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

Peter J. M. STEENBAKKERS*, Wimal UBHAYASEKERA†, Harry J. A. M. GOOSSEN*, Erik M. H. M. VAN LIEROP*, Chris VAN DER DRIFT*, Godfried D. VOGELS*, Sherry L. MOWBRAY† 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, and †Department of Molecular Biology, Swedish University of Agricultural Sciences, Biomedical Centre, P.O. Box 590, SE-751 24 Uppsala, Sweden

The cellulosome produced by *Piromyces* sp. strain E2 during growth on filter paper was purified by using an optimized cellulose-affinity method consisting of steps of EDTA washing of the cellulose-bound protein followed by elution with water. Three dominant proteins were identified in the cellulosome preparation, with molecular masses of 55, 80 and 90 kDa. Treatment of cellulose-bound cellulosome with a number of denaturing agents was also tested. Incubation with 0.5 % (w/v) SDS or 8 M urea released most cellulosomal proteins, while leaving the greater fraction of the 80, 90 and 170 kDa components. To investigate the major 90 kDa cellulosome protein further, the corresponding gene, *cel9A*, was isolated, using immunoscreening and N-terminal sequencing. Inspection of the *cel9A* genomic organization revealed the presence of four introns, allowing the construction of a consensus for introns in anaerobic fungi. The 2800 bp cDNA clone contained an open reading frame of 2334 bp encoding a 757-residue extracellular protein. Cel9A includes a 445-residue glycoside hydrolase family 9 catalytic domain, and so is the first fungal representative of this large family. Both modelling of the catalytic domain as well as the activity measured with low level expression in *Escherichia coli* indicated that Cel9A is an endoglucanase. The catalytic domain is succeeded by a putative β-sheet module of 160 amino acids with unknown function, followed by a threonine-rich linker and three fungal docking domains. Homology modelling of the Cel9A dockerins suggested that the cysteine residues present are all involved in disulphide bridges. The results presented here are used to discuss evolution of glycoside hydrolase family 9 enzymes.

Key words: chytrid, endoglucanase, evolution, cellulose-binding, homology modelling.

INTRODUCTION

Obligately anaerobic fungi are best known for their high (hemi)cellulolytic capacity [1]. They can easily be isolated from the digestive tract of herbivores, where they play an important role in the degradation and fermentation of plant cell-wall material [2]. For the degradation of crystalline cellulose the anaerobic fungus *Piromyces* sp. strain E2 is dependent on the activities of a high-molecular-mass (hemi)cellulolytic complex [3,4] that is similar to the cellulosome produced by *Clostridium thermocellum* and other anaerobic bacteria [5,6]. A key element of the bacterial cellulosome is a central, cellulose-binding scaffoldin protein that is able to interact with the catalytic components through its cohesin sites in a calcium-dependent manner. For the interaction with the cohesin sites, the catalytic components require the presence of a bacterial dockerin domain. Similarly, the fungal cellulosome consists of catalytic components, which contain at least one copy of a fungal dockerin. Although the presence of putative fungal scaffoldins has been shown for *Piromyces*, *Neocallimastix* and *Orpinomyces* species [7–9], the corresponding genes have not yet been described.

The cellulases thus far described for the *C. thermocellum* cellulosome represent members of five glycoside hydrolase (GH) families of which the GH family 48 and the GH family 9 enzymes are the most dominant [10]. These two families of GHs have been shown to degrade cellulose in a synergistic manner [11]. Experiments on the heterologously expressed GH family 48 enzyme from *C. thermocellum* indicated that this component alone accounted for most of the biochemical properties of the complete cellulosome, indicating a pivotal role for this cellulase [12]. The second prominent cellulase in the *C. thermocellum* cellulosome is the GH family 9 endoglucanase, CelD [13]. In addition to CelD, four other *C. thermocellum* cellulosome components carry a GH family 9 catalytic domain [5]. In the *Acetivibrio cellulolyticus* cellulosome, a GH family 9 catalytic domain is even fused to the organizing scaffoldin protein, an observation that is indicative of an important role for this GH family in cellulosome activity [14]. The GH family 9 catalytic domain, however, is not confined to cellulosomes. Over 200 GH family 9 encoding sequences have been deposited in databases originating from insects, plants, the cellular slime mould *Dictyostelium discoideum* and bacteria. However, no family 9 enzymes from fungi have previously been described.

Until now, genes encoding cellulases from only GH families 5, 6 and 45 have been reported to be part of the cellulosome of anaerobic fungi [15,16]. All contain a catalytic domain and one

Abbreviations used: BLAST, basic local alignment search tool; CBM, cellulose-binding module; CMC, carboxymethylcellulose; DTT, dithiothreitol; GH, glycoside hydrolase; IPTG, isopropyl β-D-thiogalactopyranoside; LB, Luria–Bertani; ORF, open reading frame; UTR, untranslated region.

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

The nucleotide sequences included in this manuscript have been submitted to the GenBank®, EMBL, DDBJ and GSDB Nucleotide Sequence Databases under the accession numbers AF459452 and AF459453.

or more fungal docking domains, but the presence of auxiliary domains has not been reported. For the family 5 and 6 GH-encoding genes, a bacterial ancestry has been suspected, because of the high sequence similarity to bacterial genes and the absence of introns [17,18]. However, introns are apparently rare in anaerobic fungi; only one intron-containing gene has been proposed [19], although not yet experimentally verified.

Genes encoding components of the cellulosome of anaerobic fungi have previously been identified through activity screening of cDNA libraries. The presence of fungal docking domains is generally viewed as an indication of involvement with the fungal cellulosome. To identify and characterize the key enzymes in the fungal cellulosome, our approach has been directed instead against the major cellulosome components specifically. In the paper we report the purification and dissociation of the *Piromyces* sp. strain E2 cellulosome and the identification of its major components. Furthermore, we report the gene encoding the dominant 90 kDa protein, and describe the expected protein product.

EXPERIMENTAL

Strain and culture conditions

Piromyces sp. strain E2 (A.T.C.C. 76762), isolated from the faeces of an Indian elephant (*Elephas maximus*), was maintained anaerobically in M2 medium with 0.5 % (w/v) barley straw as a carbon source [20]. For cellulosome purification, *Piromyces* sp. strain E2 was transferred twice to fresh M2 medium with 0.5 % (w/v) Whatman no. 1 filter paper as a carbon source before preparative culturing. Media were inoculated (1 %, v/v) from exponentially growing precultures.

Purification of the *Piromyces* sp. strain E2 cellulosome

Piromyces sp. strain E2 was grown for 4 days and biomass was removed by passing the culture through a nylon filter. The remaining culture fluid was concentrated approx. 100-fold by ultrafiltration with a Minitan cassette (30 kDa cut-off; Millipore Corp., Bedford, MA, U.S.A.) and dialysed against 10 mM Mes buffer, pH 6.5, containing 100 mM NaCl. The cellulosome was purified from the concentrated extracellular protein either by gel filtration (method A) or by a cellulose-affinity procedure using Avicel (microcrystalline cellulose, type PH 105; Serva, Heidelberg, Germany) (method B). In method (A), the cellulosome was purified using gel filtration as described by Dijkerman et al. [3]. Concentrated extracellular protein (40 mg of protein/ml) was separated on a S-300 Sephacryl HR column (2.5 cm × 95 cm; Pharmacia LKB Biotechnology, Uppsala, Sweden) at a flow rate of 0.5 ml/min using 10 mM Mes (pH 6.5)/100 mM NaCl as eluent. The high-molecular-mass fractions were pooled and constituted the gel-filtration-purified cellulosome preparation. In method (B), an Avicel suspension was added to the concentrated extracellular protein to a final concentration of 4 % (w/v) and 100 mM KH_2PO_4/K_2HPO_4, pH 6.5 (P-buffer). Protein was bound to the Avicel by stirring the suspension for 1 h at 4 °C. Avicel-bound protein was collected by centrifugation. The Avicel pellet was washed three times by resuspending the pellet in 50 ml of 100 mM P-buffer or 100 mM sodium EDTA/HCl buffer, pH 7 (E-buffer) and subsequent centrifugation (5000 ***g*** for 10 min at 4 °C), discarding the supernatant. The pellet was subjected to a final washing step with 50 ml of 25 mM P-buffer or E-buffer. The cellulosome was eluted by resuspending the pellet in 1.5 ml of deionized water and rotating it for 5 min at room temperature followed by centrifugation (10000 ***g*** for 2 min at room temperature). This was repeated ten times, after which the supernatant was free of protein. The remaining pellet no longer contained protein. The protein-containing elution fractions were pooled and constituted the cellulose-purified cellulosome.

