Vladimir Zverlov and Wolfgang Schwarz

Research Group Microbial Biotechnology, Technical University Munich, Am Hochanger 4, D-85350 Freising-Weihenstephan, Germany
e-mail: zverlov@hotmail.com

Genes for two new cellulosomal components, cellulase Cel9N and a non-hydrolytic protein CseP, are located downstream of *celI* in the genome of *Clostridium thermocellum*

Summary

The sequence of a gene cluster containing the genes *celI*, *celN* and *cseP* was determined from *Clostridium thermocellum* F7. Cel9I and Cel9N belong to glycosyl hydrolase family 9 with a CBM3c module, but show considerable differences: Cel9I is a non-cellulosomal protein with an additional carbohydrate binding module (CBM3b), whereas Cel9N contains a dockerin module and no CBM3b. Moreover, whereas Cel9I is a processive endo-glucanase, Cel9N is an endo-glucanase with no signs of processivity. CseP is the first component of the cellulosome (after CipA) for which no catalytic activity could be shown. Its sequence is homologous to the spore protein CotH of Bacillus subtilis, indicating its role in the assembly of the cellulosome.

Introduction

To hydrolyze insoluble, especially crystalline substrates, the anaerobic, thermophilic bacterium Clostridium thermocellum possesses a structurally coherent complex of enzymes arranged around a non-catalytic scaffolding protein. The catalytic components bind to the scaffoldin with the aid of a dockerin module. The large multienzyme complex, the cellulosome, ensures the synchronous co-operation of enzymes. Their different mode of activity results in the complete decomposition of the recalcitrant, crystalline substrate. In addition to the cellulosome, *C. thermocellum* also produces soluble enzymes which might aid in the degradation of cellulose.

Results

A clone, pCU108, expressing "cellulase" activity was isolated previously from a genomic library of the thermophilic *Clostridium thermocellum* strain F7. The upstream 3685 bp (of 7583 bp) were nearly identical to sequences acc.Nr. L04735 and L04736 of *C. thermocellum* strain ATCC 27405 containing the endo-glucanase gene *celI* (Hazlewood et al., 1993). The region further downstream of *celI* was obtained by cloning a fragment from restriction-endonuclease digested genomic DNA (partial Sau3A), which hybridized against the 3´-part of pCU108. A total of 10 782 bp were sequenced (genebank acc.Nr. AJ275974) containing six open reading frames (Fig. 1; Tab.1).

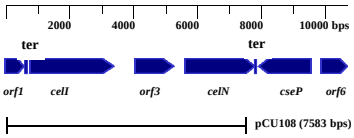

Fig. 1: Physical map of the *celI-celN-cseP* region in the *Clostridium thermocellum* chromosome. The extension of the initial clone pCU108 is indicated. *ter*, palindromic structure, potential transcription terminator.

	ORF1	CelI	ORF3	CelN	CseP	ORF6
sise (aa) Start codon Stop codon Ribos.-bind. site	178 (incompl) TAG (537)	887 ATG (783) TGA (3444) AGGAGG (772)	400 ATG (4108) TAG (5308) AGGAGGT (4090)	742 ATG (5699) TAA (7925) GGGGGT (5685)	533 ATG (9652) TAA (8053) GGAGGT (9664)	259 (incompl) ATG (10005) GGAGG (9992)
Homologous gene	*orf1*	*celI*	CAC1651	*celI*	*cotH*	*cwlU*
Bacterium	*C. thermocellum* NCIB 10682	*C. thermocellum* NCIB 10682	*Clostridium acetobutylicum*	*Clostridium thermocellum*	*Bacillus subtilis*	*Paenibacillus polymyxa*
Protein sequence identity	178/178 (100%)	860/860 (100%)*	230/396 (58%)	294/625 (47%)	97/396 (24%)	41/143 (28%)
Gene product	Putative sensory transducer protein	endo-1,4-glucanase	Putative GTPase	endo-1,4-glucanase	Inner spore coat protein	Autolysin protein
Reference or accession no.	Hazlewood et al. , 1993	Hazlewood et al. 1993	Nolling et al., 2001	Hazlewood et al. 1993; this paper	Zilhao et al., 1999	Ishikawa et al., 1999
		* deletion and frame shift in C.th. NCIB10682				

Table 1: Open reading frames in the *C. thermocellum celI* region: Homology to reported genes of other bacteria. Numbers in brackets designate the first base in *the C. thermocellum* sequence (accNo. AJ275974).

The gene product of *celI*, Cel9I, is a modular protein consisting of a leader peptide, a catalytic module of glycosyl hydrolase family 9 (GHF 9), a carbohydrate binding module CBM 3c and a CBM 3b module (Fig. 2). The gene product of *celN*, Cel9N, has a similar structure, but instead of the CBM3b module it contains a spacer peptide (PT-box) and a dockerin module (DD) for attachment to the cellulosome. Despite their similar structure Cel9I seems to be a free enzyme, whereas Cel9N is integrated in the cellulosome.

