Biochem. J. (2003) **370**, 963–970 (Printed in Great Britain)

β-Glucosidase in cellulosome of the anaerobic fungus *Piromyces* sp. strain E2 is a family 3 glycoside hydrolase

Peter J. M. STEENBAKKERS, Harry R. HARHANGI, Mirjam W. BOSSCHER, Marlous M. C. VAN DER HOOFT, Jan T. KELTJENS, Chris VAN DER DRIFT, Godfried D. VOGELS and Huub J. M. OP DEN CAMP[1]

Department of Microbiology, Faculty of Science, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands

The cellulosomes of anaerobic fungi convert crystalline cellulose solely into glucose, in contrast with bacterial cellulosomes which produce cellobiose. Previously, a *β*-glucosidase was identified in the cellulosome of *Piromyces* sp. strain E2 by zymogram analysis, which represented approx. 25 % of the extracellular *β*-glucosidase activity. To identify the component in the fungal cellulosome responsible for the *β*-glucosidase activity, immunoscreening with anti-cellulosome antibodies was used to isolate the corresponding gene. A 2737 bp immunoclone was isolated from a cDNA library. The clone encoded an extracellular protein containing a eukaryotic family 3 glycoside hydrolase domain homologue and was therefore named *cel3A*. The C-terminal end of the encoded Cel3A protein consisted of an auxiliary domain and three fungal dockerins, typical for cellulosome components. The Cel3A catalytic domain was expressed in *Escherichia coli* BL21 and purified. Biochemical analyses of the recombinant protein showed that the Cel3A catalytic domain was specific for *β*-glucosidic bonds and functioned as an exoglucohydrolase on soluble substrates as well as cellulose. Comparison of the apparent K_m and K_i values of heterologous Cel3A and the fungal cellulosome for *p*-nitrophenyl-*β*-D-glucopyranoside and D-glucono-1,5-*δ*-lactone respectively indicated that *cel3A* encodes the *β*-glucosidase activity of the *Piromyces* sp. strain E2 cellulosome.

Key words: cellulose degradation, exoglucanase, product inhibition.

INTRODUCTION

Anaerobic fungi and several Clostridia species excrete a high molecular mass (hemi)cellulolytic complex during growth on sugar polymers [1–6]. These complexes, or cellulosomes, consist of individual catalytic components and a central organizing protein, named scaffoldin, which binds the enzymes by its cohesin sites. The catalytic components in turn require a dockerin domain for incorporation into the complex.

Although both fungal and bacterial cellulosomes have a similar role during microbial growth on plant cell wall material, there are significant differences between both complexes. In addition to the differences in dockerin structure [7] and their relative position in the cellulosome components, the two cellulosome types differ in the glycoside hydrolase (GH) families present as catalytic components. The *Clostridium thermocellum* cellulosome has been extensively characterized [8] and contains cellulases from five GH families according to the classification of Henrissat and Davies [9] and Henrissat and Bairoch [10]: GH family 5 (CelG, CelB, CelE, CelH), GH family 8 (CelA), GH family 9 (CelD, CelF, CelK, cbhA, CelJ), GH family 44 (CelJ) and GH family 48 (CelS). The fungal cellulosomes analysed so far also contain components classified as family 5 [11], 9 [12] and 48 [13] GHs. However, they also contain enzymes from GH families 6 [14] and 45 [11], which have not been identified in bacterial cellulosomes.

The most profound biochemical difference between the bacterial and fungal cellulosomes is the end product produced during the hydrolysis of crystalline cellulose. Degradation of cellulose by clostridial cellulosomes results in the production of cellobiose which is taken up into the cell, hydrolysed to glucose and fermented [5]. The *Piromyces* sp. strain E2 cellulosome, however, produces solely glucose during cellulose degradation [1] as do the complexes from *Piromonas communis* P [2] and *Neocallimastix frontalis* [3]. No significant accumulation of oligosaccharides is detectable, which indicates the presence of very effective *β*-glucosidase activity. The factor in the fungal cellulosome responsible for the *β*-glucosidase activity is therefore one of the key components.

Already, a number of cellulases and *β*-glucosidases have been identified from *Piromyces* sp. strain E2 [15]. Harhangi et al. [16] showed the existence of a highly expressed *β*-glucosidase gene, encoding an extracellular, non-complexed GH family 1 enzyme with multiple copies on the *Piromyces* sp. strain E2 genome. Based on its kinetic characteristics and the transglycosylating capacity of the heterologous protein, this enzyme is presumably involved in the induction of the cellulolytic system and not in cellulose hydrolysis. Zymogram analysis of the *Piromyces* sp. strain E2 cellulosome indicated the presence of a single cellulosomal *β*-glucosidase band which represented 27 % of total extracellular *β*-glucosidase activity [17]. In the present study, we describe the isolation of the gene encoding this key component of the anaerobic fungal cellulosome and the characterization of the heterologous enzyme.

EXPERIMENTAL

Immunoscreening of the *Piromyces* sp. strain E2 cDNA library

A *Piromyces* sp. strain E2 cDNA library (described in [18]) in λZAP II (Stratagene, La Jolla, CA, U.S.A.) was screened with anti-cellulosome antibodies in accordance with the manufacturer's picoBlue instruction manual [12]. Of the rescreened immunoreactive clones, 32 were sequenced using pBluescript SK vector primers.

Abbreviations used: GH, glycoside hydrolase; ORF, open reading frame; pNP, *p*-nitrophenyl.

[1] To whom correspondence should be addressed (e-mail huubcamp@sci.kun.nl).

The nucleotide sequence data reported is available in the GenBank® Nucleotide Sequence Database under the accession number AY172977.

© 2003 Biochemical Society

Recombinant DNA techniques

The full-length *cel3A* sense strand was sequenced using nested deletions with MBI Fermentas reagents (Munich, Germany). Nested deletions subclones were selected using a rapid screening procedure [19] and sequenced using vector primers. The antisense strand was sequenced with primers, designed from the sense strand sequence. Sequencing reactions were performed with the dRhodamine sequencing kit (PerkinElmer, Norwalk, CT, U.S.A.) or the Beckman cycle sequencing kit (Beckman, Fullerton, CA, U.S.A.), using an automated ABI prism sequencer (PerkinElmer) or a Beckman CEQ 2000 sequencer respectively.