Cellulosomal protein (50 μg) was applied to a standard Laemmli SDS/7.5 %-(w/v)-polyacrylamide gel, using the Protean II or mini-Protean II system (Bio-Rad Laboratories, Richmond, CA, U.S.A.). To identify the dominant cellulosome proteins, the gel was stained with Coomassie Brilliant Blue R-250. Glycosylated cellulosomal proteins were stained using the periodic acid/silver staining method [21]. α-Galactose-specific glycosylation of purified cellulosomal protein from *Piromyces* sp. strain E2 was investigated as described previously [22], with *Clostridium thermocellum* cellulosome protein serving as a positive control.

Dissociation of cellulose-bound cellulosome

A 1 ml (1 mg of protein/ml) portion of gel filtration-purified cellulosome (molecular mass 1200–1400 kDa) was incubated with 200 μl of 20 % (w/v) microcrystalline cellulose (Avicel) for 1 h at 4 °C with continuous mixing. The suspension was centrifuged and the pellet washed once with 1.5 ml of deionized water. This washing step did not liberate protein and was used to lower the ionic strength of the preparation. The washed pellet was resuspended in 1 ml of deionized water and dispensed as 100 μl aliquots. Next, the following reagents were added individually and in all combinations: (final concns.) 25 mM EDTA/HCl, pH 7, 20 mM dithiothreitol (DTT) and 0.5 % (w/v) SDS. Other conditions tested were 0.05–5 % (w/v) SDS, 0.5 % (w/v) CHAPS, 0.5 % (v/v) Tween-20, 0.5 % (v/v) Tween-80, 0.5 % (v/v) Triton X-100, 0.5 % (v/v), Nonidet P-40 and 8 M urea, as well as P- and E-buffer in concentrations varying from 1 to 250 mM. Untreated cellulose-bound cellulosome, the cellulose-treated gel filtration-purified cellulosome preparation and a deionized water incubation were included as controls. After mixing, all samples were incubated for 30 min at 39°C. The mixtures were cooled on ice, centrifuged and the supernatant was collected. The pellet was resuspended in 100 μl of deionized water and 20 μl of 5 × SDS sample buffer [5 × SDS sample buffer is 65 mM Tris (pH 6.8)/10 % (v/v) glycerol/2 % (w/v) SDS/5 % (v/v) β-mercaptoethanol/12.5 μg/ml Bromophenol Blue] was added to the cellulose suspensions and the supernatants. Pellet and supernatant samples were boiled for 5 min, centrifuged, and the supernatants applied to a standard Laemmli SDS/7 % polyacrylamide gel.

N-terminal microsequencing of the 90 kDa dominant cellulosome protein

Extracellular protein was collected and concentrated as described above. A 10 ml portion of the concentrated culture filtrate (40 mg of protein/ml) was mixed with an equal volume of 100 mM P-buffer, pH 6.5, and 5 g of Avicel. The mixture was stirred at 4 °C for 1 h, centrifuged (5000 ***g*** for 10 min at 4 °C) and the resulting pellet was washed three times with 100 ml of 25 mM P-buffer, pH 6.5. The pellet was resuspended in 5 ml of deionized water and 2 ml of 5 × SDS sample buffer, boiled for 10 min and centrifuged (20000 ***g*** for 20 min at room temperature) and the remaining supernatant was used for gel electrophoresis. A 1 ml portion was applied to a 2.25 mm-thickness 7.5 %-polyacrylamide denaturing gel in a Bio-Rad Mini-Protean II system and run for 6 h at 20 mA. The gel was blotted for 12 h at 0.6 mA/cm², using a discontinuous buffer system [anode: 0.05 % (w/v) SDS, 190 mM glycine and 25 mM Tris buffer, pH 8.3; cathode: 20 % (v/v) methanol, 190 mM glycine and 25 mM Tris buffer, pH 8.3]

and a Bio-Rad semi-dry transfer cell. Separation resolution and transfer to the PVDF membrane was confirmed by Ponceau S staining. The 90 kDa protein was cut out and 17 pmol of protein was used to determine the N-terminal 25 amino acids by Edman degradation. Microsequencing was performed at the Sequentiecentrum Utrecht of the University of Utrecht, Utrecht, The Netherlands.

Preparation of anti-cellulosome antibodies

Culture fluid of *Piromyces* sp. strain E2 was concentrated and dialysed as described above. The cellulosome was purified using gel filtration as described by Dijkerman et al. [3]. The fractions comprising the cellulosome peak were concentrated to 0.5 mg/ml using the Amicon system equipped with an YM-10 filter. Approx. 200 μl was mixed with Freund's complete adjuvant (1:1, v/v) and used to immunize a rabbit by intracutaneous injection. The rabbit was boosted twice with the same amount of cellulosome protein mixed with Freund's incomplete adjuvant (1:1). Blood serum was collected after 4 weeks and incubated for 2 h at room temperature and overnight at 4 °C. The anti-cellulosome antibodies were collected as the supernatant after centrifugation (9000 ***g*** for 30 min at 4 °C).

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

A previously described *Piromyces* sp. strain E2 cDNA library [23] in lambda ZAP II (Stratagene, La Jolla, CA, U.S.A.) was screened and positive plaques were rescreened using the anti-cellulosome antibodies according to the picoBlue (Stratagene) immunoscreening instruction manual. Expression time was extended with an overnight incubation at room temperature. Nitrocellulose filters were blocked and incubated overnight at 4 °C with (1:200-diluted) anti-cellulosome antibodies. Immunodetection was performed with (1:5000-diluted) horseradish peroxidase-conjugated goat anti-rabbit polyclonal antibodies and diaminobenzidine staining. Immunoreactive plaques were isolated and recombinant pBlue vectors obtained by *in vivo* excision and rescued using ExAssist helper phage. Of the immunoreactive clones, 32 were sequenced using pBluescript SK vector primers.

Isolation of the genomic *cel9A* sequence

A PCR using genomic DNA from *Piromyces* sp. strain E2 as template was performed to isolate genomic DNA containing the *cel9A* coding sequence. Primers were designed using the *cel9A* cDNA sequence, which was isolated by immunoscreening. The sense primer, TCCGCTGTTGCAGCCCTGGTG, was designed to anneal to the region immediately after the putative start codon and the antisense primer, GTTCAAATTAACAACTGCTCTTAA, to anneal to the region containing the *cel9A* stop codon and a part of the 3′-untranslated region (3′-UTR). PCR was performed with Takara (Otsu, Japan) proofreading Pyrobest, 1 μg of *Piromyces* sp. strain E2 genomic DNA (isolated according to Brownlee [24]) and consisted of 40 cycles of 1 min 94 °C, 1 min 60 °C and 4 min elongation time using a Biometra T-gradient thermocycler PCR apparatus (Westburg, Leusden, The Netherlands). PCR products were cloned using the T/A cloning kit (Fermentas, Vilnius, Lithuania). Vectors containing the PCR product were identified using the rapid screening procedure [25].

