A newly described cellulosomal cellbiohydrolase, CelO, from Clostridium thermocellum: investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose

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

Author for correspondence: Wolfgang H. Schwarz. Tel: +49 8161 71 5445. Fax: +49 8161 71 5475.
e-mail: schwarz@mikro.biologie.tu-muenchen.de

1 Institute of Molecular Genetics, Russian Academy of Science, Kurchatov Sq., 123182 Moscow, Russia
2 Research Group Microbial Biotechnology, Technische Universitaet Muenchen, Am Hochanger 4, D-85350 Freising-Weihenstephan, Germany

The sequence of the celO gene from Clostridium thermocellum F7 was determined. The gene product, cellulase CelO (Ct-Cel5F), had a modular structure consisting of a carbohydrate-binding module of the CBM3 family and a catalytic domain of the glycosyl hydrolase family 5. The presence of the dockerin module indicated that the enzyme was a component of the cellulosome complex. The thermostable recombinant gene product was active on cellodextrins, barley β-glucan, carboxymethylcellulose and insoluble cellulose. Cellobiose was the only product released from amorphic and crystalline cellulose, cellotetraose and higher cello-oligosaccharides, identifying CelO as a cellbiohydrolase. The cleavage pattern of p-nitrophenyl β-D-cellotetraoside, blockage of the hydrolysis of NaBH4-reduced cellopentaose and the reduction in substrate viscosity suggested activity from the reducing end in a processive mode after making random cuts. Binding to insoluble, i.e. amorphous, and crystalline cellulose was mediated by the carbohydrate-binding module CBM3b, with a preference for the crystalline substrate.

Keywords: cellulase CelO, cellulosome, carbohydrate-binding module CBM3, reducing end

INTRODUCTION

The enzymic hydrolysis of cellulose is a complex process involving a set of enzymes. These enzymes have in common a specificity for 1,4-β-glucans, but they differ in their modes of action. Only the synchronous cooperation of enzymes with different modes of activity results in the complete decomposition of the recalcitrant, crystalline substrate. To hydrolyse insoluble, especially crystalline substrates, cellulytic anaerobic bacteria possess a structurally coherent complex of enzymes arranged around a non-catalytic scaffolding protein called scaffoldin. The term ‘cellulosome’ has been coined for this extracellularly located multiprotein complex (Lamed et al., 1983). Scaffoldin contains a number of modules (the cohesins) to which the catalytic components bind with the aid of a dockerin module (Benguin et al., 1998). Cohesin–dockerin binding is highly specific, and the presence of a dockerin module sequence in a protein indicates its location within the cellulosome (Mechaly et al., 2000; Salamitou et al., 1994).

To hydrolyse crystalline substrates efficiently, cellulase systems depend on the simultaneous presence of non-processive endo-1,4-β-glucanases, which produce new ends at random within a polysaccharide chain, and processive cellulases (exo-1,4-β-glucanases), which remain attached to the substrate and split off cellobiose (cellbiohydrolase) or multimers of cellobiose (processive endocellulase) from such ends (Reverbel-Leroy et al., 1997; Teeri, 1997). The catalytic domains of the cellulases have been assigned on the basis of sequence comparisons and hydrophobic cluster analysis to the glycosyl hydrolase families (GHFs) (Henrissat et al., 1998; CAZy server: http://afmb.cnrs-mrs.fr/~pedro/}

Abbreviations: CBM, carbohydrate-binding module; CMC, carboxymethylcellulose; GHF, glycosyl hydrolase family; PASC, phosphoric acid swollen cellulose; pNP, p-nitrophenyl.
The GenBank accession number for the sequence determined in this work is AJ275975.
CAZY). Four of these, GHF 6, 7, 9 and 48, contain cellobiohydrolases.

The endo- or exo-mode of an enzyme is determined by differences in the shape of the active-site pocket. If the hydrolytic amino acids in the active site can be freely accessed by the polysaccharide strand, i.e. if they are situated in an open, pocket-shaped cleft on the enzyme surface, the enzyme is endo-gluco-lytically active in a non-processive way. However, if the active site is blocked by a bulky extension of the protein, creating a tunnel-like structure covering the catalytic amino acids, the polymeric substrate can access the active site with only one of its ends and the enzyme acts processively by sliding along the substrate (Barr et al., 1996). The direction of this processivity, i.e. the enzyme activity on the reducing or the non-reducing end of the substrate, seems to be defined by the way that the substrate is bound and released (Parsiegla et al., 2000).

