged attack using biochemistry, molecular biology, crystallography and bioinformatics. In *Carbohydrases from Trichoderma reesei and Other Microorganisms*, pp. 39–65. Edited by M. Claeysens, W. Nerinckx & K. Piens. London: Royal Society of Chemistry.
- Bayer, E. A., Shoham, Y. & Lamed, R. (2000). Cellulose-decomposing prokaryotes and their enzyme systems. In *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community, 3rd edn* (latest update release 3, 7 September 2001). Edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt. New York: Springer.
- Béguin, P., Chauvaux, S., Chaverocche, M.-K., Guglielmi, G., Kataeva, I., Leibovitz, E. & Miras, I. (1998). The cellulosome: a versatile system for coupling cellulolytic enzymes and attaching them to the cell surface. In *Carbohydrases from Trichoderma reesei and Other Microorganisms*, pp. 66–72. Edited by M. Claeysens, W. Nerinckx & K. Piens. London: Royal Society of Chemistry.
- Boisset, C., Armand, S., Drouillard, S., Chanzy, H., Driguez, H. & Henrissat, B. (1998). Structure–function relationships in cellulases: the enzymatic degradation of insoluble cellulose. In *Carbohydrases from Trichoderma reesei and Other Microorganisms*, pp. 124–132. Edited by M. Claeysens, W. Nerinckx & K. Piens. London: Royal Society of Chemistry.
- Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. & Gilbert, H. J. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* **331**, 775–781.
- Bumazkin, B. K., Velikodvorskaya, G. A., Tuka, K., Mogutov, M. A. & Strongin, A. Y. (1990). Cloning of *Clostridium thermocellum* endoglucanase genes in *Escherichia coli*. *Biochem Biophys Res Commun* **167**, 1057–1064.
- Carrard, G., Koivula, A., Söderlund, H. & Béguin, P. (2000). Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc Natl Acad Sci U S A* **97**, 10342–10347.
- Galagan, J. E., Nusbaum, C., Roy, A. & 72 other authors (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* **12**, 532–542.
- Hazlewood, G. P., Davidson, K., Laurie, J. I., Huskisson, N. S. & Gilbert, H. J. (1993). Gene sequence and properties of Cell, a family E endoglucanase from *Clostridium thermocellum*. *J Gen Microbiol* **139**, 307–316.
- Henrissat, B., Teeri, T. T. & Warren, R. A. J. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell wall of plants. *FEBS Lett* **425**, 352–354.
- Irwin, D., Walker, L., Spezio, M. & Wilson, D. (1993). Activity studies of eight purified cellulases: specificity, synergism, and binding domain effects. *Biotechnol Bioeng* **42**, 1002–1013.
- Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A. & Wilson, D. B. (1998). Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J Bacteriol* **180**, 1709–1714.
- Ishikawa, S., Kawahara, S. & Sekiguchi, J. (1999). Cloning and expression of two autolysin genes, *cwlU* and *cwlV*, which are

- tandemly arranged on the chromosome of *Bacillus polymyxa* var. *colistinus*. *Mol Gen Genet* **262**, 738–748.
- Johnson, E. A., Madia, A. & Demain, A. L. (1981).** Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile *Clostridium thermocellum*. *Appl Environ Microbiol* **41**, 1060–1062.
- Laemmli, U. K. (1970).** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- Lamed, R., Setter, E., Kenig, R. & Bayer, E. A. (1983).** The cellulosome – a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnol Bioeng* **13**, 163–181.
- Mechaly, A., Yaron, S., Lamed, R., Fierobe, H.-P., Bélaich, A., Bélaich, J.-P., Shoham, Y. & Bayer, E. A. (2000).** Cohesin–dockerin recognition in cellulosome assembly: experiment versus hypothesis. *Proteins* **39**, 170–177.
- Meissner, K., Wassenberg, D. & Liebl, W. (2000).** The thermostabilizing domain of the modular xylanase XynA of *Thermotoga maritima* represents a novel type of binding domain with affinity for soluble xylan and mixed-linkage beta-1, 3/beta-1,4-glucan. *Mol Microbiol* **36**, 898–912.