Heterologous Cel9A activity staining

The *cel9A* cDNA clone isolated by immunoscreening was transformed into *Escherichia coli* BL21 (Novagen, Madison, WI, U.S.A.) and grown on standard Luria–Bertani (LB) plates containing 100 μg/ml ampicillin, 1% (w/v) carboxymethylcellulose (CMC; low viscosity), 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and 1 mM $CaCl_2$. After overnight growth at 37 °C the plates were stained with 0.1% Congo Red solution for 30 min and destained with 1 M NaCl. For CMC zymogram analysis, recombinant *E. coli* BL21 was grown overnight at 30 °C in 250 ml of LB medium containing 100 μg/ml ampicillin. After growth, cells were harvested by centrifugation and the cell pellet was washed with 25 ml of 100 mM citrate buffer, pH 6. The washed cells were resuspended in 10 ml of 100 mM citrate buffer, pH 6, containing protease inhibitors (protease inhibitor cocktail; Boehringer-Mannheim, Mannheim, Germany) and lysed using a French Press. The lysed cells were centrifuged (15000 ***g*** for 30 min at 4 °C) and the cell-free extract (35 mg of protein/ml) was collected as the supernatant. Cell-free extract was mixed (4:1, v/v) with 5 × SDS sample buffer without β-mercaptoethanol and incubated at room temperature for 1 h. A 20 μl sample was loaded on to a 1.5 mm-thick SDS/7.5%-polyacrylamide gel containing 0.2% (w/v) CMC and run at 20 mA at room temperature using a Bio-Rad Mini-Protean III system. After electrophoresis, the marker lane was cut off and stained with Coomassie Brilliant Blue R250. The activity gel was washed three times for 30 min at room temperature with 100 mM P-buffer, pH 6.5, and incubated overnight at 37 °C in the same buffer. Activity was visualized by Congo Red staining and NaCl destaining as described above.

Recombinant DNA techniques

The full-length *cel9A* sense strand was sequenced using nested deletions with Fermentas reagents. The antisense strand was sequenced with primers designed from the sense strand sequence. DNA was sequenced using the dRhodamine sequencing kit (Perkin-Elmer, 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.

Homology modelling

Similar sequences from GenBank® [26] were located and aligned using PSI-BLAST (position-specific iterative–basic local alignment search tool) searches [27], hidden Markov models [28] and MULTALIGN [29]. Secondary-structure predictions were carried out using a number of web-based tools, including 3D-PSSM [30] and programs available at http://npsa-pbil.ibcp.fr/. Appropriate protein structures were obtained from the Protein Data Bank [31] and structural comparisons were carried out using O [32] and LSQMAN [33]. Multiple alignments were used to generate the best pairwise alignment of the catalytic modules of the *Piromyces* sp. strain E2 Cel9A enzyme and *T. fusca* E4 endo/exo-1,4-glucanase (P26221). This pair-wise alignment was then the basis of creating a homology model, using the appropriate portion of the structure of the *T. fusca* enzyme (1TF4; [34]) as a template in the program SOD [35]. The model was modified somewhat in the graphics program O [32], using rotamers that would improve packing in the interior of the protein and to account for insertions and deletions in some loop regions. Similarly, homology models of the three dockerins were generated using the structure of a cellulosome dockerin module from *Piromyces equi* [36]. Models are available upon request from H. J. M. Op den Camp. Figure 7 (below) was prepared using Molscript [37], Molray [38] and Povray (http://www.povray.org/).

Sequence analysis

Cel9A nucleotide sequences were assembled using vector NTI Suite (Informax, Golden, CO, U.S.A.). Further analysis was performed with default settings using the on-line facilities of the Swiss ExPASy server (http://www.expasy.ch) and the CAZy site (http://afmb.cnrs-mrs.fr/CAZY/) [39]. Phylogenetic analysis with bootstrapping was performed using the GeneBee Phylogenetic Tree Prediction software [40] with default settings (http://www.genebee.msu.su/services/phtree-full.html) using the sequence alignment of the Pfam GH family 9 seed sequences. Accession numbers and the amino acid stretch used for the alignment are available at http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00759.

RESULTS AND DISCUSSION

Purification of the *Piromyces sp.* strain E2 cellulosome

Previous efforts to purify the cellulosome of *Piromyces* sp. strain E2 from filter-paper-grown culture filtrate did not result in the identification of any major components. The procedure included size fractionation followed by absorption to phosphoric acid-swollen cellulose. After complete solubilization of the acid-swollen cellulose by the attached cellulosome, no cellulosomal proteins with a molecular mass higher than 75 kDa could be identified [4]. To be able to identify the key components in the cellulosome, the purification required further optimization. When size-fractionated extracellular protein was bound to crystalline cellulose (Avicel) and immediately eluted by boiling the suspension with SDS/PAGE sample buffer, a denatured cellulosome preparation was obtained with proteins ranging in size from 50 to 250 kDa. Comparison of the amount of cellulosomal proteins with the amount of size-fractionated proteins that did not bind to cellulose suggested that the high-molecular-mass pool consisted of approx. 95% pure cellulosome. Moreover, cellulose absorption of total extracellular protein with crystalline cellulose (Avicel) followed by the denaturing elution step resulted in an identical cellulosomal protein pattern, suggesting that no significant amounts of non-cellulosomal proteins bound to cellulose.

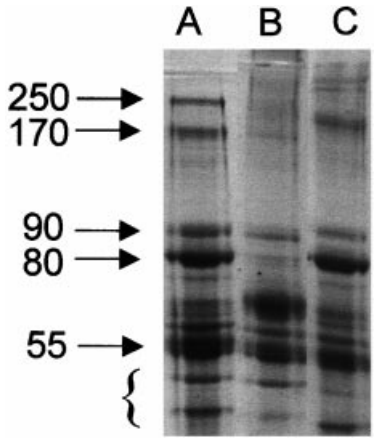

Figure 1 Dissociation/degradation of cellulose-bound cellulosome by CHAPS

Lane A, control (phosphate-buffer-washed, cellulose-bound cellulosomal proteins); lane B, cellulosomal protein that remained bound to cellulose after CHAPS treatment; lane C, cellulosomal protein that dissociated from cellulose after CHAPS treatment. The face (curly) bracket indicates degradation products in lane A as a result of the purification procedure. Sizes are indicated in kDa.

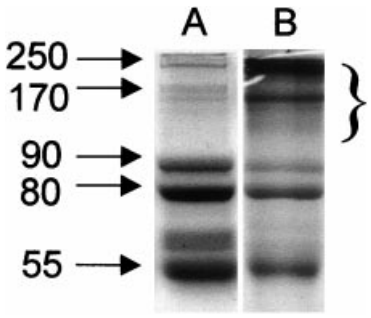

Figure 2 Coomassie and periodic acid/silver staining of EDTA-treated purified cellulosome

Lane A, Coomassie Brilliant Blue staining (50 μg); lane B, periodic acid/silver staining (20 μg). Note the absence of degradation products in comparison with Figure 1, lane A. The face (curly) bracket indicates the smear of glycosylated protein. Sizes are indicated in kDa.