Sequence analysis

Cel3A sequences were assembled using vector NTI Suite software (Informax, Golden, CO, U.S.A.). The full-length DNA and protein sequence was compared with deposited sequences in GenBank® [20] and the protein database [21] using BLAST (basic local alignment search tool) [22]. The putative signal peptide of the encoded protein was predicted using SignalP [23] at default settings. Further analysis was performed by using the online facilities of the Swiss ExPasy server (http://www.expasy.ch), the Pfam site of the Sanger Institute (Wellcome Trust Genome Campus, Hinxton, Cambridge, U.K.) (http://www.sanger.ac.uk/Software/Pfam/) and the CAZy site (http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html) [9,10]. For phylogenetic analysis of the Cel3A protein sequence with bootstrapping, an alignment was constructed with 14 entries from GH family 3, using ClustalW program [24]. The phylogenetic tree was calculated using the cluster algorithm of the GeneBee TreeTop software (http://www.genebee.msu.su/services/phtree_full.html) [25].

Expression and purification of *cel3A* in *Escherichia coli* BL21

The catalytic and auxiliary domains were amplified with *pfu* proofreading polymerase using the *cel3A* immunoclone as a template. The primers used annealed to the region immediately after the predicted signal peptide splicing site and the linker region immediately in front of the first fungal dockerin domain, yielding *rCel3A*: rCel3A forward, 5′-*CCATGG*CTACTTCTT-GGAGTGAAG-3′ and rCel3A reverse, 5′-*GGATCC*GTCAG-CGTTGTTGTTTGG-3′. The PCR product was purified (Qiagen gel extraction kit, Hilden, Germany), tailed using dATP and *Taq* polymerase and subsequently cloned in the pGEM T-Easy vector according to the manufacturer's recommendations. The amplified *rCel3A* was cloned into the expression vector pQE60 using the *Nde*I and *Bam*HI restriction sites (depicted in italics in the primer sequences), ligating it in frame with the pQE60 encoded C-terminal His tag. The sequence of the pQE60-rCel3A expression construct was confirmed by sequencing.

The pQE60-rCel3A expression construct was co-transformed with pREP4 encoding the *lac*I repressor to competent *E. coli* BL21 cells. Transformants were selected on Luria–Bertani broth plates containing 100 μg/ml of ampicillin, 25 μg/ml of kanamycin and 0.5% (w/v) glucose. A single colony was transferred to 5 ml of 2 × YT medium (16 g/l of tryptone, 10 g/l of yeast extract and 5 g/l of NaCl) containing the same amounts of antibiotics and glucose and was cultured for 16 h at room temperature (20 °C). The culture (1 ml) was transferred to 250 ml of prewarmed 2 × YT medium containing the same amounts of antibiotics and glucose, and grown at 37 °C until the A_{600} reached 1.0. The culture was cooled for 15 min to room temperature and isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 1 mM. Induction was continued for 4 h at room temperature with steady stirring (250 rev./min). After induction, cells were harvested by centrifugation at 10000 ***g*** for 20 min at 4 °C and resuspended in 1 pellet volume of 200 mM Na_2HPO_4/NaH_2PO_4 buffer (pH 8) and 2 × protease inhibitor cocktail (mini tablets complete; Boehringer Mannheim, Mannheim, Germany). Resuspended cells were lysed using a French press and cell-free extract was collected as the supernatant after centrifugation at 20000 ***g*** for 30 min at 4 °C. Before subsequent purification of the recombinant protein, the concentration of NaCl and imidazole was adjusted to 300 and 10 mM respectively. The rCel3A protein containing a C-terminal His tag was purified from the cell-free extract using a 10 ml Ni^{2+}-nitrilotriacetate column, following the manufacturer's recommendations. Eluted protein was pooled and stored in 40% (v/v) glycerol at −20 °C.

Characterization of rCel3A

Purified rCel3A was separated on a standard 7.5% (w/v) Laemmli SDS gel and stained with Coomassie Blue. All biochemical assays with rCel3A and the purified cellulosome of *Piromyces* sp. strain E2 (purification is described in [12]) were performed in duplicate or triplicate. Protein concentration was determined with the Bio-Rad protein assay solution using BSA as a standard. The optimum pH was determined at 40 °C in 0.1 M KH_2PO_4/K_2HPO_4 buffer at different pH values using *p*-nitrophenyl-β-D-glucopyranoside (pNP-β-D-glucopyranoside) as the substrate. The substrates tested were pNP-β-D-glucopyranoside, 5 mM pNP-β-D-xylopyranoside, 5 mM pNP-β-D-fucopyranoside, 5 mM pNP-β-D-galactopyranoside, pNP-β-D-cellobiose, 5 mM pNP-α-D-glucopyranoside, cellobiose and Avicel PH 105 (cellulose). The assays for the pNP compounds and cellobiose were performed as described previously [1]. The activity of rCel3A against cellulose was performed by mixing 1 ml of a 10% (w/v) Avicel suspension in 50 mM Na_2HPO_4/NaH_2PO_4 buffer (pH 6.0) with 990 μl of double-distilled water. After prewarming at 37 °C, the reaction was started by adding 10 μl of 2.75 mg/ml rCel3A stock. The assay was continued under continuous mixing at 37 °C. At $t = 0$ min and every 15 min, 150 μl of sample was taken and boiled for 5 min. The glucose concentration was determined in the supernatant after centrifugation. The Avicel suspension and incubations of rCel3A with only the supernatant of the Avicel suspension were included as controls. A 24 h incubation of rCel3A with cellulose was performed in a 2% (w/v) Avicel suspension in 25 mM Na_2HPO_4/NaH_2PO_4 buffer (pH 6.0) at 37 °C under continuous mixing with 10 μl of 2.75 mg/ml rCel3A stock. The glucose concentration was determined using the Sigma glucose detection kit. Apparent K_m and V_{max} values were calculated using linear regression for pNP-β-D-glucopyranoside and cellobiose from Lineweaver–Burk plots. The correlation coefficients were higher than 0.99 in all cases. Apparent K_i values for glucose and D-glucono-1,5-δ-lactone were determined with pNP-β-D-glucopyranoside as the substrate. Activity is given in international units, corresponding to 1 μmol of glucose or *p*-nitrophenol formed per min, unless mentioned otherwise.

