Lic16A of Clostridium thermocellum, a non-cellulosomal, highly complex endo-β-1, 3-glucanase bound to the outer cell surface

Klaus-Peter Fuchs,1 Vladimir V. Zverlov,2 Galina A. Velikodvorskaya,2 Friedrich Lottspeich3 and Wolfgang H. Schwarz1

1Research Group Microbial Biotechnology, Technical University of Muenchen, Am Hochanger 4, D-85350 Freising, Germany
2Institute of Molecular Genetics, Russian Academy of Science, Kurchatov Sq., 123182 Moscow, Russia
3Max-Planck-Institute for Biochemistry, Martinsried, Germany

INTRODUCTION

Non-cellulosic carbohydrates constitute the second largest fraction of biomass on earth. In plants they are for example important structural components in the cell wall or nutritional components of the endosperm. However, they are extraordinarily heterogeneous, consisting of a number of different sugar residues which are partially derivatized and linked. Complex systems of distinct enzymes are necessary to hydrolyse each of their constituents. One fraction of these polysaccharides, β-1,3-glucan, can be isolated in pure form, for example from the reproductive structures of plants (callose), from the marine macro-alga Laminaria saccharina (laminarin), from the cell walls of yeasts and fungi (zymosan or schizophyllan) or as an exopolysaccharide from bacteria like Alcaligenes faecalis var. myxogenes (curdlan) (Harada, 1992). β-1,3-Glucosidic linkages are also found in the linear mixed-linkage β-glucans of the Graminaceae (e.g. barley) or in the lichen Cetraria islandica (lichenan), alternating with β-1,4-linkages (30 : 70 %).

Bacteria hydrolyse β-1,3-linkages by specific endoglucanases. Some of them have been characterized, e.g. from Bacillus circulans, Clostridium thermocellum, Rhodothermus marinus or Thermotoga maritima (Schwarz et al., 1988a; Spilliaert et al., 1994; Zverlov et al., 1997a). They are biochemically classified either as 1,3-β-D-glucan glucohydrolases (EC 3.2.1.39) or as 1,3-(1,3;1,4)-β-D-glucan 3(4)-glucanohydrolases (EC 3.2.1.6). Their sequences are homologous and belong to glycosyl hydrolase family GHF16 (Henrisatt & Bairoch, 1996). But despite a similar protein fold, the β-1,3-glucanases cleave only β-1,3-linkages, whereas the β-1,3-1,4-glucanases hydrolyse the β-1,4-linkage adjacent to a β-1,3-linkage (Keitel et al., 1993).

Abbreviations: CBM, carbohydrate-binding module; DNSA, dinitrosalicylic acid; GHF, glycosyl hydrolase family; GuHCl, guanidinium hydrochloride; SLH, S-layer homologous module.

The GenBank accession numbers for the sequences reported in this paper are AJ307315 and X89732.

Clostridium thermocellum produces one major β-1,3-glucanase. Genomic DNA fragments containing the gene were cloned from two strains, DSM12377 (6848 bp) and F7 (9766 bp). Overlapping sequences were 99-9 % identical. The nucleotide sequences contained reading frames for a putative transposase, endo-β-1,3-1,4-glucanase CeIC, a putative transcription regulator of the LacI type, β-1,3-glucanase Lic16A and a putative membrane protein. The licA genes of both strains encoded an identical protein of 1324 aa with a calculated molecular mass of 148 kDa. Lic16A is an unusually complex protein consisting of a leader peptide, a threefold repeat of an S-layer homologous module (SLH), an unknown module, a catalytic module of glycosyl hydrolase family 16 and a fourfold repeat of a carbohydrate-binding module of family CBM4a. The recombinant Lic16A protein was characterized as an endo-1,3(4)-β-glucanase with a specific activity of 2680 and 340 U mg⁻¹ towards barley β-glucan and laminarin, respectively. It was specific for β-glucans containing β-1,3-linkages with an optimum temperature of 70 °C at pH 6.0. The N-terminal SLH modules were cleaved from the protein as well in Escherichia coli as in C. thermocellum, but nevertheless bound tightly to the rest of the protein. Lic16A was located on the cell surface from which it could be purified after fractionated solubilization. Its inducible production allowed C. thermocellum to grow on β-1,3- or β-1,3-1,4-glucan.
The three-dimensional structure of some members of GHF16 has been resolved (Hahn et al., 1995; Krah et al., 1998). The β-1,3-glucanases produce a characteristic pattern of β-1,3-oligosaccharides, which are further degraded by β-glucosidases. In contrast to the highly specific β-glucanases, β-glucosidases degrade laminarin-oligosaccharides (β-1,3) as well as cello-oligosaccharides (β-1,4) (Gräbnitz, et al. 1989; Zverlov et al., 1997b).

β-1,3-Glucanases have a commercial potential, for example for yeast extract production or the conversion of algal biomass to fermentable sugars. They also have, in combination with chitinases, antimiycotic activity for disease protection of plants (Nogi & Horikoshi, 1990). Partial hydrolysis of β-1,3-glucans can lead to biologically active products, for example to therapeutic biological response modifiers with antiviral, antitumour or related activity (Hagiwara & Kikuchi, 1994; Misaki & Kakuta, 1997).