Cellulose-bound cellulosome was incubated with different buffers and reagents and at different pH values. This revealed that binding of the complex to cellulose was dependent on a buffer concentration of at least 25 mM and a pH between pH 3 and 11. Incubation of the buffer-washed cellulose-bound complex with deionized water eluted the complex. Directly eluting the complex with water, rather than first washing with buffer, resulted in a preparation that did not resemble that obtained from the gel filtration-purified cellulosome. Elution fractions did not contain all cellulosomal protein from 50 to 250 kDa and showed proteins with an, on average, lower molecular mass. Interestingly, washing of the cellulose-bound protein with EDTA-containing buffer did not dissociate or degrade components of the complex as was shown for the *Clostridium thermocellum* cellulosome [41]. Apparently, the interaction of the fungal scaffoldin and dockerins does not require Ca^{2+} or other bivalent cations, which is consistent with previous results on the structure of the fungal dockerin [36]. Washing with EDTA instead of phosphate buffer prevented the appearance of small amounts of low-molecular-mass proteins, which could indicate inhibition of low levels of proteolytic activity (Figures 1A and 2A). After washing of cellulose-bound complex with EDTA/buffer, the cellulosome was eluted with deionized water with a simultaneous release of cellulose particles, as was judged from an opalescent supernatant. Subsequent cellulose absorption of the purified cellulosome indicated that the complex was not dissociated by the procedure and did not remain bound to the released cellulose particles. SDS/PAGE analysis of the EDTA-washed purified cellulosome revealed the presence of proteins ranging from 55 to 250 kDa and allowed, for the first time, the identification of the 55, 80 and 90 kDa proteins as the major components (Figure 2A).

Dissociation of cellulose-bound cellulosome

To investigate the dissociative behaviour of the cellulosome, cellulose-bound complex was treated with a number of reagents, and dissociation of the cellulosome was judged by the distribution of eluted and cellulose-bound proteins. Of the chemicals tested, Tween-20, Tween-80, Triton X-100 or Nonidet P40 did not dissociate or elute the complex from cellulose. CHAPS, however, did result in partial dissociation/degradation of the complex,

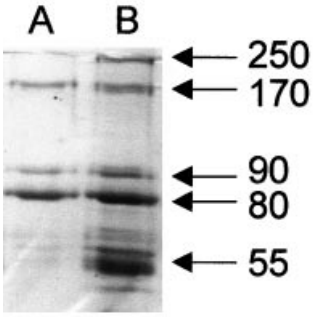

Figure 3 Dissociation of cellulose-bound cellulosome by SDS

Lane A, cellulose-bound cellulosome after mild SDS treatment; lane B, control of untreated cellulose-bound cellulosome.

proteins with a molecular mass up to 90 kDa being eluted (Figure 1). Incubation of the cellulose-bound complex with 20 mM DTT caused the release of only 50 % of the cellulose-bound protein (results not shown). With the recent publication of the first structure of a fungal dockerin, it became apparent that disulphide bridges are essential for their function. Interaction of the complex with cellulose or the association of the fungal dockerin with its native catalytic domain apparently protects it from reduction of its disulphide bridges. The disulphide bridges in the dockerins, therefore, are likely to exist in the natural environment of the cellulosome, which is the digestive tract of herbivores, a strongly reducing environment. Treatment of the complex with up to 5 % (w/v) SDS in the presence of EDTA and treatment with 8 M urea (not shown) only partially released it from cellulose. Approx. 50 % of the major 80 and 90 kDa proteins and a minor 170 kDa component remained cellulose-bound, with the 55 kDa and the 250 kDa proteins serving as internal standards (Figure 3). Apparently, the interaction of the 80, 90 kDa and 170 kDa proteins with cellulose differs from that of the other cellulosomal proteins.

Isolation of *cel9A*, encoding the 90 kDa cellulosome component

To investigate the identity of the 90 kDa dominant fungal cellulosome component, the N-terminal amino acid sequence was determined by Edman degradation: KSQDYARHIELSL-LFYEAQRSGKLP. The N-terminal sequence was compared with database sequences and showed similarity to the N-terminal regions of family 9 GHs. Because the 90 kDa protein is one of the dominant proteins in the cellulosome preparation, a relatively high expression level was expected. Therefore, immunoscreening using anti-cellulosome antibodies was chosen to isolate the corresponding gene. Approx. 60000 plaques from the *Piromyces* sp. strain E2 cDNA library were screened using anti-cellulosome antibodies. A total of 50 plaques positive in the primary screening were rescreened and the plasmids from 32 positive plaques were isolated and sequenced. The sequences determined were compared with database sequences and with the previously determined N-terminus of the 90 kDa cellulosome component. One 2.8 kb cDNA clone exactly matched the N-terminus and was designated *cel9A*, and the encoded protein, Cel9A, according to the GH nomenclature of Henrissat and Davies [39]. Five incomplete cDNA clones encoding family 9 GHs identical with Cel9A were also identified.

cDNA sequence of the *cel9A* gene

The 2.8 kb cDNA clone contained an open reading frame (ORF) of 2334 nucleotides with a GC content of 40.2 %. The ORF was flanked by a 83-nt upstream region and a 387-nt downstream region (Figure 4) with an average AT content of 88 %, consistent with the low GC content of non-coding regions in genes from anaerobic fungi [24]. Both up- and down-stream regions contained multiple stop codons in all three reading frames. The ATG start codon was preceded by the nucleotide stretch CAAA also present in other genes from anaerobic fungi. The ORF encoded a 778-residue protein in which the previously determined N-terminus of the 90 kDa dominant cellulosome protein was found, starting at amino acid position 22. These results provided evidence that the *cel9A* coding region was complete. The experimentally determined N-terminus revealed the presence of a 21-amino-acid signal peptide, MKFQSIISAVAALVAPMAV-GA, consistent with the extracellular location of the cellulosome. It contained the typical features of charged/hydrophilic amino acids at the N-terminal end, followed by a hydrophobic amino acid stretch. Also, the residues near the cleavage site followed the (-1, -3) rule which states that residues at these positions should be small and hydrophobic [42]. The resulting mature Cel9A protein of 757 amino acids has a calculated pI of 5.1 and molecular mass of 83811 Da, slightly less than the 90 kDa estimated by SDS/PAGE. To determine whether the 90 kDa protein was glycosylated, the SDS/PAGE-separated complex was subjected to periodic acid/silver staining. This revealed the presence of glycosylated proteins as a smear ranging from an apparent molecular mass from 150 kDa to 250 kDa, with a distinct staining of a 170 kDa protein (Figure 2B). Apparently the 90 kDa protein was not glycosylated. To investigate whether the glycosylation in the *Piromyces* sp. strain E2 complex was similar to that observed for the bacterial scaffoldin protein, the fungal cellulosome was separated by SDS/PAGE and blotted together with the *Clostridium thermocellum* cellulosome [22]. Staining of the blotted proteins with an α-galactose-specific lectin only stained the clostridial scaffoldin, but not the fungal cellulosome, indicating that a different type of glycosylation is present in the fungal cellulosome.

Chromosomal organization of *cel9A*

To investigate the chromosomal organization of the *cel9A* gene, the corresponding genomic clone was isolated by PCR using a proofreading polymerase. Primers were designed to anneal to the sequences encoding the Cel9A signal peptide and the stop codon. The AT-rich, non-coding regions were not used in the design of the primers because of an anticipated lower primer specificity in these regions. A 2746 bp genomic *cel9A* PCR product was isolated and 2323 bp could be aligned to the *cel9A* cDNA sequence. The coding region of the genomic *cel*9A clone was interrupted by four introns of 127, 93, 90 and 113 nucleotides (intron I–IV; Figure 4) with an average AT content of 85 %, rendering five exons of 175, 21, 172, 49 and 1906 nt. Only intron IV was not situated inside a codon. All introns identified were located in the 5′-end of the coding region, as was also observed for the introns present in the *cel9A* clone isolated from the Australian red-claw crayfish *Cherax quadricarinatus* [43]. The sizes of the introns corresponded well to the relatively small size of approx. 100 nt of introns from aerobic filamentous fungi [44].