- Morag, E., Bayer, E. A. & Lamed, R. (1992).** Affinity digestion for the near-total recovery of purified cellulosome from *Clostridium thermocellum*. *Enzyme Microb Technol* **14**, 289–292.
- Nolling, J., Breton, G., Omelchenko, M. V. & 16 other authors (2001).** Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* **183**, 4823–4838.
- Nutt, A., Sild, V., Pettersson, G. & Johansson, G. (1998).** Progress curves. A mean for functional classification of cellulases. *Eur J Biochem* **258**, 200–206.
- Pagès, S., Gal, L., Bélaich, A., Gaudin, C., Tardif, C. & Bélaich, J.-P. (1997).** Role of scaffolding protein CipC of *Clostridium cellulolyticum* in cellulose degradation. *J Bacteriol* **179**, 2810–2816.
- Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Bélaich, J. P., Driguez, H. & Haser, R. (2000).** Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. *Biochem* **39**, 11238–11246.
- Reverbel-Leroy, C., Pages, S., Bélaich, A., Bélaich, J.-P. & Tardif, C. (1997).** The processive endocellulase CelF, a major component of the *Clostridium cellulolyticum* cellulosome: purification and characterization of the recombinant form. *J Bacteriol* **179**, 46–52.
- Riedel, K., Ritter, J., Bauer, S. & Bronnenmeier, K. (1998).** The modular cellulase CelZ of the thermophilic bacterium *Clostridium stercorarium* contains a thermostabilizing domain. *FEMS Microbiol Lett* **164**, 261–267.
- Salamitou, S., Lemaire, M., Fujino, T., Ohayon, H., Gounon, P., Béguin, P. & Aubert, J.-P. (1994).** Subcellular localization of *Clostridium thermocellum* ORF3p, a protein carrying a reporter for the docking sequence borne by the catalytic components of the cellulosome. *J Bacteriol* **176**, 2828–2834.
- Schwarz, W. H. (2001).** The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol* **56**, 634–649.
- Sedmak, J. J. & Grossberg, S. E. (1977).** A rapid, sensitive assay for protein using Coomassie brilliant blue G250. *Anal Biochem* **79**, 544–552.
- Teeri, T. T. (1997).** Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends Biotechnol* **15**, 160–167.
- Tomme, P., Boraston, A., McLean, B. & 7 other authors (1998).** Characterization and affinity applications of cellulose-binding domains. *J Chromatogr* **715**, 283–296.
- Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996).** Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J* **15**, 5739–5751.
- Wood, T. M. (1988).** Preparation of crystalline, amorphous and dyed cellulase substrates. *Methods Enzymol* **160**, 19–25.
- Wood, T. M. & Bhat, K. M. (1988).** Methods for measuring cellulase activities. *Methods Enzymol* **160**, 87–112.
- Zilhao, R., Naclerio, G., Heriques, A. O., Baccigalupi, L., Moran, C. P., Jr & Ricca, E. (1999).** Assembly requirements and role of CotH during spore coat formation in *Bacillus subtilis*. *J Bacteriol* **181**, 2631–2633.
- Zverlov, V. V., Piotukh, K., Dakhova, O., Velikodvorskaya, G. & Borris, R. (1996).** The multidomain xylanase A of the hyperthermophilic bacterium *Thermotoga neapolitana* is extremely thermostable. *Appl Microbiol Biotechnol* **45**, 245–247.
- Zverlov, V. V., Velikodvorskaya, G. A., Schwarz, W. H., Kellermann, J. & Staudenbauer, W. L. (1999).** Duplicated *Clostridium thermocellum* cellobiohydrolase gene encoding cellulosomal subunits S3 and S5. *Appl Microbiol Biotechnol* **51**, 852–859.
- Zverlov, V. V., Fuchs, K. P. & Schwarz, W. H. (2002).** Chi18A, the endochitinase in the cellulosome of the thermophilic, cellulolytic bacterium *Clostridium thermocellum*. *Appl Environ Microbiol* **68**, 3176–3179.