C. thermocellum produces a highly active β-1,3-glucanase. It is a moderately thermophilic anaerobic Gram-positive bacterium which is common in soil, self-heated rotting biomass or hot springs (Bender et al., 1985; Chuvilskaya et al., 1987; Stainthorpe & Williams, 1988). The thermostability of its hydrolytic exo-enzymes promises considerable biotechnological application potential. Different strains of this species received increasing attention for their ability to ferment ligno-cellulosic biomass to ethanol and organic acids in a one-step process, a trait rare among bacteria (Tailliez et al., 1989; Lynd et al., 2002). It produces a sophisticatedly arranged enzyme complex for the degradation of the resilient crystalline cellulose, called the cellulosome (Bayer et al., 2000; Schwarz, 2001). This is located extracellularly on the cell surface. Most polysaccharide hydrolases of C. thermocellum are known to reside in the cellulosome, which contains cellulases and hemicellulases. However, some exo-enzymes are not bound to this multi-enzyme complex. Examples are the endoglucanases CelC, Cell, CelM or the xylanase XynX. These proteins may be part of a second, a ‘soluble’ enzyme system for the hydrolysis of polysaccharides.

Surprisingly C. thermocellum does not readily grow on glucose, the building block of cellulose. However, β-1,4-oligomers and polymers like cellobiose, celloextrins and cellulose (but not α-1,4-glucans) are used by virtue of the energy-saving phosphorolytic cleavage of celloextrins (Lynd et al., 2002). Other β-glucans, e.g. β-1,3-glucans, have not been tested for growth so far. Pentose sugars or their polymers are not fermented at all, even though hydrolysis of xylan takes place (Nochur et al., 1990; Hernández, 1982).

Genomic libraries have previously been constructed from two C. thermocellum strains, DSM1237T and F7 (Schwarz et al., 1985; Tuka et al., 1990). Clones expressed in Escherichia coli were extensively screened for enzymic activities related to β-1,3-glucan (laminarin) and mixed-linkage glucan (lichenan) hydrolysis. In this paper we describe the primary structure of the gene licA, its product Lic16A and the flanking genomic area in both strains. We discuss a possible evolutionary mechanism of the extraordinarily complex β-glucanase Lic16A, the relation of the β-1,3- to the β-1,3,1,4-glucanases and the localization of the protein. We show growth of C. thermocellum on β-1,3-glucan and present a characterization of the purified recombinant and the native β-1,3-glucanase (laminarinase), which is thermostable and highly active.

METHODS

Strains and growth conditions. E. coli XL-1 Blue was used for cloning and expression of clones. C. thermocellum DSM1237T (corresponds to ATCC 27405) was obtained from the German Collection of Micro-organisms (DSMZ) and strain F7 from the All-Russian Collection of Micro-organisms (VKMB 2203). Recombinant E. coli was grown at 37°C in Luria–Bertani broth (LB) containing ampicillin (100 µg ml⁻¹). Schizosaccharomyces pombe (DSM2791) was grown in LB. Bacillus subtilis MW10, which is deleted for the β-glucanase gene, was obtained from R. Borriss and grown in LB (Wolf et al., 1995). dLB medium was LB containing 20 g glucose l⁻¹. C. thermocellum was grown in rubber-stoppered glass bottles with pre-reduced, anaerobic GS-2 medium containing 1% (w/v) cellobiose or other carbohydrates at 60°C (Johnson et al., 1981).

Recombinant DNA techniques. Clones pLIC7 and pCU309, containing the laminarinase gene of strains DSM1237T and F7, respectively, have been described previously (Schwarz, et al. 1985; Tuka et al., 1990). DNA preparation, cloning, restriction-endonuclease digestion, hybridization and PCR were carried out by standard procedures (Sambrook et al., 1989). The Expand-High-Fidelity PCR-System was used according to the recommendations of the manufacturer (Roche Diagnostics). The intact LicA gene (pKL112) was amplified with primers pLICA1 (5'--aaccggct ataacctccg attccaacc gagg-3') and pLICA4 (5'--cgcttacaata ctaactgct cgctc-3'), and the DNA of the Lic16A(Cat) module with pLICA7 (5'--gcgctgac taactctca ccaacaaga atccg-3') and pLIC8 (5'--aacttacct tagctgctt ctctgctgg ggt-3'). The resulting DNA fragments were cloned with the expression vector pQE32 (Qiagen). To amplify a 420 bp fragment of Orf1 with the PCR DIG Labelling MixªLCS (Roche Diagnostics) the oligonucleotide primers pTra1 (5'-aagaa- cagtt aatggaaat ggg-3') and pTra2 (5'-tactctat ataacttagct gtc-3') were used.

DNA sequences were determined from supercoiled double-stranded plasmid DNA for both strands by using the Thermosequenase Cycle Sequencing Kit (Amersham Biosciences) for extension of 5'-biotinylated oligonucleotide primers. DNA fragments were detected with the GATC 1500 Direct-Blotting-Electrophoresis apparatus using streptavidin-conjugated alkaline phosphatase and NBT-BCIP (nitro blue tetrazolium-5-bromo-4-chloro-3-indolyolphosphate) as chromogenic substrate (Roche Diagnostics). Sequence data were analysed, edited and compared with the DNASIS/PROSIS for Windows package (Hitachi Software Engineering). Nucleotide and protein sequence databases were screened using FASTA and BLAST software at the NCBI server (http://www.ncbi.nlm.nih.gov).