Thus far, only a single intron has been described for anaerobic fungi. The inducible gene, *enol*, encoding a putative enolase from the anaerobic fungus *Neocallimastix frontalis* [19] was shown to contain a 331-nt, AT-rich (86 %) intron. However, the presence

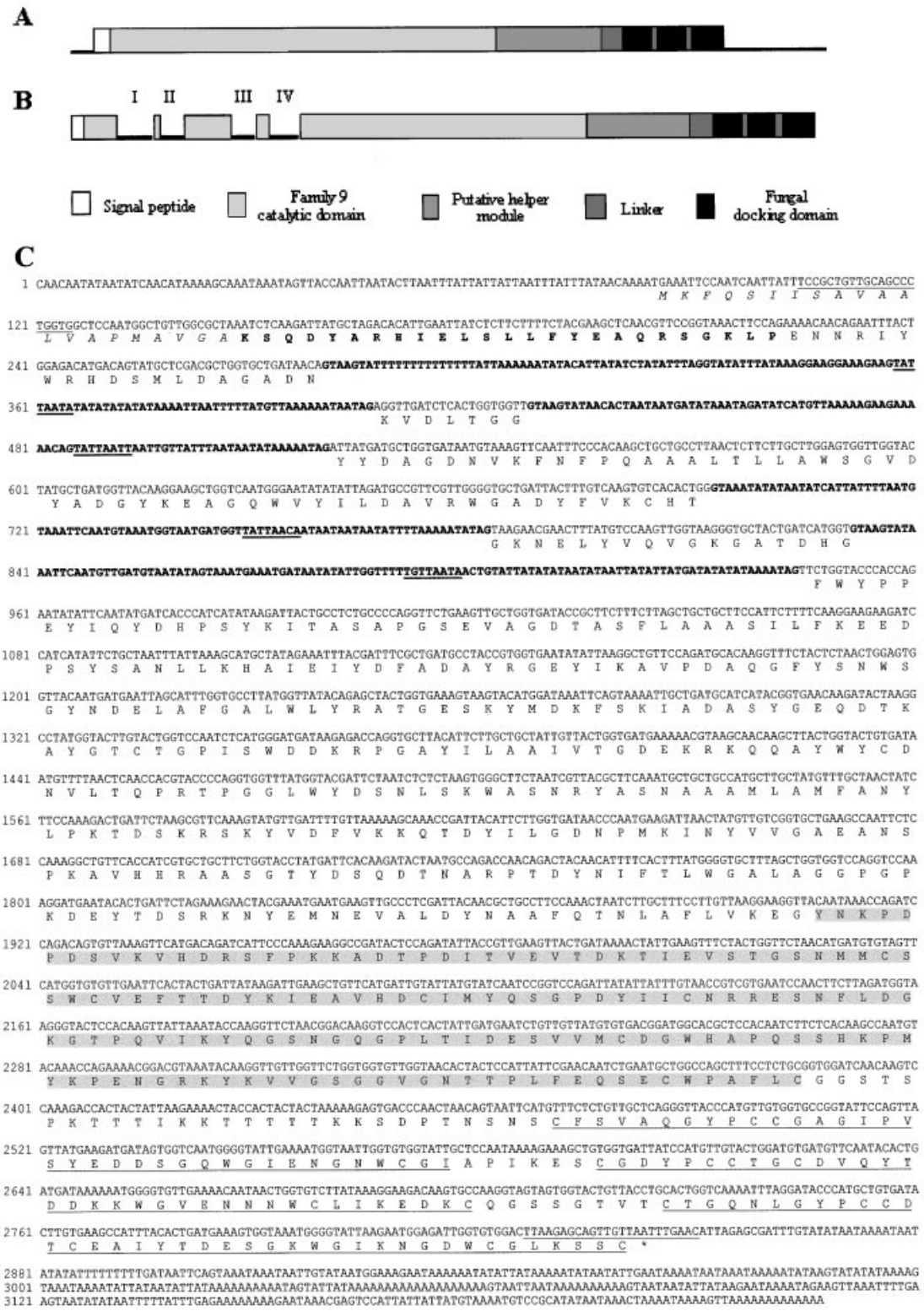

Figure 4 ***Cel9A*** **sequence**

(**A** and **B**) Schematic representations of the *cel9A* cDNA and gDNA clone respectively. The introns are indicated in **bold** and are numbered from I to IV. (**C**) composite sequence of the cDNA and genomic sequence of *cel9A*. Intron nucleotides are depicted in **bold** and the nucleotide stretch probably involved in lariat formation is underlined. The primers used for amplification of the genomic *cel9A* clone are underlined. The stop codon is indicated by an asterisk (*). The signal peptide is depicted *italic* and the experimentally determined N-terminus is depicted **bold**. The 160-residue module of unknown function is shaded and the three fungal dockerins are underlined.

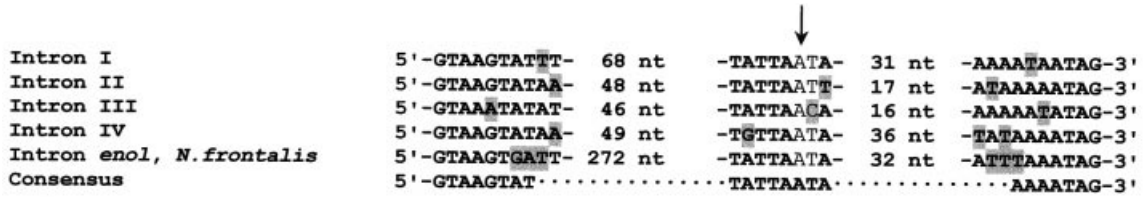

Figure 5 Alignment of the *cel9A* introns and the *enol* intron from *Neocallimastix frontalis*

The consensus was derived from conservation of residues in four out of five sequences and is given underneath the alignment. The arrow indicates the adenosine residue presumably involved in lariat formation during splicing.

of this intron could not be experimentally verified, because *enol* mRNA instability prevented the synthesis of full-length cDNA. The *cel9A* introns were compared with the predicted one from *N. frontalis enol* and analysed for the conservation of the 5′- and 3′-splicing sites and the consensus (U/C)N(U/C)U(G/A)A(U/C) for nucleotides involved in lariat formation during pre-mRNA splicing (Figure 5). No significant sequence conservation immediately upstream or downstream of the introns was found. Close inspection of the 3′ part of the introns showed the presence of the consensus of nucleotides involved in lariat formation. This enabled the construction of a consensus for introns from anaerobic fungi, including the exon/intron boundaries and the 8-nt palindromic sequence considered to be involved in lariat formation during splicing:

5′-GTAAGTAT-$(nt)_{n>40}$-TATTA<u>A</u>TA-$(nt)_{20–40}$-AAAATAG-3′,

with the putative adenosine residue involved in lariat formation underlined.

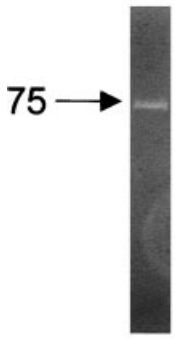

Figure 6 Cel9A CMC zymogram

The cell-free extract of the *cel9A* immunoclone expressed in *E. coli* BL21 was analysed. The corresponding molecular mass is indicated in kDa.

Activity of heterologous Cel9A

To investigate the production of functional Cel9A from the immunoclone, the isolated immunoclone was transformed into *E. coli* BL21. Transformants were selected on LB plates containing CMC and activity was detected after Congo Red staining, indicating functional expression of heterologous Cel9A. Plates with or without 1 mM Ca^{2+} did not reveal any significant difference in activity. Clearest activity was achieved on plates which did not contain IPTG. Possibly traces of lactose present in the casein used for the medium or leakage of the *lacZ* promoter, caused a low level of expression, resulting in small amounts of functional Cel9A protein. This was supported by the observation that growth of the transformants in the presence of 1% (w/v) glucose did not result in any detectable CMCase activity in a parallel experiment. Functional expression in a similarly structured cDNA clone was also observed for a xylanase in *Piromyces equi* [45]. Zymogram analysis of *E. coli* BL21 cell-free extract was investigated and showed an activity band with an apparent molecular mass of about 75 kDa after overnight incubation (Figure 6). Activity was very low in *E. coli*, and attempts to express the family 9 catalytic domain intra- or extracellularly, using the *Pichia pastoris* pPICZ vector system, were unsuccessful. These results, however, indicate that *cel9A* encodes a functional protein and that Cel9A has CMCase activity.