RESULTS AND DISCUSSION

Cellulosomal β-glucosidase from anaerobic fungi

Fungal complexes are the most efficient cellulose-degrading complexes described and they exceed the cellulose-degrading capacity of cellulolytic systems of aerobic fungi and bacterial cellulosomes [26]. One possible explanation for the superior cellulose-degrading capacity of fungal cellulosomes could be the

© 2003 Biochemical Society

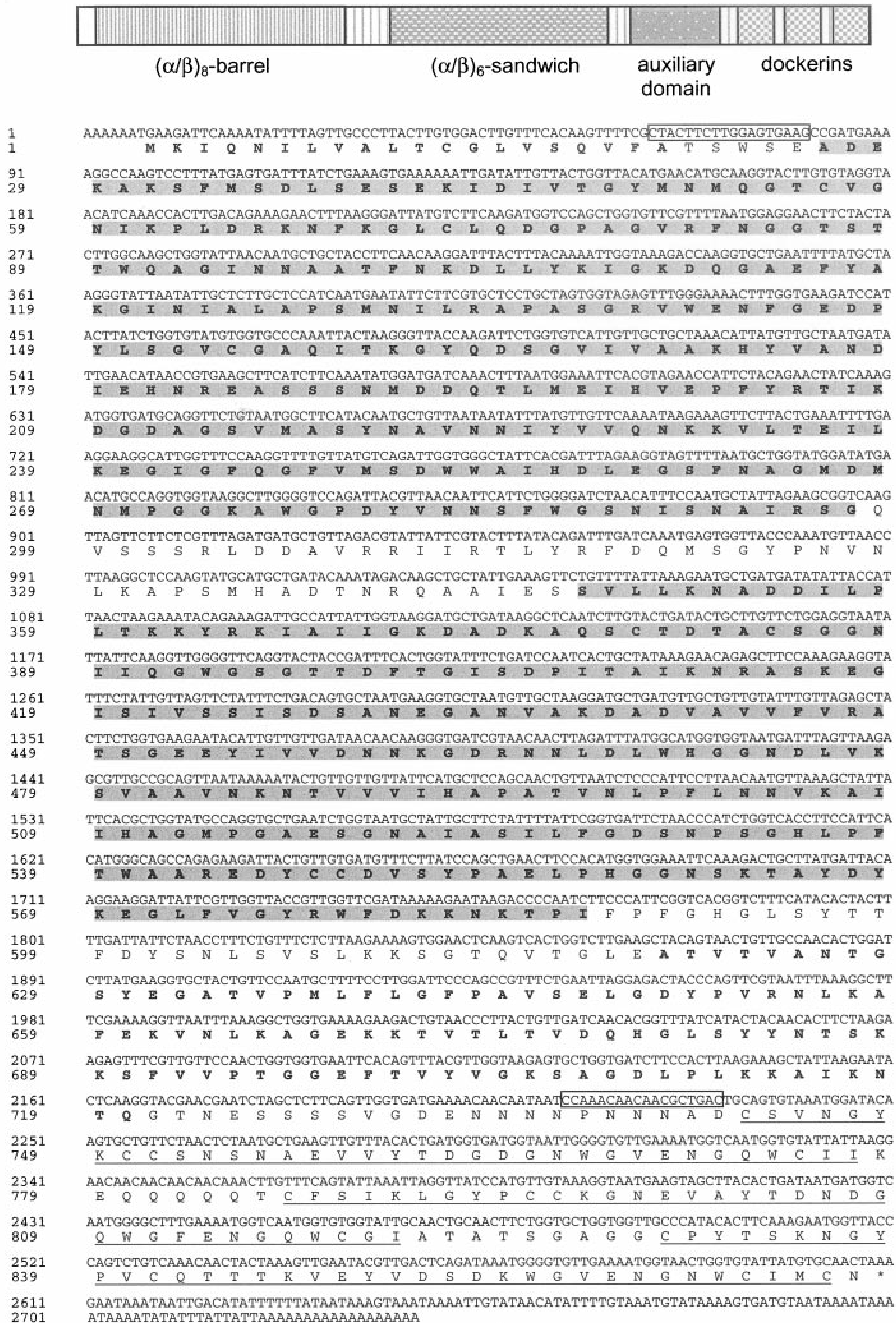

Figure 1 Cel3A, DNA and protein sequence

A scale representation of the modular architecture of Cel3A and the nucleotide and deduced amino acid sequences of *cel3A*. Nucleotides and amino acids are numbered on the left-hand side. The stop codon is indicated by an asterisk. The putative signal sequence and the auxiliary domain are depicted in boldface. The N- and C-terminal substructure sequences are shaded. The three fungal dockerins are underlined. Primer sequences used for rCel3A are enclosed by boxes.

© 2003 Biochemical Society

presence of a high β-glucosidase activity. β-Glucosidase would prevent the accumulation of cellobiose, which is known to be a potent inhibitor of the cellulases in bacterial cellulosomes and uncomplexed cellulases [27,28]. To isolate the gene encoding the β-glucosidase component of the fungal cellulosome, immunoscreening of a cDNA library was applied using anti-cellulosome antibodies.

Isolation of *cel3*A

Approximately 60000 plaques from the *Piromyces* sp. strain E2 cDNA library were screened using anti-cellulosome antibodies. Fifty plaques positive in the primary screening were rescreened and 32 plasmids were sequenced. The nucleotide sequences were compared with database sequences using BlastX. Six out of 32 clones appeared to originate from the same allele and the encoded proteins showed high similarity (approx. 55%) to members of GH family 3. Five cDNA clones were incomplete and one presumed full-length gene was further analysed. The full-length clone was designated *cel3A*, and the encoded protein Cel3A, according to the GH classification of Henrissat and Davies [9] and Henrissat and Bairoch [10].

The nucleotide sequence of *cel3A*

The *cel3A* cDNA clone contained a 2739 bp insert with an open reading frame (ORF) of 2604 bp (Figure 1). The gene contained all features typical for full-length genes from anaerobic fungi. The putative start codon was preceded by a 3 bp adenosine stretch. Secondly, the 130 bp 3′-flanking region was AT-rich (92%) in comparison with the 63% A+T in the ORF and also the codon usage showed a typical preference for A or T in the wobble position [29].

Amino acid sequence analysis of Cel3A

Translation of the *cel3A* ORF resulted in a 867 residue putative preprotein, designated Cel3A, with a calculated mass of 93.588 kDa (Figure 1). Analysis of the encoded protein using SignalP for eukaryotic proteins predicted the presence of a 20-amino-acid signal peptide, consistent with the extracellular nature of the cellulosome: MKIQNILVALTCGLVSQVFA (Figure 1). It contained the typical charged/hydrophilic residues at the N-terminus, followed by a hydrophobic stretch. Also, the signal peptidase site conformed with the (−1, −3) rule stated by von Heijne [30]. The predicted mature Cel3A protein had a calculated molecular mass of 91.457 kDa and a pI of 5.4.