Construction of the ileX-argU plasmid. The E. coli ileX gene, obtained from H. Inokuchi (Komine et al., 1990) and encoding tRNA⁵₅leu, was cloned, including its own promoter, by ligating the SpI₂ fragment into the SpI₂ site of plasmid pSB161 (courtesy S. Baumann & R. Matthes; Schenk et al., 1995) to yield plasmid pKL162. This plasmid thus contains the strongly expressed argU and
icleX genes from E. coli, and the kanamycin resistance and p15A replication origin from pACYC177 (Chang & Cohen, 1978) for compatibility with the pBR322-derived expression system in E. coli.

Purification of the enzyme. E. coli pKL112 (pQE32::licA)/pKL162 cells from an IPTG-induced 3 l culture in dLB with ampicillin (100 mg l⁻¹) and kanamycin (25 mg l⁻¹) were collected by centrifugation, washed twice in 20 mM Tris/HCl, pH 8-0, suspended in the presence of Benzonase and Pefabloc protease inhibitor (both Merck) and disintegrated by two to six passages through a French pressure cell (AmInCo) at 110 MPa with ice cooling. Completeness of disintegration was controlled microscopically. Cell fragments were removed by centrifugation (Sorvall SS34, 18,000 r.p.m., 30 min). The clarified cell extract was heated to 55°C for 30 min, centrifuged and filtered through a 0-2 μm membrane filter. The filtrate was applied to a Sepharose FF anion-exchange column, equilibrated with the same buffer, and eluted with an increasing NaCl gradient (0–1000 mM). Protein fractions having β-glucanase activity were applied to a hydrophobic interaction chromatography column Phenyl-Sepharose HP 16/10 (Amersham Biosciences) in 1-2 M (NH₄)₂SO₄ and eluted with a decreasing gradient of (NH₄)₂SO₄ in 20 mM Tris/HCl, pH 8-0. The active fractions were applied in 100 mM NaCl, 20 mM Tris/HCl, pH 8-0 to a gel filtration on a Superdex 200 prep grade XK 16/60 column (Amersham Biosciences).

Western blotting and protein sequencing. SDS-PAGE slabs were blotted on PVDF membranes (Roche Diagnostics). Monoclonal anti-His-IgG (Qiagen) or polyclonal rabbit antibodies elicited against purified Lic16A(Cat) were used for detection. N-terminal sequences were determined from protein eluted from cut-out pieces of SDS gel electrophoresis slabs by Edman degradation using a Procise 492 protein sequencer (Applied Biosystems). The phenylthiohydantoin derivatives were identified by reversed-phase HPLC. Endoproteinase LysC (Roche Diagnostics) was used to digest degradation-resistant native Lic16A. The peptide mixture was separated by reversed-phase HPLC.

Zymogram technique. Proteins were separated by denaturing SDS-PAGE (10% polyacrylamide) and renatured by three successive washings in 50 mM phosphate-chlorite buffer (pH 6-2) with and without i-propanol as described previously (Schwarz et al., 1987). The gel slabs were overlaid with agarose gel containing polymeric substrates (1%, w/v), incubated at 55°C and stained with 0. The gel slabs were overlaid with agarose gel containing polymeric substrates (1%, w/v), incubated at 55°C and stained with 0. The gel slabs were overlaid with agarose gel containing polymeric substrates (1%, w/v), incubated at 55°C and stained with 0. The gel slabs were overlaid with agarose gel containing polymeric substrates (1%, w/v), incubated at 55°C and stained with 0.

Enzymic assays and detection of hydrolysis products. Enzyme aliquots in standard assays were incubated in succinate buffer (100 mM, pH 5.0) with 1% (w/v) substrates at 60°C. Reducing sugars released from polymeric substrates were detected by the 3,5-dinitrosalicylic acid (DNSA) method (Wood & Bhat, 1988). Insoluble substrates were precipitated by centrifugation and the supernatant with DNSA.

Mono- and oligosaccharides were separated by TLC on 0.2 mm silica gel 60 plates (Merck) in multiple runs in acetonitrile/water (80:20, v/v). Sugars were detected by spraying with a solution of 1 g diphenylamine and 1 ml aniline in 100 ml acetic acid plus freshly added 10 ml orthophosphoric acid. Colour was developed on heating at 120°C.

Preparation of yeast cell walls. For isolation of Saccharomyces cerevisiae cell walls, commercial bakers yeast was used. Sch. pombe cells were cultured and collected by centrifugation. Cells were washed by centrifugation, homogenized and disrupted by eight to twelve passages through a French Press (AmInCo). The cell walls were sedimented by centrifugation, washed five times with pure water and lyophilized. No reducing sugars could be detected in the supernatant with DNA.

Substrates. Oat spelt xylan, carboxymethyl-cellulose, lichenan, CMC (low viscosity), laminarin, soluble starch and chitin were obtained from Sigma-Aldrich; barley β-glucan, pachymann and CM-pachymann from Megazyme; pustulan from Roth; curdlan from Wako; soluble starch from Merck; and Avicel CF1 from Serva. Laminaridextrins as size standards for HPLC and TLC were obtained from Seikagaku. Phosphoric acid swollen cellulose was prepared from Avicel CF1 by the method of Wood & Bhat (1988).