Structure and function of the catalytic domain

A BLAST search with the encoded mature protein indicated that Cel9A has a modular organization, starting with a domain that shows similarity to family 9 GHs. To obtain further insight into the structure and function of the *Piromyces* sp. strain E2 Cel9A catalytic domain, homology modelling was applied. The highest sequence similarity of Cel9A was not to the family 9 components of clostridial cellulosomes, but to the family 9 catalytic domain of the E4 enzyme from *Thermobifida fusca*. Residues 1–445 of the mature sequence comprise the catalytic module, based on comparisons with the *T. fusca* enzyme, and the high amino acid sequence homology (45% identity and 60% similarity; Figure 7A) made this protein an excellent template for homology modelling of the *Piromyces* catalytic module (Figure 7B). Backbone atoms in the core regions of the actual structure would be expected to be very similar to those of the model, with a root-mean-square difference of $\approx$1.0 Å ($\approx$0.1 nm)[46]. The structure of the GH family 9 domain is that of an $(\alpha/\alpha)_6$ barrel, with a shallow binding cleft running along one face. The residues directly involved in catalysis were all conserved. As with other GH family 9 enzymes, the *Piromyces* sp. strain E2 enzyme is expected to convert the β-glycosidic bond of the substrate into the α-enantiomer in the product during hydrolysis as an 'inverting' enzyme. Glu^{425} of the Cel9A sequence was identified as the acid for catalysis, while Asp^{54} and/or Asp^{57} could act as base to activate the catalytic water in the reaction. The Ca^{2+}-binding site was also conserved, involving seven ligands that are contributed by two loops near the active site, namely Ser^{210}-OG, ^{211}-O, Asp^{214} (OD1 and OD2), Glu^{215} (OE1 and OE2), and ^{262}-O.

Interpretation of the model was further enhanced by inspection of the six subsites observed in the *T. fusca* structures [34]. The residues lining the substrate-binding cleft are highly conserved. A number of small insertions and deletions could have minor

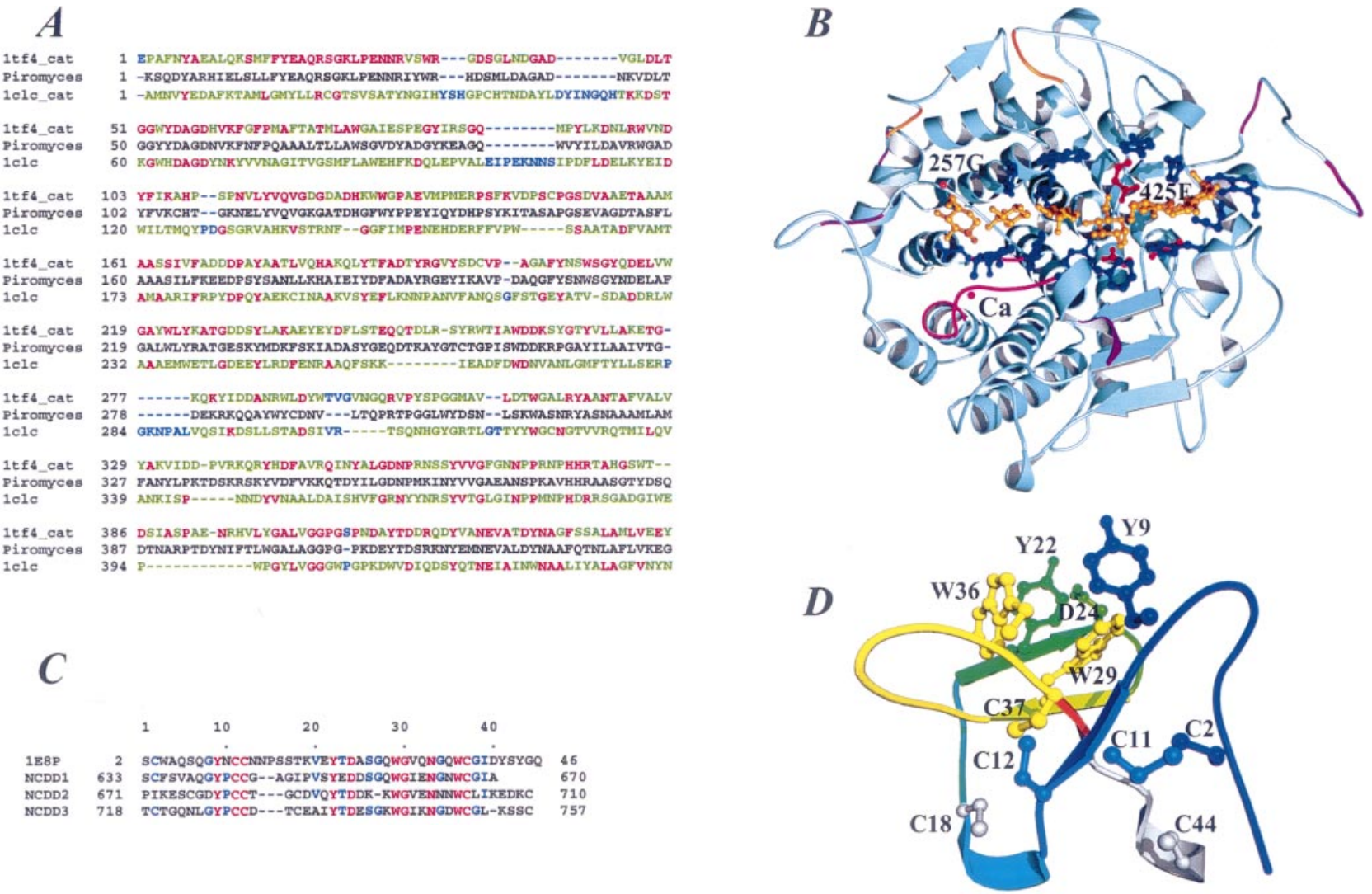

Figure 7 Homology modelling

(**A**) Sequences of the catalytic domains of *T. fusca* E4 (1tf4_cat) and the *C. thermocellum* CelD (1clc_cat) were aligned using their structures, together with the pairwise alignment of E4 and Cel9A, as described in the Experimental section. Residues identical with Cel9A are coloured red, insertions are blue. (**B**) Homology model of the Cel9A catalytic module; carbohydrate (gold) is modelled into six subsites using the complex structures of *T. fusca* E4 (Protein Data Bank entry code 4TF4 [34]). Residues directly implicated in catalysis are shown in red, and conserved residues that line the substrate-binding cleft are blue, while those of the Ca^{2+}-binding site are pink. The positions and approximate extent of insertions and deletions are marked in purple and orange respectively. The glycine residue that replaces the tryptophan of the −4 site in *T. fusca* E4 is shown as a coral sphere. (**C**) Sequences of the three *Piromyces* sp. strain E2 dockerins are aligned with that corresponding to the NMR structure of the *P. equi* dockerin (1E8P; [36]). Numbering at the top indicates the relative position, while at left and right are found the positions in the actual sequence of the mature protein. Absolutely conserved residues are red, and ones that are conserved in three out of four sequences are blue; (**D**) A homology model of dockerin I of Cel9A, based on 1E8P [36]. Colouring through the sequence follows a 'rainbow', beginning with blue at the N-terminus, and ending with red at the C-terminus. Two pairs of disulphide-bonding residues are shown in the same colouring, as are conserved side chains that have been implicated in ligand binding [36]. Additional features that will be present in dockerins II and III (type 3) are approximately placed in grey, including two potential disulphide partners.