The amino acid sequence was compared with database sequences and revealed a modular architecture. The N-terminal 700 residues showed extensive sequence homology with fungal β-glucosidases classified as family 3 GH. The catalytic domain was followed by a 25 residue linker region, rich in serine and asparagine residues and the C-terminal end consisted of three copies of the fungal docking domain (Figure 1).

The Cel3A catalytic domain

A comparison of Cel3A with database sequences revealed that the *Piromyces* sp. strain E2 protein resembled fungal β-glucosidases classified as family 3 GH. The GH family 3 contains more than 200 members which have been identified in higher eukaryotes, plants, bacteria and fungi. From these enzymes, 144 have been analysed in detail and are included in the GH family 3 description of Pfam (Sanger Institute). Only one structure has been resolved for GH family 3, i.e. the exoglucanase Exo1 of barley (*Hordeum vulgare*) [31]. The GH family 3 barley enzyme contains two domains connected by a large hinge region. The N-terminal domain, approx. 250–300 residues, forms an $(\alpha/\beta)_8$ barrel and is connected to the second domain by a 50-amino-acid linker. The second domain of approx. 200–250 amino acids forms an $(\alpha/\beta)_6$ sandwich and is located adjacent to the N-terminal domain. Both domains share a large interface which forms the catalytic site, with the nucleophile located in the $(\alpha/\beta)_8$ barrel and the acid/base in the sandwich structure. In the β-glucosidase of *Butyrivibrio fibrisolvens* (M31120) and one of *Ruminococcus albus* (X15415), the order of these domains is reversed.

Coccidioides immitis	275 -	GFIMS**D**WQAHH	- 285
Ajellomyces capsulata	276 -	GFIMS**D**WQAHH	- 286
Aspergillus aculeatus	275 -	GFVMS**D**WGAHH	- 285
Gaeumannomyces graminis	274 -	GFVVS**D**WAATH	- 284
Trichoderma reesei, β-xylosidase	306 -	GYVSS**D**CDAVY	- 316
Aspergillus niger, β-xylosidase	310 -	GYVSS**D**CDAAY	- 320
Flavobacterium meningosepticum	242 -	GFIVT**D**YTGIN	- 252
Ruminococcus albus	234 -	GLVMS**D**WGAVD	- 244
Clostridium thermocellum	226 -	GFVVS**D**WGAVN	- 236
Hordeum vulgarum	305 -	GFVIS**D**WEGID	- 315
Aspergillus niger, β-glucosidase	275 -	GFVMS**D**WAAHH	- 285
Trichoderma reesei, β-glucosidase	261 -	GYVMT**D**WNAQH	- 271
Phanerochaete chrysosporium	330 -	GYVMS**D**WWATH	- 340
Piromyces sp. strain E2	217 -	GFVMS**D**WWAIH	- 227

Figure 2 Cel3A catalytic nucleophile

Partial alignment of the N-terminal part of the Cel3A catalytic domain to other members of GH family 3. The glutamate indicated by the arrow has been shown to act as the nucleophile in the hydrolysis mechanism of the retaining GH family 3 enzymes from *Aspergillus niger* [34], *Hordeum vulgare* [31] and *Flavobacterium meningosepticum* [35]. The database entries are given in Figure 3.

To investigate the architecture of the *Piromyces* Cel3A enzyme the sequence was submitted at the Pfam site of the Sanger Institute. Both structural halves could be identified. The N-terminal domain forming the $(\alpha/\beta)_8$ barrel (residues 26–297) was separated from the $(\alpha/\beta)_6$-sandwich domain (residues 347–587) by a 50-amino-acid linker. A BlastP search against the deposited database sequences showed that Cel3A shared highest homology with the cellulose-binding domain containing β-glucosidase from *Phanerochaete chrysosporium* [32] and the β-glucosidase from *Trichoderma reesei* QM9414. Residues 21–716 in the *Piromyces* sequence showed approx. 40% sequence identity and 55% sequence similarity to both sequences. Sequence similarity started at residue 21, supporting the predicted signal peptide of Cel3A.

GH family 1 and 3 enzymes are retaining glycohydrolases, which require amino acids in the active site that act as the nucleophile and the general acid/base respectively in the reaction mechanism [33]. To investigate further sequence conservation and to identify the catalytic residues, Cel3A was compared with the β-glucosidase BGL1 from *Aspergillus niger* (AJ132386) [34], barley exoglucanase Exo1 (AF102868) [31] and Fbgl from *Flavobacterium meningosepticum* [35,36] because of the kinetic and structural knowledge that has been published for these proteins. The catalytic nucleophile has been identified in all three sequences but the general acid/base residue has only been identified in the *F. meningosepticum* enzyme. Analysis of the alignment showed Cel3A residue D251 as the catalytic nucleophile, which was situated in a conserved motif: SDW (Figure 2). However, a putative catalytic acid/base residue could not be deduced from the sequence comparison because of low-sequence homology in the C-terminal part of the catalytic domain. It should be noted that β-*N*-acetylglucosaminidases from GH family 3 (e.g. β-*N*-acetylglucosaminidases from *E. coli*, SwissProt