RESULTS

The licA region in the genome of two C. thermocellum strains

Cosmid clone χLIC7 expressing β-1,3-glucanase activity has been isolated previously from a genomic library of the C. thermocellum type strain DSM1237T (Schwarz et al., 1985, 1988a). The nucleotide sequence of a subcloned DNA fragment was determined (5595 bp; accession no. X89732). It overlapped the previously obtained sequence of another subclone that encoded the endoglucanase gene celC (accession no. M19422; Schwarz et al., 1988b). The connected sequences cover 6698 bp (Fig. 1). To verify the unusual structure of the genes we also determined the corresponding sequence of clone pCU309 from strain F7, another strain of C. thermocellum (9760 bp; accession no. AJ307315). The F7 sequence covered that from DSM1237T completely and extended beyond it at both ends (Fig. 1). The sequences were 99-9% identical in the overlapping area, with the majority of the mismatches in intergenic regions.

![Fig. 1. The CelC-Lic16A region in the genome of two C. thermocellum strains.](image-url)
The reading frames and the structure of the genes were equivalent and are described below for the longer F7 sequence.

Six ORFs were identified and all but Orf2 were preceded by potential Shine–Dalgarno sequences (Fig. 1). All frames were encoded on the same DNA strand. The corresponding protein sequences were almost completely conserved between the two strains, because most mismatches were in the third codon position and did not change the amino acid composition or the reading frames (as, for example, all mismatches in licA and celC). The celC gene region could be compared with a third sequence from strain F1, isolated in Japan (Sakka et al., 1991), which contained only four amino acid exchanges (of 343 aa residues) compared to the corresponding sequences of DSM12377 and F7.

**Putative transposase gene**

The first reading frame in the F7 sequence (Orf1, bp 257–712) had a high sequence identity with IS605/IS200 transposases of family 17 (Pfam accession no. PF01797; Bateman et al., 2002), e.g. 92 % similarity to the C-terminal part of the Clostridium acetobutylicum transposase (aa 58–151; accession no. AE007849). However, the alignment of amino acid residues 29–151 with different transposases, including that of Clostridium perfringens or Thermotoga maritima, was perfect if a gap of 10 aa residues between L57 and K58 was introduced. It should be mentioned that the overall size of the homologous bacterial transposase genes was identical to the transposase gene proposed here.

To verify the presence of IS605/200-homologous DNA sequences in the genome of C. thermocellum F7, an amplified 420 bp DNA fragment corresponding to the well conserved C-terminal region was hybridized to three complete endonuclease restriction enzyme digests of genomic DNA from strain F7. The multitude of resulting bands of hybridizing DNA sequences indicated the presence of a number of different insertion sites in the chromosome of strain F7 (Fig. 2). The intensity of the hybridizing bands may vary due to the occurrence of double and triple bands.

**Other reading frames**

Orf2 is dealt with in the Discussion section. Orf3 encodes the non-cellulosomal β-1,3-1,4-glucanase CelC (bp 2350–3381), which was described previously (Schwarz et al., 1988b). Orf4 shows up to 52 % similarity to bacterial transcription repressors of the GalR/LacI family over the whole length of the gene. Orf6 is truncated at the C terminus. It is up to 62 % similar to the N-terminal 322 aa residues of hypothetical membrane proteins, which consist of a leader peptide and eight putative transmembrane regions, e.g. from C. perfringens or Staphylococcus aureus.

**The modular structure of the licA gene**

Orf5 encodes laminarinase A and is called licA due to the high activity of the gene product on lichenan which it shares with LicB (Zverlov et al., 1994). A putative ribosome-binding site is located at base 4632 upstream of a coding sequence starting at ATG (base 4645) and leading to a large protein of 1323 aa (148 116 Da). licA is followed by a downstream 13 bp hairpin structure and a run of T residues, presumably a rho-factor independent transcription terminator. The translated amino acid sequence consists of a hydrophobic N terminus (25 aa) with the typical features of a signal peptide (von Heijne, 1986), three repeated S-layer homologous modules (SLH; aa 27–227), a module of no significant similarity to the databases which is connected via a short stretch of hydroxy-amino acids (PTS-box) to a catalytic module (aa 421–671), and followed by four consecutive carbohydrate-binding modules of family 4a (CBM4a) (Fig. 1b; Zverlov et al., 2001). The catalytic module aligns well with bacterial glycosyl hydrolase family 16 modules, which were found in β-1,3-glucanases and mixed-linkage glucanases of Bacillus, Thermotoga and other bacteria (Coutinho & Henrissat, 1999). Hence the gene product was named Lic16A and can be described by the structural formula LP/SLH/SLH/SLH/X/PT/GH16/CBM4a/CBM4a/CBM4a/CBM4a according to a recently proposed nomenclature (Henrissat et al., 1998).