The processive activity of cellobiohydrolases on modified cellulose, such as carboxymethylcellulose (CMC), is blocked, or at least strongly inhibited, by the occurrence of derivatized glucose residues. This leads to severe substrate limitation. Cellobiohydrolases acting from the non-reducing end of the substrate can be assayed with chromogenic arylcellobiosides, from which the chromogenic aryl residue, attached to the reducing end of the substrate, is liberated. However, more recently, processive enzymes acting from the reducing end of the carbohydrate chain have been identified, for which chromogenic substrates are not yet commercially available (Boisset et al., 1998). Although only a few members of this enzyme type have been characterized, models for the hydrolysis of crystalline cellulose predict that the presence of processive cellulases with both types of chain-end-specificity is essential for productive and complete degradation of the substrate (Barr et al., 1996; Schwarz, 2001).

Besides the presence of different catalytic modules, the activity of cellulases on crystalline cellulose also depends on the presence of non-catalytic modules which contribute to the tight binding of the substrate (Tomme et al., 1998). These carbohydrate-binding modules (CBMs) have a stimulating effect by increasing enzyme-substrate proximity (Bolam et al., 1998); they help the enzyme to overcome the liquid-solid interface (Nutt et al., 1998) and also enhance the accessibility of the substrate surface by modifying the structure of the cellulose crystal (Páges et al., 1997). On the basis of sequence comparison, 26 families of substrate-binding modules from bacteria, with between 40 and 180 amino acids, have been defined (Coutinho & Henrissat, http://afmb.cnrs-mrs.fr/~pedro/DB/db.html; Tomme et al., 1998).

In this study we report the structure of the celO gene and the corresponding cellulase, CelO, from Clostridium thermocellum F7. CelO was shown to consist of a leader peptide, a substrate-binding domain CBM3b (which bound to cellulose), a short proline-rich linker, a catalytic domain of GHF5 and a dockerin domain. We present evidence which suggests that the enzyme is a cellobiohydrolase which hydrolyses cellodextrins from the reducing end.

METHODS

Bacterial strains and plasmids. Clone pCU107, from C. thermocellum F7 containing the celO gene, has been described by Bumazkin et al. (1999). Escherichia coli XL-1 Blue was used for cloning, and E. coli M15 containing plasmid pREP-4 was used for the overexpression of cloned genes (Qiagen). Recombinant cells were cultivated and overexpressed according to the manufacturer’s recommendations. Plasmids pQE30 and pQE31 (Qiagen) were used for the cloning of PCR products.

Molecular biological methods. The DNA sequence of clone pCU107 was determined from supercoiled, double-stranded plasmid DNA from both strands (Thermosteque Sequence Cycle Sequencing Kit, Amersham) with 5’-biotinylated oligonucleotide primers. The DNA sequence was determined with a GATC 1500 Direct-Blotting Electrophoresis sequencer (GATC, Konstanz) using streptavidin-conjugated alkaline phosphatase and the chromogenic substrate NBT-BCIP (Promega). Sequence data were analysed with the DNA/PROSIS software package (Hitachi Software Engineering). Nucleotide and protein sequence databases were screened using FASTA and BLAST (http://www.ebi.ac.uk).

The clustalw program (http://www.ebi.ac.uk/clustalw) was applied for progressive multiple sequence alignment by calculating maximum-parsimony, maximum-likelihood and position-specific gap-penalties (Thompson et al., 1994), and by using significant sequence similarity based on the amino acid substitution matrix of Henikoff & Henikoff (1992).