© 2003 Biochemical Society

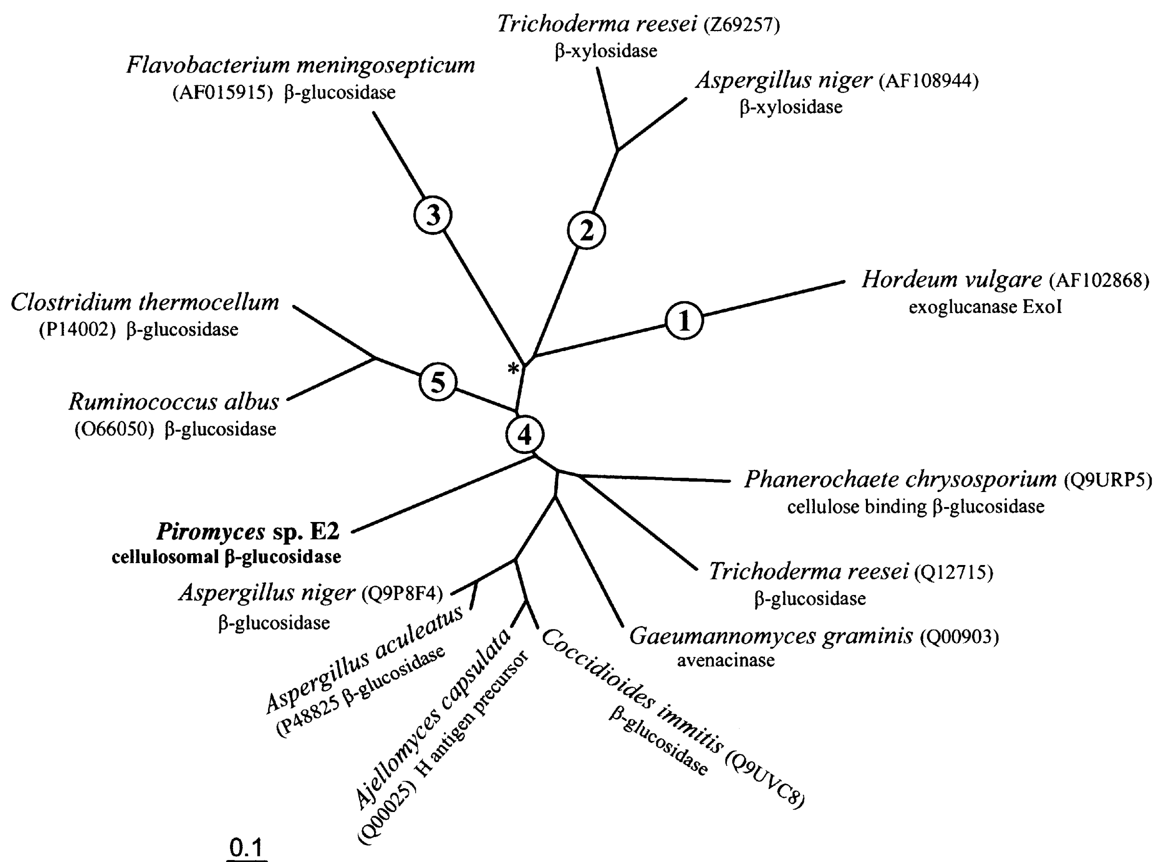

Figure 3 Cel3A phylogeny

Phylogenetic analysis including bootstrapping of the $(\alpha/\beta)_8$-barrel sequence of the catalytic domain of Cel3A. The sequences of five subfamilies of GH family 3 were included (indicated by the encircled number). The Cel3A sequence showed highest similarity to members of subfamily 4 containing enzymes originating from aerobic fungi. The *Piromyces* β-glucosidase also associates with this group. The database entry codes are depicted in parentheses. Bootstrap values (100 trees) were all higher then 97, except for the branch indicated by an asterisk (bootstrap value 69). Scale bar, 10 substitutions/100 residues.

P75949) only contain the $(\alpha/\beta)_8$-barrel domain which harbours both catalytic residues. Since this indicates that the $(\alpha/\beta)_6$-sandwich domain of GH family 3 enzymes is not necessarily the domain of the acid/base residue, this might explain why a conserved acid/base in the $(\alpha/\beta)_6$-sandwich domain of the *Piromyces* sp. strain E2 β-glucosidase could not be deduced from the sequence alignment.

Cel3A phylogeny

Recently, phylogenetic analysis of GH family 3 sequences showed that the enzymes are grouped into six clades which correlate with their enzyme activities and origins. Homology modelling of members of each of the six clades showed that the barley structure was conserved throughout GH family 3 [37]. To investigate the phylogenetic relationship of the *Piromyces* sp. strain E2 Cel3A, the sequence encompassing the N-terminal domain was aligned to selected members of the different clades, including enzymes from cellulosome-producing bacteria. For simplicity, sequences from clade 6 were not included since these enzymes are most distant. The C-terminal domain of Cel3A was not included because of the low sequence conservation in this region. The alignment was used to calculate a phylogenetic tree including bootstrapping using the Tree Top cluster algorithm (Figure 3). The resulting tree corresponds to the phylogeny described by Harvey et al. [37] and the selected sequences nicely distributed into separate clades. The *Piromyces* sequence associated with sequences originating from aerobic fungi (clade 4) and not with the β-glucosidases from cellulosome-producing bacteria. This suggests that the *Piromyces* Cel3A gene was not obtained by lateral gene transfer from bacteria. The majority of genes encoding cellulosome components from anaerobic fungi are thought to be obtained by horizontal gene transfer [38] and so far only Cel45A from *Piromyces equi* appeared to originate from a eukaryote [11]. The association of the *Piromyces* Cel3A with group 4, which contains neither β-xylosidases nor *N*-acetyl-β-glucosaminidases, further suggests that the *Piromyces* Cel3A protein is a β-glucosidase.

Auxiliary domains of Cel3A

Further analysis of the domain organization of Cel3A (Pfam site, Sanger Institute) showed the presence of an auxiliary domain C-terminally flanking the catalytic domain (Pfam accession no. PB000239), of unknown function. This amino acid stretch, running from residues 620–716 in Cel3A (Figure 1), has been identified only in other GH family 3 proteins. The auxiliary domain has been identified in 61 out of 110 sequences in the Pfam database that contain both the $(\alpha/\beta)_6$-sandwich and the $(\alpha/\beta)_8$-barrel domain. The PB000239 auxiliary domain has been identi-

© 2003 Biochemical Society

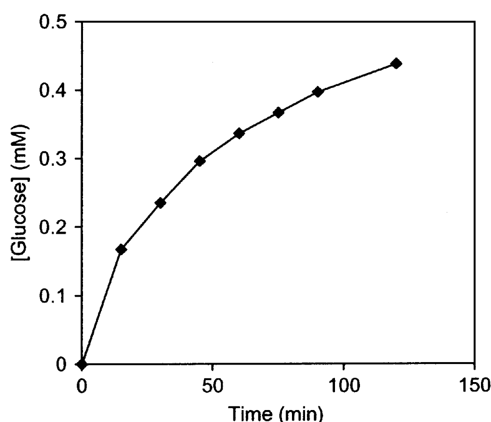

Figure 4 Degradation of cellulose by rCel3A

Time course of the hydrolysis of a 5% (w/v) Avicel (cellulose) suspension by 13.5 μg/ml of rCel3A. The release of glucose (◆) was followed with time.