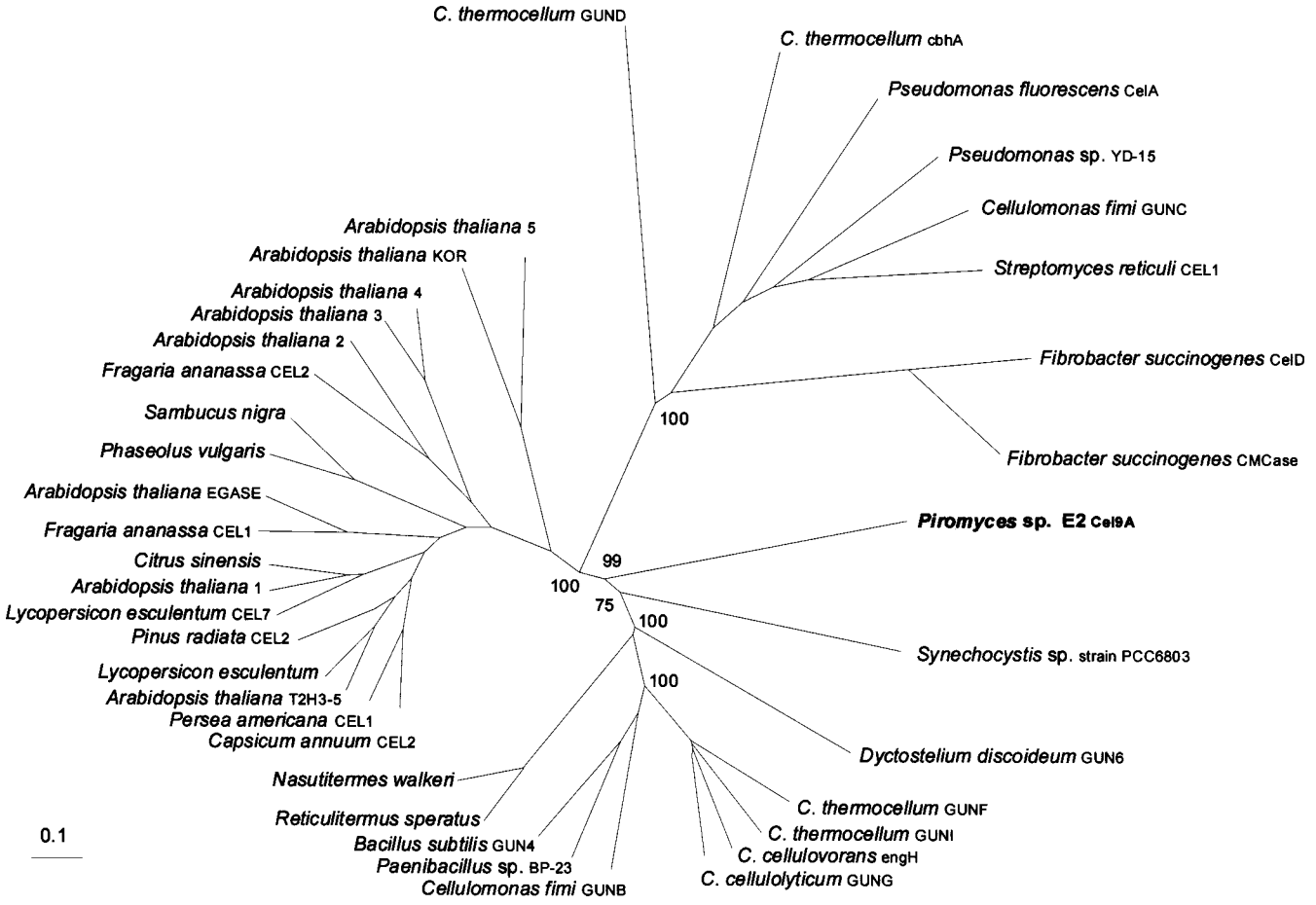

Figure 8 Cel9A phylogeny

Phylogenetic tree of the *Piromyces* sp. strain E2 catalytic domain with seed sequences available from pf00759. The Cel9A sequence from *Piromyces* sp. strain E2 clusters with bacterial sequences. Relevant bootstrap values are given (as percentages). The bar represents 1 nt substitution per 100 nucleotides.

effects on the local structure, but only few would be expected to influence substrate binding. A small change near residue 390 could have some impact on the +2 (and potential +3) site. Probably the most interesting difference is the replacement of Trp^{256} by Gly^{257}, which eliminates an aromatic stacking interaction with substrate that is the basis of the −4 site of the *T. fusca* enzyme. A one-residue insertion near position 255 and a three-residue deletion near position 295 could also alter the shape of the substrate-binding cleft near the −4 subsite. The *T. fusca* enzyme is unusual in that it has the rather open substrate-binding cleft generally associated with endoglucanases, while at the same time possessing a significant amount of exo activity. This property has been partly accounted for by blocking of the cleft at the −4 end, as well as by the interaction with Trp^{256}. Additional contributions to E4′s exo activity, as well as its processivity, are provided by the cellulose binding modules (CBMs), which apparently help bind and orientate substrate within the enzyme [47]. The tryptophan, the CBMs, and possibly the blockage as well, are missing from the *Piromyces* sp. strain E2 enzyme, without any compensating loop extensions that could close the active-site cleft. There is only one other available structure for this GH family, that from a *Clostridium thermocellum* endoglucanase, CelD [13]. Although this enzyme is more distantly related than *T. fusca* E4 (amino acid sequence identity ≈23%), like *Piromyces* sp. strain E2 Cel9A it has an open active-site cleft and lacks the aromatic residue that would be characteristic for a −4 site. We conclude that the predominant activity of the Cel9A from *Piromyces* sp. strain E2 will be that of an endoglucanase.

Evolutionary relationships of the Cel9A catalytic domain

To investigate the phylogenetic relationships of the *Piromyces* sp. strain E2 Cel9A catalytic domain, the sequence was analysed, using the Pfam facility at the Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K. (www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00759). The Cel9A catalytic domain was identified and aligned to the family 9 seed sequences present in the Pfam protein family pf00759. This alignment was used to calculate a phylogenetic tree with bootstrap values, using the Tree Top cluster algorithm (Figure 8). The *Piromyces* sp. strain E2 Cel9A catalytic domain associated with bacterial sequences as well as with the cellulases identified from termites (consisting of only a catalytic domain). This indicated that the eukaryotic sequences were probably introduced through horizontal gene transfer from bacteria and that apparently introns were subsequently incorporated.