PCR was carried out using chromosomal DNA as a template and the Expand High Fidelity PCR System (Boehringer Mannheim). The PCR oligonucleotide primers were designed to truncate the gene for the leader peptide and the dockerin domain (rCelO, primers cel4-1-Bam and cel4-2-Sal), or the leader peptide, the CBM and the dockerin domain (rCelO-Cat, primers cel4-1-Bam and cel4-2-Sal). The primers were also designed to introduce recognition sites for the restriction endonucleases SalI (GTCGAC) or BamHI (GGATCC), allowing for cloning into the expression vectors pQE30 and pQE31, respectively. The primers used were: cel4-1-bam (3’-CCGTGATTTGGATCCTGCTTTTCAG-5’); cel4-2-bam (3’-AACCGACGGATCCTGGAAACAAG-5’); and, cel4-2-sal (3’-GTAAACCTCGAGATATTAAAGGTATCC-5’).

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

Enzyme assays. Enzyme aliquots for standard assays were incubated in sodium phosphate buffer (50 mM, pH 6.6) at 65 °C. The concentration of the substrates used was 1% (w/v) for soluble polysaccharides and 2% (w/v) for insoluble polysaccharides. Reducing sugars released from the 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 (mg protein)⁻¹. Specific activities were determined in the linear range of the reaction. The protein concentration was determined with Coomassie brilliant blue as described by Sedmak & Grossberg (1977). p-Nitrophenol liberation from p-nitrophenyl (pNP)-glycosides was measured by its absorbance in alkaline solution.
The equilibrium binding constant ($K_b$) was calculated as:

\[ K_b = \frac{[N_a] K_a [F]/(1 + K_a [F])} \]

where $[B]$ is the concentration of bound CBM (mol (g cellulose)$^{-1}$), $[F]$ is the concentration of free CBM (M), $[N_a]$ is the total concentration of available binding sites on the cellulose surface, also called the capacity value for cellulose (mol (g cellulose)$^{-1}$), and $K_a$ is the equilibrium association constant (M$^{-1}$). The concentration of bound protein was determined as the difference between the initial protein concentration and the free protein concentration after incubation with the substrate. The measurements were done at least in triplicate and the standard error was calculated.

**TLC.** Polymeric and oligomeric substrates were hydrolysed to completion under the conditions stated above. Hydrolysis products were separated on 0.2 mm aluminium sheet silica gel 60 plates (Merck) with acetonitrile/water (80:20, v/v) as eluent. Sugars were detected by spraying the plates with a freshly prepared mixture of 10 ml stock solution and 1 ml orthophosphoric 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.

**Viscometry.** A 1 % (w/v) solution of CMC in 0.1 M phosphate buffer of the appropriate pH, according to the optimal activity profile of the enzyme, was hydrolysed at 50 °C. The flow-time of the reaction mixture was determined at intervals in an Oswald viscometer at 50 °C, together with the reducing sugar content (3,5-dinitrosalicylic acid method). The relative fluidity was calculated as:

\[ \frac{[T_o/(T_o - T_p)] - [T_o/T_o]} \]

where $T_o$ is the flow-time measured for the buffer, $T_p$ the flow-time of the CMC solution without enzyme, and $T$ is the flow-time of the reaction mixture with the enzyme.

**Substrates.** Birchwood xylan, Avicel CFI, CMC (low viscosity), and pNP-glycosides were obtained from Sigma, cellobextrins from Merck, barley β-glucan from Megazym, and pustulan from Roth. Phosphoric acid swollen cellulose (PASC) was prepared from Avicel CFI by the method of Wood (1988). NaBH$_4$-reduced cellobextrins were prepared from cellobiose, cellobiose and cellobiose epimerase. The crude enzyme was purified by gel filtration on Sephadex G-100 (Sigma) or Sephadex G-50 (Pharmacia). The protein was eluted with 0.1 M sodium phosphate buffer, pH 6.6, containing 0.5 M NaCl. The protein concentration was determined by the method of Lowry et al. (1951) with bovine serum albumin as the standard. The specific activity of the enzyme was determined as the rate of hydrolysis of 1 mg substrate in 0.1 ml buffer pH 6.6, using 0.1 M sodium phosphate buffer, pH 6.6, containing 0.5 M NaCl.
site was detected between amino acids A-32 and D-33, leading to a mature protein of 628 aa with a molecular mass of 71677 Da. A BLAST search and subsequent sequence alignment revealed a complex architecture consisting of several protein modules (Fig. 1b): a 150 aa stretch with homology to the CBM3 family (amino acids S-35 to G-185); a short proline-rich linker; 384 aa of a protein domain (amino acids N-197 to G-580) with homology to enzymes of GHF5 (also known as family A); and 74 aa containing a well-conserved duplicated 24 aa region composing a dockerin module type I (amino acids K-587 to E-660). According to the nomenclature suggested recently the protein CelO can be unequivocally described as Ct-CelSF with the structural formula CBM3b/CD5/DD; Cd designates its C. thermocellum origin, CD the catalytic domain and DD the dockerin domain (Henrissat et al., 1998). The presence of the dockerin module sequence is considered a clear indication that CelO can dock to a cohesin module of the cellulosome scaffold protein CipA and thus is present in the cellulosome complex (Bayer et al., 1998a; Tokatlidis et al., 1991).