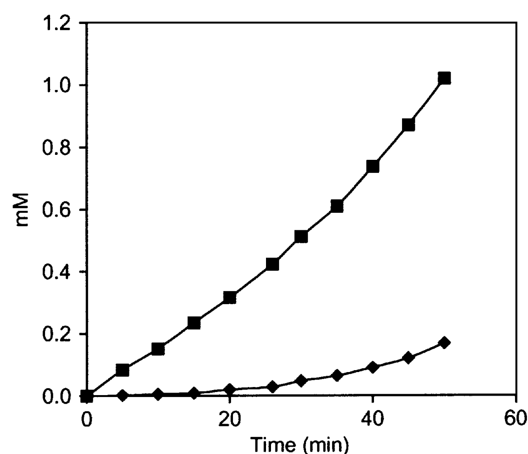

Figure 5 Hydrolysis of pNP-β-D-cellobiose by rCel3A

Time course of the hydrolysis of pNP-β-D-cellobiose by rCel3A (triplicate). The release of glucose (■) and *p*-nitrophenol (◆) was followed over time. The initial specific production activities were calculated from the values at $t = 0$ and 25 min: 3.8 units/mg and 0.1 unit/mg for the release of glucose and *p*-nitrophenol respectively.

fied exclusively in the β-glucosidases of microbial origin, except for one *Arabidopsis thaliana* β-glucosidase (AAF17692). The auxiliary domain therefore seems to be associated with β-glucosidases that are involved in generating a substrate for growth. In the 'reversed' β-glucosidases of *B. fibrisolvens* and *R. albus* (see above), the auxiliary domain is located between the sandwich and barrel domains.

The GH family 3 catalytic domain and auxiliary domain are followed by a 25 residue linker region, rich in serine and asparagine residues. The C-terminal end of the Cel3A protein consists of three tandem repeats of the cysteine-rich fungal docking domain (Figure 1). Fungal docking domains have been shown to function as the eukaryotic analogue of the bacterial dockerins and putative scaffoldins have been detected for cellulosomes from *Piromyces*, *Neocallimastix* and *Orpinomyces* species [39–41]. Thus far, most fungal cellulosome components contain two fungal dockerins, and only a limited number of cellulosome components containing three consecutive dockerins have been identified: Cel45A [11] and ManA [39] from *Piromyces equi* and *Piromyces* sp. strain E2 Cel9A [12]. The possible consequences of the number of fungal dockerins in the components for cellulosome assembly remain to be elucidated.

Biochemical analyses of purified rCel3A

To investigate the biochemical characteristics of the Cel3A protein, a truncated form, rCel3A, containing the catalytic and auxiliary domain, was heterologously produced in *E. coli*. SDS/PAGE analysis of the purified rCel3A protein indicated 90% purity (results not shown) and this preparation was used for further biochemical analyses. The pH and temperature optima of rCel3A were found at 6 and 40 °C respectively. To investigate substrate specificity, a number of pNP-linked substrates were tested. The rCel3A protein was active against pNP-β-D-glucopyranoside and pNP-β-D-cellobiose (approx. 10% of pNP-β-D-glucopyranoside activity, not linear), but not against pNP-β-galactopyranoside, pNP-β-fucopyranoside or pNP-α-D-glucopyranoside and poorly against pNP-β-D-xylopyranoside (1%) in comparison with the activity for pNP-β-D-glucopyranoside. This indicated that rCel3A has a narrow substrate specificity. The activity of rCel3A against pNP-β-D-glucopyranoside and cellobiose was investigated in detail. With pNP-β-D-glucopyranoside as a substrate, an apparent K_m of 0.6 mM and an apparent V_{max} of 67 units/mg were determined. With cellobiose as the substrate, an apparent K_m and V_{max} of 1.8 mM and 25 units/mg were determined respectively.

The rCel3A protein was also tested for activity against cellulose (Figure 4). rCel3A liberated glucose at an apparent initial activity of 0.9 unit/mg, which decreased to 0.04 unit/mg after 2 h of the incubation. Overnight incubation showed that rCel3A was able to hydrolyse approx. 0.5% of the total amount of cellulose to glucose, under the conditions tested. Presumably, rCel3A is able to liberate glucose from microfibril ends only. Because of the liberation of glucose during hydrolysis of cellulose, rCel3A was analysed for exoglucohydrolase activity. This was investigated by testing whether the enzyme preferred to hydrolyse pNP-β-D-cellobiose to glucose and pNP-β-D-glucopyranoside or if rCel3A was able to hydrolyse pNP-β-D-cellobiose to cellobiose and *p*-nitrophenol directly. The initial rates of glucose and *p*-nitrophenol production during hydrolysis of pNP-β-D-cellobiose were followed and compared (Figure 5). This showed that rCel3A had an initial glucose production activity of 3.8 units/mg compared with an apparent *p*-nitrophenol production activity of 0.1 unit/mg. These results indicated that rCel3A preferred the release of glucose from the cellobiose moiety of pNP-β-D-cellobiose instead of releasing cellobiose directly. During the incubation both rates increased, quite probably because of the increasing concentration of the product pNP-β-D-glucopyranoside for which rCel3A has a high-specific activity (67 units/mg). rCel3A can therefore be regarded as an exo-glucohydrolase, releasing subsequent glucose residues from the substrates tested.

Cel3A activity in the cellulosome

Previous experiments on the hydrolysis of cellulose by the cellulosome showed that cellobiose accumulated when initial glucose (60 mM) was present [1]. These results could be due to product inhibition of the β-glucosidase component of the complex. To assess the role of Cel3A in the cellulosome of *Piromyces* sp. strain E2, the kinetic values of the purified complex for pNP-β-D-glucopyranoside were determined and compared with the kinetic data of rCel3A. The apparent K_m value of the fungal cellulosome for pNP-β-D-glucopyranoside was 0.7 mM, which is

© 2003 Biochemical Society

almost identical with the value obtained for rCel3A (0.6 mM). The apparent V_{max} of the complex for the same substrate was 1 unit/mg, in comparison with 67 units/mg for the heterologous protein. Also the sensitivities for inhibition by glucose and by the specific β-glucosidase inhibitor D-glucono-1,5-δ-lactone were identical. Using pNP-β-D-glucopyranoside as the substrate both the cellulosome and the rCel3A protein lost 50 % of their activity at 30 mM glucose or 0.1 mM D-glucono-1,5-δ-lactone respectively. These results strongly indicate that Cel3A is responsible for the β-glucosidase activity in the cellulosome of *Piromyces* sp. strain E2.