**Heterologous expression and purification of Lic16A<sub>DSM1237</sub>**

The codon usage in gene licA is as biased as it is in other genes of C. thermocellum, especially for the arginine codons...
AGG and AGA and the isoleucine codon AUA, for which tRNAs in E. coli are scarce. A succession of these codons in a heterologous gene impedes overexpression in E. coli host cells (Del Tito et al., 1995; Zdanovsky & Zdanovskaia, 2000). Accordingly, the heterologous expression of licA in E. coli was low and IPTG induction resulted in severe growth retardation and heavy protein degradation. Expression of licA in Bacillus subtilis MW10 in the shuttle-vector pHPS9 resulted in similar difficulties (Haima et al., 1990; Wolf et al., 1995). In the presence of the helper plasmid pKL162, which supplemented the limiting IleX and ArgU tRNAs, the licA-mediated induction in E. coli was increased 10- to 12-fold and Lic16A protein was less degraded. Upon IPTG induction, the isolated catalytic module of Lic16A [LicA(Cat)] in plasmid pQE32 amounted to 50 % of the soluble cell proteins (data not shown).

Lic16ADSM1237 was purified from induced E. coli cells by liquid chromatography. In denaturing SDS-PAGE two major proteins forming bands of 120 and 29 kDa were detected which co-purified with the enzymic activity and could not be separated by gel filtration or affinity chromatography (data not shown, but similar to Fig. 3). The molecular mass calculated from calibrated gel filtration chromatography was about 130 kDa, identifying Lic16A as a monomeric enzyme. The 22 N-terminal amino acids of the 120 kDa protein band were determined by micro-sequencing (SIHFINTKIN RVVNVNTGVR IV) and verified that the 120 kDa protein was a fragment of Lic16A, truncated by its 266 N-terminal amino acids. Taking the processing site between aa residues 266 and 267, the calculated masses for the cleaved products would be 118-3 kDa for the C-terminal fragment (including the catalytic module) and 29-5 kDa for the N-terminal protein (including the leader peptide). This was corroborated by the experimental data (Fig. 3) and suggested a single, specific processing site, cleaving the complete 148 kDa protein of Lic16A, truncated by its 266 N-terminal amino acids. Taking the N terminus was split off the catalytic module, an N-terminal hexahistidyl tag added by cloning into vector pQE32 in the place of the leader peptide could help to purify the complete Lic16A protein of 148 kDa by metal-chelating affinity chromatography. But again the two bands of 120 and 29 kDa were obtained in equimolar amounts in denaturing SDS-PAGE as described above (Fig. 3). Only the 120 kDa band, but not the 29 kDa band, was active in zymograms with barley β-glucan and thus contained the catalytic module, whereas only the smaller polypeptide reacted with anti-His tag antibodies, identifying it as the N terminus of Lic16A. A band of the complete 148 kDa protein could not even be detected in the highly sensitive Western blot with anti-Lic16A(Cat)-antibodies. Despite the occurrence of two bands in denaturing SDS-PAGE we have purified a homogeneous protein: the protein obviously was quantitatively cleaved but the two fragments of Lic16A bound tightly to each other.

If the N terminus was split off the catalytic module, an N-terminal hexahistidyl tag added by cloning into vector pQE32 in the place of the leader peptide could help to purify the complete Lic16A protein of 148 kDa by metal-chelating affinity chromatography. But again the two bands of 120 and 29 kDa were obtained in equimolar amounts in denaturing SDS-PAGE as described above (Fig. 3). Only the 120 kDa band, but not the 29 kDa band, was active in zymograms with barley β-glucan and thus contained the catalytic module, whereas only the smaller polypeptide reacted with anti-His tag antibodies, identifying it as the N terminus of Lic16A. A band of the complete 148 kDa protein could not even be detected in the highly sensitive Western blot with anti-Lic16A(Cat)-antibodies. Despite the occurrence of two bands in denaturing SDS-PAGE we have purified a homogeneous protein: the protein obviously was quantitatively cleaved but the two fragments of Lic16A bound tightly to each other.

Biochemical characterization of Lic16A

Purified recombinant Lic16A was capable of degrading barley β-1,3-1,4-glucan and with a lower reaction velocity, lichenan and laminarin (Table 1). Activity on pure β-1,3-glucans, e.g. pachyman, CM-pachyman and insoluble curdlan, was present but low. Activity towards Sch. pombe cell walls which also contain β-1,3-glucan, was obtained, but surprisingly not with Sac. cerevisiae cell walls. β-1,4-Glucans (like crystalline cellulose, acid swollen cellulose or CMC), other β-1,4-glycans (like xylan or chitin) or p-nitrophenyl-β-D-glycosides were not hydrolysed. This indicated that the enzyme catalysed the hydrolysis of β-1,3-glucosidic linkages and is a β-1,3-glucanase (laminarinase). A Vmax of 3175 and 417 U mg⁻¹, a kcat of 7474 and 836 s⁻¹, a Km of 0-94 and 2-11 mg ml⁻¹ and an apparent kcat Km⁻¹ of 7951 and 396 ml s⁻¹ mg⁻¹ were determined with the substrates barley β-glucan and laminarin, respectively, at 60°C and pH 6-0.