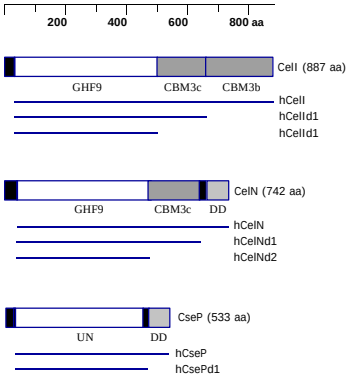

Figure 2: Structure of the gene products CelI, CelN and CseP. LP, leader peptide; CBM3c, carbohydrate binding module family 3, subfamily c; GHF9, glycosyl hydrolase family 9; DD, dockerin domain. The extension of the amino acid sequences expressed by different deletion clones is indicated.

The GHF9-CBM3c parts, common to both proteins, share 47 % sequence identity. To compare their biochemical traits, *celI* and *celN* were recombinantly expressed and purified as His-tagged proteins (Tab. 2).

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

The pH optimum and the specific activity were not changed by deleting the C-terminal modules. However, the temperature optimum of hydrolytic activity on barley ß-glucan dropped by 15 °C with both proteins, if the CBM 3c module was removed.

	hCelI	hCelN
pH_{opt}	6,0	5,4
T_{opt} (°C)	60	70
	Substrate specificity (U mg^{-1})	
Barley ß-glucan	300	360
CMC	2,3	3,4
PASC	2,4	2,6
Avicel	nd (< 0,001)	nd (< 0,001)
Laminarin	nd (< 0,002)	nd (< 0,002)
pNP-G1	nd (< 0,002)	nd (< 0,002)
pNP-G2	nd (< 0,002)	nd (< 0,002)
pNP-G3	nd (< 0,002)	nd (< 0,002)
pNP-G4	0,016	0,013

Table 2. Substrate specificity. Recombinant proteins hCelI and hCelN were incubated and reducing sugars or *p*-nitrophenol released were determined. pH_{opt} and T_{opt} were determined with PASC. nd, activity not detectable.

To investigate the hydrolytic mode, the products of hydrolysis were identified by thin layer chromatography. The amorphous polymer PASC was hydrolyzed by hCelI to G4; no larger cellodextrins were detected even early in the reaction. G4 was subsequently degraded to G3+G1 and less G2 (Fig. 3). This suggests a processive action mode of CelI. In contrast, hCelN produced from PASC larger cellodextrins first, which were lateron degraded to G4 and subsequently to G2, and less G3 and G1. This suggests an endo-hydrolytic mode of action for CelN.

To investigate the endo-mode of Cel9N further and to compare it with that of Cel9I, the processivity of both enzymes was compared on PASC by measuring the occurrence of reducing ends in the soluble and the insoluble fraction on incubation with the enzyme (Fig. 4). Cel9N thus behaves as expected for an endo-glucanase. The simultaneous appearance of reducing power in the soluble and insoluble phase with hCelI indicates that this enzyme can be regarded as a processive endo-glucanase. Both enzymes thus differ in the hydrolytic mode, despite the fact that they are closely related and possess the same GHF 9-CBM 3c arrangement.

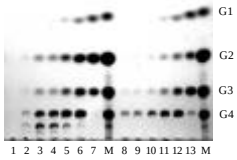

Figure 3: Hydrolysis of phosphoric acid swollen cellulose (PASC). Thin layer chromatography of the products from a digestion with hCelN (lanes 1-7) and hCelI (lanes 8-13). Incubation times are increasing from left to right from 30 sec (lanes 1 and 8) to overnight (lanes 7 and 13).

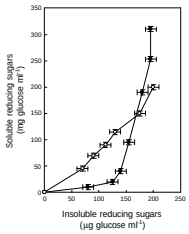

Figure 4: Estimation of the increase in reducing ends (glucose equivalents) in PASC by hCelI (open circles) and hCelN (closed circles). Soluble and insoluble fractions were separated and reducing equivalents were estimated.

The fifth reading frame in the sequenced genomic fragment was called *cseP*, **c**ellulo**s**omal **e**lement **P**. All so far investigated dockerin containing cellulosome components had hydrolytic activity. CseP was therefore incubated with 10 polysaccharides, 14 aryl-glycosides and pNP-acetate, but no hydrolytic activity could be detected, in agreement with the lack of sequence homology to any hydrolytic enzymes. However, the amino acid sequence of the N-terminal module is sufficiently homologous to the complete CotH protein from *Bacillus subtilis* (table 2), a structural component of the spore coat. CseP also had significant sequence homology to a number of putative proteins of unknown function, derived from genomic sequences of the following bacteria (sequence identity in % / fragment length): *Clostridium difficile* (29%/410), *C. perfringens* (26%/189), *Bacillus cereus* (23%/411), *B. anthracis* (24%/321), *Streptococcus gordonii* (22%/328), *Fibrobacter succinogenes* (23%/276), *Ruminococcus albus* (21%/293), and the archaeon *Methanosarcina acetivorans* (25%/221 and 24%/256).

The expression of the *cseP* gene and the presence of CseP in the cellulosome was shown by Western blot with polyclonal antibodies against hCsePd1 which is stripped of the dockerin domain. In the separated cellulosomal proteins only one band of 61 kDa was recognized, in accordance with the expected molecular mass for native CseP (61 532 Da). This indicated that CseP is present in the cellulosome.

The absence of hydrolytic activity and the homology to the spore-assembly protein CotH of *Bacillus subtilis* implies a similar role of CseP in the assembly of the cellulosome.