To characterize CelO, and the role of its modules, two truncated proteins were constructed by PCR: rCelO, with the leader peptide and the dockerin module deleted (587 aa, 67272 Da), and rCelO-Cat, containing only the catalytic domain devoid of the CBM of CelO (415 aa, 47946 Da). The resulting recombinant enzymes were purified to homogeneity by affinity chromatography of the hexahistidyl-tagged recombinant proteins and showed a molecular mass consistent with that calculated from the sequence.

The hydrolytic activities of rCelO and rCelO-Cat were identical within the error limits of the assay if molar activities were calculated (Table 1). rCelO hydrolysed soluble and insoluble 1,4-β-linked substrates, with the highest activity shown for the soluble barley β-glucan, containing alternating 1,3-β- and 1,4-β-linkages in a ratio of 1:3, whereas no activity could be observed with laminarin, a 1,3-β-glucan. rCelO released reducing sugar from soluble CMC and from PASC, an amorphous cellulose preparation. Only minor activity near the detection limit was observed with xylan (1,4-β-p-xylan) or pustulan (1,6-β-D-glucosidase, 1,4-β-xyloside, 1,3-β-mannoside, lactoside and 1,4-β-galactoside were not hydrolysed. With an optimum temperature of 65°C at its optimal pH of 6.6 for hydrolysis of barley β-glucan, CelO is a relatively thermostable enzyme. These values are comparable to those found for other exoenzymes from the thermophilic bacteria C. thermocellum and Clostridium stercorarium.

The hydrolytic activity of rCelO on CMC showed severe substrate limitation with a maximum hydrolysis of 4% (w/w) of the substrate (Fig. 2). The addition of new substrate resulted in an equal increase of reducing equivalents, excluding inactivation of the enzyme or product inhibition. This substrate limitation suggested an exo-glucanolytic mode of enzyme action. Further evidence for an exo-mode of action came from viscosity

---

**Table 1. Substrate specificity of rCelO**

One unit of activity is defined as the amount of enzyme producing 1 μmol reducing sugar min⁻¹. The measurements were done at least in triplicate and the standard errors were calculated.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>rCelO Specific activity [U (mg protein)⁻¹]</th>
<th>Relative activity (%)</th>
<th>rCelO-Cat specific activity [U (mg protein)⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barley β-glucan</td>
<td>340 (±16)</td>
<td>100</td>
<td>485 (±22)</td>
</tr>
<tr>
<td>CMC</td>
<td>7.5 (±0.4)</td>
<td>2.2</td>
<td>10.6 (±0.5)</td>
</tr>
<tr>
<td>PASC</td>
<td>2.1 (±0.1)</td>
<td>0.62</td>
<td>2.9 (±0.2)</td>
</tr>
<tr>
<td>Xylan</td>
<td>0.3 (±0.03)</td>
<td>0.09</td>
<td>0.4 (±0.04)</td>
</tr>
<tr>
<td>Pustulan</td>
<td>0.2 (±0.02)</td>
<td>0.06</td>
<td>0.3 (±0.03)</td>
</tr>
<tr>
<td>Avicel</td>
<td>0.002 (±0.0007)</td>
<td>0.001</td>
<td>0.003 (±0.001)</td>
</tr>
</tbody>
</table>

**Fig. 2.** Substrate limitation during hydrolysis of CMC by CelO. Five milligrams of CMC was digested with rCelO, and the amount of reducing sugars released was determined. The arrow indicates the addition of new substrate after 60 min.
measurements on CMC, where rCelO displayed a smaller slope of specific fluidity increase than the control encoglucanase Cel from *C. thermocellum* (Fig. 3) (Hazlewood et al., 1993). However, as for Ct-CbhA (Tuka et al., 1990), the considerable reduction in viscosity suggested that CelO acts in a processive mode after making random cuts in cellulose chains.