---

**Fig. 3.** Recombinant Lic16A protein purified with His-tag affinity chromatography, denaturing SDS-PAGE and Western blotting. Lanes: 1, purified recombinant Lic16A stained with Coomassie blue (bands of 120 and 29 kDa); 2 and 3, Western blot with anti-Lic16A(Cat) (only 120 kDa; lane 2) and anti-His-tag antibodies (only 29 kDa; lane 3). The size of marker proteins is indicated.

http://mic.sgmjournals.org
The specific activity of supernatants and cell extracts of all cultures was determined on β-1,3,glucan and barley β-1,3,1,4-glucan. Zymogram investigations had shown that mixed-linkage β-glucan is hydrolysed by all endoglucanases, whereas a comparison with Western blots revealed that Lic16A is the prevalent if not the only β-1,3-glucanase in C. thermocellum (see below). The ratio of the specific activities on laminarin and barley β-glucan can therefore be regarded as an indication for the specific expression of Lic16A: whereas this ratio was 0·6 and 11·3 % in the cell extracts of Avicel- and cellobiose-grown cells, it was significantly higher in laminarin- and β-1,3,1,4-glucan-grown cultures (26·1 and 21·4 %), respectively, (Table 2). This indicates a specific induction of Lic16A expression by β-glucans containing β-1,3-linkages. This result was supported by Western blot experiments with antibodies prepared against purified Lic16A(Cat) where the amount of reacting protein increased upon growth on β-1,3,1,4-glucan and laminarin (Fig. 4). Fig. 4 also reveals that the larger band of native Lic16A had a molecular mass of 125 kDa, slightly higher than in recombinant E. coli. No 148 kDa protein was detected with Lic16A(Cat)-antibodies, neither in culture supernatants nor in cell extracts, indicating complete processing not only in E. coli but also within the C. thermocellum cell (prior to secretion), although probably at a different site. Bands with a smaller mass (92, 80 and 61 kDa) may be the result of further processing and roughly fit to a gradual release of the C-terminal CBM4α modules.

To investigate the localization of mature Lic16A and to purify the protein from the native C. thermocellum DSM1237\(^{T}\), cultures were raised on CMC which showed an eight times higher induction of β-1,3-glucanase activity than cellobiose, but was much more affordable than within 4 days to an optically dense culture on cellobiose, Avicel, CMC, barley β-glucan and laminarin, but not (within 10 days) on xylan, insoluble curdlan, pustulan, yeast cell walls, chitin and starch. The cultures reached the highest density on laminarin and barley β-glucan: the protein yield of cell extracts was 50–100 % higher than on Avicel and even higher than on cellobiose (Table 2).

Table 2. Enzymic activities in cell extracts of C. thermocellum cultures grown on different carbohydrate sources

<table>
<thead>
<tr>
<th>Carbohydrate source</th>
<th>Protein (mg ml(^{-1}))</th>
<th>β-Glucan (U ml(^{-1}))</th>
<th>Laminarin (U ml(^{-1}))</th>
<th>Lam/ Gluc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avicel (β-1,4)</td>
<td>3-63</td>
<td>122</td>
<td>0-75</td>
<td>0-6</td>
</tr>
<tr>
<td>Cellobiose (β-1,4)</td>
<td>5-15</td>
<td>13-3</td>
<td>1-9</td>
<td>11-3</td>
</tr>
<tr>
<td>CMC (β-1,4)</td>
<td>3-14</td>
<td>12-8</td>
<td>1-7</td>
<td>12-9</td>
</tr>
<tr>
<td>Barley β-glucan (β-1,3,1,4-glucan)</td>
<td>7-69</td>
<td>43-3</td>
<td>9-3</td>
<td>21-4</td>
</tr>
<tr>
<td>Laminarin [β-1,3,(1,6)-glucan]</td>
<td>5-97</td>
<td>29-8</td>
<td>7-8</td>
<td>28-1</td>
</tr>
</tbody>
</table>
During growth, cell-bound laminarinase activity increased proportionally with the optical density of the culture (OD600) and the cell protein content, and thus was not growth-phase-dependent (data not shown).

After 4 days of growth the cells were collected and washed differentially: 150 mM NaCl or distilled water released only 7% of the laminarinase activity, whereas 2 M guanidinium hydrochloride (GuHCl) solubilized Lic16A completely, but only 10% of the cell protein. No more laminarinase was washed off the cells by 5 M GuHCl and the cells did not retain β-1,3-glucanase activity. A control experiment showed that laminarinase activity recovered at least partially after 5 M GuHCl treatment and would have been detected if present. The 5 M GuHCl wash contained a dominant 130 kDa protein band, presumably the major S-layer protein (Lemaire et al., 1995). Western blots with anti-Lic16A(Cat)-antibodies and zymograms with barley β-glucan corroborated that Lic16A could be solubilized selectively from the cell surface by 2 M GuHCl treatment and contained an enzymically active protein of 125 kDa. The comparison of Lic16A-Western blots and β-1,3-glucan zymograms revealed that Lic16A was the major if not the only β-1,3-glucanase present.