These results were also used to estimate the number of Cel3A proteins per cellulosome. Based on the V_{max} of the complex for pNP-β-D-glucopyranoside, approx. 1.5 % of total cellulosome protein consists of Cel3A (corrected for the impurity of the rCel3A preparation and the lower molecular mass in comparison with full-length Cel3A). With an average mass of 1.3 MDa of the *Piromyces* sp. strain E2 cellulosome, a ratio of Cel3A per cellulosome of approx. 0.22 can be calculated. This indicates that Cel3A is, quantitatively, a minor component in the *Piromyces* sp. strain E2 cellulosome.

To estimate the contribution of Cel3A to glucose production during cellulose degradation by the cellulosome, earlier data were used. Dijkerman et al. [17] showed a linear correlation between the amounts of Avicelase, CMCase and β-glucosidase activities of the *Piromyces* sp. strain E2 cellulosome binding to cellulose. The ratio between the Avicelase (glucose production rate with cellulose as a substrate) and β-glucosidase (pNP-β-D-glucopyranoside hydrolysis rate) activity were recalculated. Dijkerman et al. [17] measured cellulosomal Avicelase and β-glucosidase activities of 16.8 and 33.5 units binding to cellulose respectively under the conditions applied. Recalculating the β-glucosidase activity value (calculated as activity against pNP-β-D-glucopyranoside) to the apparent V_{max} for cellobiose consumption yields 7 units of cellobiose-consuming activity. This assumes, however, that the concentration of cellobiose during cellulose degradation, immediately in the vicinity of the complex, exceeds the apparent K_m value of Cel3A for cellobiose. The Avicelase activity, which is measured as μmol of glucose produced from cellulose per min, is considered to be twice the cellobiose-forming activity from cellulose. Therefore the recalculated value for the cellobiose-producing activity of the cellulosomal cellulases yields an activity of 8.4 units. This means that the ratio of cellobiose consumption activity to cellobiose production activity during cellulose hydrolysis is 0.83, which indicates that Cel3A is responsible for the major part of the glucose-producing activity of the complex during cellulose degradation. This is corroborated further by the recent identification and characterization of the two major exoglucanases of the *Piromyces* sp. strain E2 cellulosome. The dominant 55 and 80 kDa proteins of the cellulosome were identified as a GH family 6 (H. R. Harhangi, A. C. J. Freelove, W. Ubhayasekera, M. van Dinther, P. J. M. Steenbakkers, A. Akhmanova, C. van der Drift, M. S. M. Jetten, S. L. Mowbray, H. J. Gilbert & H. J. M. Op den Camp, unpublished work) and a GH family 48 cellulase [13] respectively. Based on homology modelling and the high sequence homology of the catalytic sites to characterized members of the same family, these proteins are predicted to produce cellobiose and not glucose during hydrolysis of cellulose.

In summary, the genetic and biochemical results show that Cel3A is the first β-glucosidase identified from a cellulosome. Furthermore, quantitatively, Cel3A is a minor component in the fungal complex, and Cel3A β-glucosidase activity is responsible for the glucose-producing activity of the complex during cellulose hydrolysis.

REFERENCES

1 Dijkerman, R., Op den Camp, H. J. M., van der Drift, C. and Vogels, G. D. (1997) The role of the cellulolytic high molecular mass (HMM) complex of the anaerobic fungus *Piromyces* sp. strain E2 in the hydrolysis of microcrystalline cellulose. Arch. Microbiol. **167**, 137–142

2 Wood, T. M. and Wilson, C. A. (1995) Studies on the capacity of the cellulase of the anaerobic rumen fungus *Piromonas communis* P to degrade hydrogen bond-ordered cellulose. Appl. Microbiol. Biotechnol. **43**, 572–578

3 Wilson, C. A. and Wood, T. M. (1992) The anaerobic fungus *Neocallimastix frontalis*: isolation and properties of a cellulosome-type enzyme fraction with the capacity to solubilize hydrogen-bond-ordered cellulose. Appl. Microbiol. Biotechnol. **37**, 125–129

4 Doi, R. H. and Tamaru, Y. (2001) The *Clostridium cellulovorans* cellulosome: an enzyme complex with plant cell wall degrading activity. Chem. Rec. **1**, 24–32

5 Schwarz, W. H. (2001) The cellulosome and cellulose degradation by anaerobic bacteria. Appl. Microbiol. Biotechnol. **56**, 634–649

6 Shoham, Y., Lamed, R. and Bayer, E. A. (1999) The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microbiol. **7**, 275–281

7 Raghothama, S., Eberhardt, R. Y., Simpson, P., Wigelsworth, D., White, P., Hazlewood, G. P., Nagy, T., Gilbert, H. J. and Williamson, M. P. (2001) Characterization of a cellulosome dockerin domain from the anaerobic fungus *Piromyces equi*. Nat. Struct. Biol. **8**, 775–778

8 Bayer, E. A., Chanzy, H., Lamed, R. and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes. Curr. Opin. Struct. Biol. **8**, 548–557

9 Henrissat, B. and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. **7**, 637–644

10 Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. **316**, 695–696

11 Eberhardt, R. Y., Gilbert, H. J. and Hazlewood, G. P. (2000) Primary sequence and enzymic properties of two modular endoglucanases, Cel5A and Cel45A, from the anaerobic fungus *Piromyces equi*. Microbiology **146**, 1999–2008

12 Steenbakkers, P. J. M., Ubhayasekera, W., Goossen, H. J., van Lierop, E. M., van der Drift, C., Vogels, G. D., Mowbray, S. L. and Op den Camp, H. J. M. (2002) An intron-containing glycoside hydrolase family 9 cellulase gene encodes the dominant 90 kDa component of the cellulosome of the anaerobic fungus *Piromyces* sp. strain E2. Biochem. J. **365**, 193–204

13 Steenbakkers, P. J. M., Freelove, A., van Cranenbroek, B., Sweegers, B. M. C., Harhangi, H. R., Vogels, G. D., Hazlewood, G. P., Gilbert, H. J. and Op den Camp, H. J. M. (2003) The major component of the cellulosomes of anaerobic fungi from the genus *Piromyces* is a family 48 glycosyl hydrolase. DNA sequence, in the press