To analyse the enzymic exo-mode of CelO further, the degradation products appearing from hydrolysis of PASC were determined; cellobiose could be detected immediately at the onset of the reaction and was the only product, even after long incubation times, as is typical for cellobiohydrolases (Fig. 4). Cellobiose was also observed after prolonged digestion of crystalline cellulose (Avicel) with a high enzyme concentration, indicating activity on crystalline cellulose. The hydrolysis of cellodextrins corroborated these results; cellobiose was produced from cellotetraose and cellohexaose, and cellobiose and cellotriose were produced from cellopentaose. Cellobiose and cellotriose were not hydrolysed further, even after extended incubation (Fig. 4).

Many cellobiohydrolases (exoglucanases) hydrolyse pNP-cellobioside readily, a reaction which was, until recently, used as an indicator for cellobiohydrolase activity. Surprisingly, rCelO completely lacked chromogenic activity on the pNP-cellodextrins from pNP-glucoside to pNP-cellopentaoside. Although p-nitrophenol was not released, pNP-glucoside was indeed split from pNP-cellotetraoside and pNP-cellotrioside together with cellotriose and cellobiose, respectively, as shown by TLC (Fig. 5). The cellobiose also detected in the chromatogram was obviously the hydrolysis product of the contaminating cellotetraose present in equal amounts in the pNP-cellotetraoside preparation. pNP-cellobioside was not degraded (data not shown).

To confirm the unusual result of pNP-glucoside release, hydrolysis was repeated with a substrate labelled by NaBH₄-reduction at the reducing end. Non-reduced cellopentaose was hydrolysed readily to cellobiose and cellotriose by rCelO; however, reduced cellopentaose was not hydrolysed (no reduced cellobiose or cellotriose could be detected) (Fig. 6). To verify the degradability of the chemically treated substrate, cellobiohydrolase (CbhA) was applied; cellobiose and reduced cellotriose were produced from the reduced cellopentaose, as was expected for CbhA, which hydrolyses cellodextrins from the non-reducing end.
The celO gene contained, at its N-terminus, a sequence with similarity to the CBM family III (CBM3). In a preliminary experiment the purified complete enzyme was found to bind to Avicel, whereas no binding was observed with rCelO-Cat. The binding of rCelO to PASC and to Avicel was quantitatively assayed, by determining the concentration of bound protein. The adsorption constants \( [K_a = 1.7 \pm 0.3] \times 10^6 \text{ M}^{-1}\) and \(3.7 \pm 0.5) \times 10^6 \text{ M}^{-1}\), respectively] and the capacity values for Avicel and PASC \(1.9 \pm 0.3 \text{ and } 3.3 \pm 0.5 \mu\text{M} \text{ g}^{-1}\), respectively] were calculated from the plot of the equilibrium adsorption isotherm.