Purification of Lic16A from C. thermocellum and interaction with cellular proteins

The native Lic16A protein was purified from the 2 M GuHCl fraction by liquid chromatography as described for the recombinant Lic16A protein. A protein active on xylan was separated from Lic16A by anion-exchange chromatography (MonoQ; Amersham Biosciences). It was probably xylanase XynX, which is bound to the cell wall via its terminal SLH modules (Lemaire et al., 1995), similar to Lic16A. A hydrophobic interaction chromatography step finally resulted in a homogeneous native Lic16A protein. The amount of Lic16A represented 0·23% of the total cell-bound protein if purification losses were taken into account. Since the N terminus was blocked for Edman degradation, the identity of the purified protein with Lic16A was verified by sequencing the N terminus of an internal LysC-protease fragment: the amino acid residues KDGALVIEARKEDI were identified, corresponding to amino acid residues 468–481.

The SLH modules of extracellular proteins were previously shown to interact with cellular components (Lemaire et al., 1995). To investigate binding of Lic16A to specific proteins, C. thermocellum extracts were separated in denaturing SDS-PAGE, blotted on PVDF membranes and incubated in blocking buffer for 1 h with purified LicA protein. With the complete Lic16A, but not with Lic16A(Cat) which is devoid of SLH modules, three bands with a molecular mass of 110, 85 and 61 kDa were detected with anti-His IgG (Fig. 5). Faint bands which could be the result of unspecific binding were neglected. The major bands represent three proteins...
of the cell to which SLH modules of Lic16A are specifically binding.

**DISCUSSION**

The chromosomal region around licA and celC was sequenced in two different strains of *C. thermocellum*, the type strain DSM1237T (corresponds to ATCC 27405 and NCIB 10682), isolated in the USA, and strain F7, isolated in Armenia (former Soviet Union; Chuvilskaya et al., 1987). Despite the isolation from different continents, nucleotide sequences were surprisingly similar and the amino acid sequences of LicA and CelC were identical. This corroborates data obtained with other pairs of genes: LicB and ChiA in DSM1237T and F7 (Schimming et al., 1992; Zverlov et al., 2002), two xylanase genes in strains ATCC 27405 and F1 (>99-3%), cellobiose phosphorylase Cbp in strains YM4 and F1 (99-6%) and CelK in strains F7 and YW20 (99-8%), which contain only a few differences. Thus the *C. thermocellum* strains isolated in the Northern hemisphere seem to be quite homogeneous, as was also suggested from the physiological characteristics of different isolates (McBee, 1954).

Two glycosyl hydrolase genes, celC and licA, are located on the same strand of the sequenced DNA fragments. Both genes encode proteins involved in non-cellulosomal β-glucan degradation. According to the arrangement of chromosomal DNA fragments and the restriction map obtained by DNA–DNA hybridization, plasmids pKS305 and pCU309 are located within a close distance (Guglielmi & Beguin, 1998). pKS305 contains the genes manB and celT (Kurokawa et al., 2002). It should be noted that according to a comparison of the restriction map given in Guglielmi & Beguin (1998) and the DNA sequences, the cellulosomal genes manB and celT are located downstream of orf6 (Fig. 1) and transcribed from the same DNA strand. A cluster of four genes involved in cellulose degradation is therefore located in this chromosomal area of *C. thermocellum*.

The two proteins CelC and LicA differ greatly in structure: Lic16A has a complex structure of one catalytic and eight non-catalytic modules, whereas CelC contains only a single (catalytic) module, like the majority of the extracellular hydrolases of the aerobic Gram-positive bacteria, e.g. the *Bacillus* amylases. The high complexity of the Lic16A structure is comparable to the most complex cellulolytic proteins from the non-cellulosomal cellulase system of *Caldicellulosiruptor* sp. or its relative *Anaerocellum thermophilum* (Gibbs et al., 2000; Zverlov et al., 1998). A comparison between the two strains and partial resequencing of sequence differences between the genes made the correctness of the unusual structure of *licA* reliable beyond reasonable doubt.

It can be speculated that the arrangement of the completely different hydrolase structures in a narrow region of the *C. thermocellum* genome, CelC and LicA, is the result of the evolution of new catabolic genes by genetic recombination events, by which new catabolic pathways are assembled (Mahillon et al., 1999). Hence, it is not surprising that a transposase gene [at least a precursor or remnant of an insertion element (orf1)] was found to be located upstream of licA and celcin the genome of *C. thermocellum* F7. Similar arrangements of cellulase gene clusters and a transposase gene can be observed in the chromosomes of *Clostridium cellulolyticum*, *Clostridium cellulovorans* and *C. acetobutylicum* (Lynd et al., 2002). At present there is no experimental evidence that the multiple copies of the IS605/200-type insertion elements are actively involved in the evolution of *C. thermocellum* or if orf1 is a IS200-transposase variant with an unusual N-terminal sequence. However, it should be mentioned that the DNA region immediately downstream of the putative transposase gene shows homology to the C-terminal part of the cell wall lysis gene *cwU* of *Paenibacillus polymyxa* (orf2 in Fig. 1a), beginning with amino acid residue 68 (bp 954). The homology region begins with a typical hexameric sequence (5′-TTTAA-3′) which was reported to be the duplicated end of the IS605/200 element in *Helicobacter pylori* (Kersulyte et al., 1998). This suggests that an IS605/200 element (orf1) may have inserted in a *cwU*-homologous gene in *C. thermocellum* and could have contributed to the module shuffling, resulting in the complex structure of nearby licA.