14 Li, X. L., Chen, H. and Ljungdahl, L. G. (1997) Two cellulases, CelA and CelC, from the polycentric anaerobic fungus *Orpinomyces* strain PC-2 contain N-terminal docking domains for a cellulase–hemicellulase complex. Appl. Environ. Microbiol. **63**, 4721–4728

15 Teunissen, M. J., de Kort, G. V., Op den Camp, H. J. M. and Huis in 't Veld, J. H. J (1992) Production of cellulolytic and xylanolytic enzymes during growth of the anaerobic fungus *Piromyces* sp. on different substrates. J. Gen. Microbiol. **138**, 1657–1664

16 Harhangi, H. R., Steenbakkers, P. J. M., Akhmanova, A., Jetten, M. S. M., van der Drift, C. and Op den Camp, H. J. M. (2002) A highly expressed family 1 β-glucosidase with transglycosylation capacity from the anaerobic fungus *Piromyces* sp. E2. Biochim. Biophys. Acta **1574**, 293–303

17 Dijkerman, R., Vervuren, M. B., Op den Camp, H. J. M. and van der Drift, C. (1996) Adsorption characteristics of cellulolytic enzymes from the anaerobic fungus *Piromyces* sp. strain E2 on microcrystalline cellulose. Appl. Environ. Microbiol. **62**, 20–25

18 Akhmanova, A., Voncken, F. G. J., Harhangi, H. R., Hosea, K. M., Vogels, G. D. and Hackstein, J. H. P. (1998) Cytosolic enzymes with a mitochondrial ancestry from the anaerobic chytrid *Piromyces* sp. E2. Mol. Microbiol. **30**, 1017–1027

19 Ohyama, T., Tagashira, H. and Tsujibayashi, H. (1997) Easy screening of recombinant plasmids. Tech. Tips Online **1**, T40073

20 Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. and Wheeler, D. L. (2000) GenBank. Nucleic Acids Res. **28**, 15–18

21 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. **28**, 235–242

22 Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402

23 Nielsen, H., Brunak, S. and von Heijne, G. (1999) Machine learning approaches to the prediction of signal peptides and other protein sorting signals. Protein Eng. **12**, 3–9

© 2003 Biochemical Society

24 Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive, multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673–4680

25 Brodskii, L. I., Ivanov, V. V., Kalaidzidiz, Y. L., Leontovich, A. M., Nikolaev, V. K., Feranchuk, S. L. and Drachev, V. A. (1995) GeneBee-NET: an internet based server for biopolymer structure analysis. Biokhimiia **60**, 1221–1230

26 Dijkerman, R., Bhansing, D. C. P., Op den Camp, H. J. M., van der Drift, C. and Vogels, G. D. (1997) Degradation of structural polysaccharides by the plant cell-wall degrading enzyme system from anaerobic fungi: an application study. Enzyme Microb. Technol. **21**, 130–136

27 Kruus, K., Andreacchi, A., Wang, W. K. and Wu, J. H. D. (1995) Product inhibition of the recombinant CelS, an exoglucanase component of the *Clostridium thermocellum* cellulosome. Appl. Microbiol. Biotechnol. **44**, 399–404

28 Tuohy, M. G., Walsh, D. J., Murray, P. G., Claeyssens, M., Cuffe, M. M., Savage, A. V. and Coughlan, M. P. (2002) Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. Biochim. Biophys. Acta **1596**, 366–380

29 Brownlee, A. G. (1994) The nucleic acids of anaerobic fungi. In Anaerobic Fungi (Mountford, D.O. and Orpin, C.G., eds.), pp. 241–256, Marcel Dekker Inc, New York

30 von Heijne, G. (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. **133**, 17–21

31 Varghese, J. N., Hrmova, M. and Fincher, G. B. (1999) Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase. Structure **7**, 179–190

32 Li, B. and Renganathan, V. (1998) Gene cloning and characterization of a novel cellulose-binding β-glucosidase from *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. **64**, 2748–2754

33 Rye, C. S. and Withers, S. G. (2000) Glycosidase mechanisms. Curr. Opin. Chem. Biol. **4**, 573–580

34 Dan, S., Marton, I., Dekel, M., Bravdo, B. A., He, S., Withers, S. G. and Shoseyov, O. (2000) Cloning, expression, characterization, and nucleophile identification of family 3, *Aspergillus niger* β-glucosidase. J. Biol. Chem. **275**, 4973–4980

35 Chir, J., Withers, S., Wan, C.-F. and Li, Y.-K. (2002) Identification of the two essential groups in the family 3 β-glucosidase from *Flavobacterium meningosepticum* by labelling and tandem mass spectrometric analysis. Biochem. J. **365**, 857–863

36 Li, Y.-K., Chir, J., Tanaka, S. and Chen, F.-Y. (2002) Identification of the general acid/base catalyst of a family 3 β-glucosidase from *Flavobacterium meningosepticum*. Biochemistry **41**, 2751–2759

37 Harvey, A. J., Hrmova, M., De Gori, R., Varghese, J. N. and Fincher, G. B. (2000) Comparative modelling of the three-dimensional structures of family 3 glycoside hydrolases. Proteins **41**, 257–269

38 Zhou, L. Q., Xue, G. P., Orpin, C. G., Black, G. W., Gilbert, H. J. and Hazlewood, G. P. (1994) Intronless Celb from the anaerobic fungus *Neocallimastix patriciarum* encodes a modular family a endoglucanase. Biochem. J. **297**, 359–364

39 Fanutti, C., Ponyi, T., Black, G. W., Hazlewood, G. P. and Gilbert, H. J. (1995) The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. J. Biol. Chem. **270**, 29314–29322

40 Fillingham, I. J., Kroon, P. A., Williamson, G., Gilbert, H. J. and Hazlewood, G. P. (1999) A modular cinnamoyl ester hydrolase from the anaerobic fungus *Piromyces equi* acts synergistically with xylanase and is part of a multiprotein cellulose-binding cellulase–hemicellulase complex. Biochem. J. **343**, 215–224

41 Steenbakkers, P. J. M., Li, X. L., Ximenes, E. A., Arts, J. G., Chen, H., Ljungdahl, L. G. and Op den Camp, H. J. M. (2001) Noncatalytic docking domains of cellulosomes of anaerobic fungi. J. Bacteriol. **183**, 5325–5333

Received 12 November 2002/13 December 2002; accepted 16 December 2002
Published as BJ Immediate Publication 16 December 2002, DOI 10.1042/BJ20021767

© 2003 Biochemical Society