**DISCUSSION**

The identified CBM contained in CelO, CelO-CBM3, was homologous to the CBM3 family, which can be subdivided into three subfamilies (Tomme et al., 1995). Many of the proposed cellulose-binding amino acids of CelO-CBM3 were conserved, but the ‘scaffoldin’ loop was missing, suggesting its assignment to subfamily CBM3b (Bayer et al., 1998b). Investigation of the binding properties of CelO-CBM3 showed that it bound to amorphous and crystalline cellulose substrates, as was expected by extrapolation from the known structure of Ct-CipA-CBM3 (subfamily 3a) (Tormo et al., 1996). Binding constants were twofold higher for the crystalline substrate Avicel than for the amorphous cellulose preparation PASC. For the binding module of another cellulosomal celllobiohydrolase, Ct-CipA-CBM3 (subfamily 3a), a 20-fold higher \( K_a \) value towards Avicel has been determined (Pages et al., 1997). However, the lower \( K_a \) value of CelO-CBM3 is in conformity with the lack of a ‘scaffoldin’ loop within the structure. The only other binding domain of subfamily 3b with known binding parameters is the C-terminal Cs-CelZ-CBM3b module from C. stercorarium (Riedel et al., 1998). However, due to the different methodology and buffer systems used, comparison of the data is problematic. Nevertheless, the higher \( K_a \) for Cs-CelZ-CBM3b might be explained by the fact that C. stercorarium does not produce a cellulosome which could assist in substrate binding: Cs-CelZ (Jauris et al., 1990) is a free cellulase, like the majority of the enzymes containing CBM3b modules. Such enzymes may have developed an optimized substrate-binding capacity of their own, and the cellulose-binding module may have evolved to support the catalytic domain in the hydrolysis of the crystalline substrate.

No negative effect on the catalytic activity or on the thermal stability of the enzyme was observed on deletion of the CBM3b in rCelO-Cat. This might be explained by the fact that CelO, unlike the non-cellulosomal enzymes, is tightly integrated in the cellulosome complex, which already contains the very strongly binding CBM3a of the scaffoldin Ct-CipA. It can be assumed that the cellulosomal enzyme components in C. thermocellum would not be dependent on additional CBMs. However, their presence could stabilize the whole complex on the substrate surface, having a back-up-function and supporting the overall activity of the cellulosome.

The catalytic GHF5 module of CelO belongs to a heterogeneous family of glucanases, subdivided into five subfamilies with amino acid sequence similarity above 25%. CelO showed a narrow substrate specificity, restricted to 1,4-β-linked substrates like CMC, cellobiose, and amorphous and microcrystalline cellulose. The CelO sequence grouped with subfamily 1 (formerly called family A1), all members of which are hitherto described as endoglucanases, mostly based on the fact of CMC hydrolysis (Lemaire & Béguin, 1993; Meinke et al., 1993; Sakon et al., 1996). Only the hydrolytic mode of cellobiose and cellulose (Ferreira et al., 1991). Despite hydrolysing CMC with a high initial reaction velocity, CelO released exclusively cellobiose from Avicel, PASC, and cellobiase. The CelO activity on CMC showed pronounced substrate limitation and a relatively small decrease in viscosity, typical for a processive mode of hydrolysis. CelO behaved as an exo-glucanolytic celllobiohydrolase (EC 3.2.1.91), although the decrease in viscosity and the amount of reducing sugars released...
from CMC does not exclude a limited ability for random attack.

The high activity on mixed-linkage β-glucan is a trait CelO has in common with most of the cellobiohydrolases described so far. The mechanism by which sequentially active cellobiohydrolases overcome the potential block of 1,3-β-linkages in the linear substrate molecule deserves further analysis. The apparent low activity of CelO on Avicel, only one-tenth of the value of the main cellulosomal component Ct-CelS (Kruus et al., 1995), may be drastically enhanced if CelO is correctly integrated into the synergistically acting cellulosome. On the other hand, Avicel might not be its natural substrate.

The lack of chromogenic activity on pNP-cellobiose (pNP-cellobioside, -tri-, -tetra- and -pentao- side) was at first glance interpreted as an absence of activity, because cellobiose was also not hydrolysed by CelO. However, TLC analysis of the reaction products detected the hydrolysis of pNP-cellobioside to pNP-glucoside and cellobiose, and could be explained by hydrolysis from the reducing end: if the substrate were hydrolysed from the non-reducing end, the expected products would be cellobiose and pNP-cellobioside. It cannot be ruled out that interactions with the aromatic aglycone may affect the pattern of degradation of pNP-cellobiose or Avicel. However, the model was further supported by another experiment using reduced cellobiose as a substrate. Whereas cellobiohydrolase Ct-CbhA hydrolysed the substrate from the unmodified non-reducing end, CelO degraded only the unmodified substrates, while activity on the reduced substrates was blocked by modification at the reducing end. Fig. 7 explains the proposed model of the hydrolytic mode schematically.