The licA gene was introduced into *E. coli*. Obviously, the initially low expression level was due to the occurrence of multiple codons for rare host aminoacyl-tRNAs clustered in the 5′ region of the gene upstream of the catalytic module. This problem was overcome by introducing a helper plasmid containing the underrepresented tRNA genes. A similar improvement upon introduction of the same helper plasmid was found in expressing the *C. thermocellum* genes *licB* or *chiA*, and the *Clostridium stercorarium* gene celY, and facilitated an efficient heterologous production with a high yield of unprocessed protein (unpublished observations).

In its biochemical characteristics Lic16A is a typical endo-β-1,3-glucanase of glycosyl hydrolase family 16 (EC 3.2.1.39) with an extraordinarily high activity on soluble mixed-linkage glucans and an affinity to barley β-glucan three times higher than to laminarin. The sequence of Lic16A has overall homology to the thermostable *R. marinus* β-1,3-glucanase BglA (42-8% identity in a 257 aa overlap) rather than to the *Bacillus*-type mixed-linkage β-glucanases, which cleave almost exclusively β-1,4-glucosyl linkages on 3-O-substituted gluco-pyranose units (Anderson & Stone, 1975; Henrissat & Bairoch, 1996; Spilliaert et al., 1994). The difference is reflected in its active site sequence which contains four additional amino acid residues (Fig. 6), as was found with the β-1,3-glucanases of *B. circulans* KCTC3004, *R. marinus* (LamR) or *Thermotoga neapolitana* (LamA) which have a similar substrate specificity and pattern of products (Yahata et al., 1990; Spilliaert et al., 1994; Aono et al., 1995; Lee & Chang, 1995; Zverlov et al., 1997a; Krah et al., 1998).
was not growth-phase dependent. Whether the putative b
The expression of extra-cellulosomal Lic16A was induced whether studying.
worthwhile studying.
connected to it. The mechanism of this phenomenon is worthwhile studying.
the putative b
The mechanism of this phenomenon is worthwhile studying.

His-tag is fused to the SLH module, the catalytic part of the putative b
His-tag is fused to the SLH module, the catalytic part of worthwhile studying.

The N terminus of Lic16A, which contains the SLH worthwhile studying.
The SLH modules of bacterial exoenzymes, involved in worthwhile studying.

The sequence of Lic16A is shown in the bottom line. The worthwhile studying.

The sequence of Lic16A is shown in the bottom line. The worthwhile studying.

The sequence of Lic16A is shown in the bottom line. The worthwhile studying.

The N terminus of Lic16A, which contains the SLH worthwhile studying.
The N terminus of Lic16A, which contains the SLH worthwhile studying.

The N terminus of Lic16A, which contains the SLH worthwhile studying.

The N terminus of Lic16A, which contains the SLH worthwhile studying.

The N terminus of Lic16A, which contains the SLH worthwhile studying.

The N terminus of Lic16A, which contains the SLH worthwhile studying.

The expression of extra-cellulosomal Lic16A was induced worthwhile studying.

C. thermocellum was described as a cellulase-hydrolysing worthwhile studying.
bacterium, which grows on cellulose and celldextrins, but worthwhile studying.

C. thermocellum was described as a cellulase-hydrolysing worthwhile studying.
bacterium, which grows on cellulose and celldextrins, but worthwhile studying.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Deutsche worthwhile studying.

REFERENCES


Fig. 6. Consensus site of glycosyl hydrolase family 16 proteins: worthwhile studying.
consensus sequences of the two subfamilies. Above: the worthwhile studying.
sequence of mixed-linkage glucanases of the ‘Bacillus’ type worthwhile studying.
(β-1,3-1,4-glucanase); below: sequence of the β-1,3-glucanases.

The SLH modules of bacterial exoenzymes, involved in worthwhile studying.
intermolecular interaction and in binding to cell wall worthwhile studying.
components, are usually located at the C terminus, whereas worthwhile studying.
in Lic16A and in the alkaline cellulase of Bacillus KSM-635 worthwhile studying.
they are found N-terminally (Ozaki et al., 1990).Lic16A worthwhile studying.
indeed interacted with a number of cell proteins of a worthwhile studying.
similar size as those described for OlpB in C. thermocellum worthwhile studying.
or EngB in C. cellulovorans, which also contain SLH worthwhile studying.
modules (Lemaire et al., 1995; Kosugi et al., 2002). However, worthwhile studying.
it did not interact with the 130 kDa major S-layer component. The worthwhile studying.
interacting proteins were not identified, but worthwhile studying.

The N terminus of Lic16A, which contains the SLH worthwhile studying.
modules, was cleaved off upon expression in E. coli. The worthwhile studying.
cleavage was quantitative and site-specific. A similar worthwhile studying.
cleavage mechanism can reasonably be surmised for the worthwhile studying.

The expression of extra-cellulosomal Lic16A was induced worthwhile studying.
by glucans containing β-1,3-linkages and its production worthwhile studying.
was not growth-phase dependent. Whether the putative worthwhile studying.

The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community, 3rd


Schimming, S., Schwarz, W. H. & Staudenbauer, W. L. (1992). Structure of the Clostridium thermocellum gene licB and the encoded β-1,3-1,4-glucanase. A catalytic region homologous to Bacillus...
lichenases joined to the reiterated domain of clostridial cellulases. 