A model explaining how enzymes with similar primary sequences can be converted to either endo- or exo-

Cellobiohydrolase was proposed previously (Divne et al., 1994; Zverlov et al., 1999). A multiple sequence alignment with CelO revealed two insertions compared to the sequence of E1 of Acidothermus cellulolyticus, both of which lie on the same side of the (α/β)-barrel around amino acids 175 and 230, and which form loop regions in the neighbourhood of the catalytic and substrate-interacting amino acids (Fig. 8) (Sakon et al., 1996). These loop extensions may form a tunnel hindering free access of the polymeric substrate chain and thus provoking a threading through of the substrate from its end and along the catalytic site, which would lead to a presumably processive catalytic mode.

Four other cellulases of subfamily 1 have loop extensions similar to CelO: Csa-GunB of Caldocellulosa saccharolyticum; Ct-CelB and Ct-CelG of Clostridium thermocellum; and Cf-CenD of Cellulomonas fimi (Sakon et al., 1996). According to the above model, an exo-mode of action can be predicted for these enzymes. However, no evidence for the production of cellobiose has been presented. In general, the hydrolysis of CMC, together with the lack of activity on pNP-cellobiose or Avicel, is taken as an indication for an endo-mode of action. If investigated further, some of these enzymes could well turn out to be cellobiohydrolases with a reducing-end specificity, as was the case with CelO. The following data are indications for such an assumption. For Csa-GunB and Ct-CelG the hydrolysis of pNP-cellobiose has been reported, indicating that these enzymes could have at least some cellobiohydrolase activity from the non-reducing end (Lemaire & Béguin 1993; Saul et al., 1990). The direction of hydrolysis may, therefore, not be conserved within an enzyme family. Ct-CelB was shown to hydrolyse CMC with a lesser reduction in viscosity than was measured for control endoglucanases, a clear indication for an exo-mode (Joliff et al., 1986). Cf-CenD hydrolysed cellobiose predominantly to cellobiose and cellobiose, with only the β-isomer of cellobiose, before the onset of mutarotation (Meinke et al., 1993). This can only be interpreted as cleavage of a cellobiose residue from the reducing end, producing a cellobiose

Fig. 7. Model of the proposed mode of action of CelO on cellopentaose. Upper line, cellopentaose; lower line, pNP-cellotetraoside. NR and R represent the non-reducing end and the reducing end, respectively.

Fig. 8. Proposed secondary structure of the catalytic domain of CelO. α-Helices, β-sheets and loops are arranged according to a sequence alignment with the endoglucanase E1 from Acidothermus cellulolyticus (Sakon et al., 1996). Arrows indicate β-sheets; cylinders indicate α-helices. Stars represent potential amino acids of the active side and solid circles represent conserved amino acids. Triangles depict insertions present in CelO.
with a new reducing end. Ct-CelB and Cf-CenD would therefore be candidates for further investigations into the potential exo-mode of reducing-end activity in cellulose hydrolysis.

The data show that GHF5 is a cellulase family containing both endo- and exoglucanases. CelO is the first cellobiohydrolase in the cellulosome of *C. thermocellum* for which evidence of reducing-end specificity has been presented. As far as the cellulosome of *C. thermocellum* is concerned, two exoglucanases each seem to be active from either side of the polymeric substrate: CelK and CbHα from the non-reducing side (Kataeva et al., 1999; Zverlov et al., 1998), and CelO and probably Cels (Teeri, 1997) from the reducing end. This, together with the presence of binding domains for crystalline cellulose, would enable the cellulosome to widen a gap in a cellulose molecule on the crystal surface in both directions.

**ACKNOWLEDGEMENTS**

This work was supported by a grant from RFFI to V.V.Z. and G.A.V. (reference no. 00-04-48197), grants from the Deutsche Forschungsgemeinschaft DFG to V.V.Z. (reference no. 436 RUS 15/57/99 and 17/88/00), and a grant from the Leonhard-Lorenz Foundation to W.H.S. We are very grateful to W.L. Staudenbauer for useful comments and for critically reading the manuscript.

**REFERENCES**


complex with inhibitors and substrates give insights into its processive action. Biochemistry 39, 11238–11246.