A second indication for horizontal gene transfer is the observation that Cel9A from *Piromyces* sp. strain E2 is the first GH family 9 enzyme-encoding gene isolated from a fungus, although over 200 sequences have been reported from a wide range of other organisms ranging from bacteria [13,48], plants [49], termites, cockroaches [50], the crayfish *C. quadricarinatus* [43] to the cellular slime mould *Dictyostelium discoideum* [51]. In bacteria and in cellulose-feeding insects, family 9 enzymes act as endoglucanases in cellulose degradation, but in plants and in *D. discoideum* they are involved in morphological events. Apparently, the family 9 catalytic module was distributed very widely during evolution, but not maintained to a large extent in fungi. Interestingly, the major 80 kDa protein of the *Piromyces* sp. strain E2 cellulosome was also recently shown to include (Figure 2) a GH family 48 catalytic domain, making it the first sequence of eukaryotic origin in a group that otherwise contains only bacterial sequences (P. J. M. Steenbakkers, A. Freelove, B. van Cranenbroek, B. M. C. Sweegers, H. R. Harhangi, G. D. Vogels, G. P. Hazlewood, H. J. Gilbert, and H. J. M. Op den Camp, unpublished work). Apparently, the major components of this anaerobic fungal cellulosome resemble the key cellulosome components of anaerobic bacteria. With the discovery of both GH families in the *Piromyces* sp. strain E2 cellulosome, the possibility that these enzymes are in some way incompatible with fungi can be disregarded. Moreover, the GH family 48 bacterial enzymes are so-called processive endocellulases that have always been found associated with GH family 9 endoglucanases. The two enzymes occur as separate proteins and even as fusion proteins, and are known to function synergistically [11,13]. Previously characterized aerobic fungal cellulolytic systems do not contain GH family 48 enzymes, but instead use a biochemically equivalent enzyme from GH family 7. This class of enzymes is restricted to aerobic fungi and shows synergy with endoglucanases that do not belong to GH family 9. A possible explanation for these observations could be that cellulases occur as functional partners, which are not strictly confined to fungi or bacteria, but are bound to each other by the fact that they lead to more efficient, synergistic degradation of cellulose. Transfer of either of the cellulases would benefit the recipient most if their respective partner accompanies them. Such interdependence of cellulases would favour the existence of one of the cellulase pairs in a receiving organism. This theory, based on the unequal distribution of fungal members in GH families and the observed synergy between certain families, could possibly also explain the redundancy in biochemical properties of evolved cellulase families.

Auxiliary domains in Cel9A

The C-terminal regions of Cel9A consist of a serine/threonine-rich linker followed by three fungal dockerin domains discussed below. The middle portion of the protein, namely that coming between the catalytic module and the linker (residues 446–605 in the mature sequence), could not be assigned a function. No detectable sequence identity with proteins of known structure was detected, using either PSI-BLAST or HMM ('hidden Markov models') searches. The only entry in GenBank® with significant similarity was a partial sequence from *Orpinomyces* sp. PC-2, designated as CelJ (accession no. AF177207) [9]. Of the 75 residues of the CelJ sequence that correspond to the middle segment of Cel9A, 57% are identical. However, the function of this region is also unknown for CelJ. The consensus of secondary structure prediction algorithms suggested that this region is likely to consist primarily of β-sheet structure. It is rich in cysteine residues (seven out of 160 residues), of which five lie within the predicted strand segments, an observation that is also consistent with an all-β-sheet structure. A number of other cellulases contain unrelated β-sheet domains of the immunoglobulin or fibronectin types, for which the function has also not yet been assigned.

The C-terminal end of Cel9A consists of a linker and three fungal dockerins (designated I, II and III). Recently, a division of fungal dockerins into three subfamilies was proposed on the basis of the number, location and sequence context of cysteine residues [9]. According to this classification, dockerin I classifies as type 2 (with four cysteine residues), and dockerins II and III as type 3 dockerins (with six cysteine residues). The NMR structure of a type 2 fungal dockerin from *Piromyces equi* was recently published [36]. The structure revealed that the four cysteine residues are all involved in the disulphide bridges essential for binding the putative fungal scaffoldin protein. The Cel9A dockerins were modelled using this structure. The sequences of the three dockerins of Cel9A are aligned, together with the *P. equi* dockerin in Figure 7(C). Amino acid sequence identity is 52% in the equivalent regions of dockerin I and the *P. equi* module; the homology model generated for dockerin I is shown in Figure 7(D). Dockerins II and III are of type 3, which differs slightly more from the template (identity 35–47% with equivalent portions of the *P. equi* module). The residues on the surface implicated in binding are totally conserved (at positions Tyr^{9}, Tyr^{22}, Asp^{24}, Trp^{29} and Trp^{36} of the alignment). Deletion of two to three residues near position 15 will shorten one loop in all of the Cel9A dockerins, as will smaller deletions near positions 27 and 40 in dockerins II and III respectively. Neither change is likely, however, to alter the binding surface substantially. Gly^{30} is conserved for structural reasons, while the need for conservation of Asn^{31} is not apparent. Residues 12 and 37 make a disulphide bond in the *P. equi* structure, which is expected to be present in each of the Cel9A dockerins. Residues 2 and 11 also make a disulphide bond in the *P. equi* dockerin, which should be present in Cel9A dockerins I and III. Dockerin II could instead have a disulphide bond between Cys^{6} and Cys^{11}, as these are likely to be sufficiently close in the structure. In dockerins II and III (type 3), there are two additional cysteine residues at positions 18 and 44, the latter of which is associated with a small C-terminal extension of the sequence. In the homology models these are expected to be close together and provide a good possibility for an additional disulphide bond (Figure 7D). These two residues are found at or near the surface of the module and could instead form a pair of interdomain disulphide bonds (e.g. between dockerins II and III, or between two molecules of Cel9A). It should be noted, however, that some enzymes from anaerobic fungi contain only a single copy of this dockerin type, which indicates that interdomain disulphides within a single protein cannot be an essential feature in their function. In any case, these results confirm the importance of the cysteine residues in fungal dockerin structure, and suggest a physical basis for this.

In summary, an improved method of purification of *Piromyces* sp. strain E2 cellulosome allowed Cel9A to be identified as a dominant component. The protein includes a GH family 9 catalytic domain at the N-terminus, which probably functions as an endoglucanase, and three fungal dockerin domains at the C-terminus; the function of the 160-residue region between these segments has not yet been assigned. Cel9A has a modular organization most similar to the cellulosome components GUNF (P26224), engH (O65987) and engG (P37700) from *C. thermocellum*, *C. cellulovorans* and *C. cellulolyticum* respectively. Furthermore, this paper is the first describing a family 9 GH from a fungus. Sequence analysis suggests that the gene has a bacterial origin, indicating that its introns have been incorporated by the

anaerobic fungus itself. These results also show that Cel9A as a major component of the fungal cellulosome is similar to key components of clostridial cellulosomes, supporting the view that bacterial and eukaryotic cellulosomes are the result of convergent evolution.

We thank Professor Y. Shoham (Technion-Israel Institute of Technology, Haifa, Israel) and Dr E. A. Bayer (The Weizmann Institute of Science, Rehovot, Israel) for kindly providing the purified cellulosome from *Clostridium thermocellum* and H. R. Harhangi (University of Nijmegen) for valuable discussions. We are also grateful for the assistance of Robert Kerkhoven and Jelle Eygenstein during the analysis of protein and DNA samples. S.L.M. and W.U. were supported by the Swedish Natural Sciences Research Council and by the Glycoconjugates in Biological Systems (GLIBS) network of the The Swedish Foundation for Strategic Research (Stiftelsen för Strategisk Forskning).

REFERENCES

1 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

2 Lee, S. S., Ha, J. K. and Cheng, K. (2000) Relative contributions of bacteria, protozoa, and fungi to *in vitro* degradation of orchard grass cell walls and their interactions. Appl. Environ. Microbiol. **66**, 3807–3813

3 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

4 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

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

6 Ding, S. Y., Bayer, E. A., Steiner, D., Shoham, Y. and Lamed, R. (2000) A scaffoldin of the *Bacteroides cellulosolvens* cellulosome that contains 11 type II cohesins. J. Bacteriol. **182**, 4915–4925

7 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

8 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

9 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

10 Kruus, K., Wang, W. K., Ching, J. and Wu, J. H. D. (1995) Exoglucanase activities of the recombinant *Clostridium thermocellum* CelS, a major cellulosome component. J. Bacteriol. **177**, 1641–1644

11 Bhat, S., Goodenough, P. W., Owen, E. and Bhat, M. K. (1994) Isolation of four major subunits from *Clostridium thermocellum* cellulosome and their synergism in the hydrolysis of crystalline cellulose. Int. J. Biol. Macromol. **16**, 335–342

12 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

13 Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Beguin, P. and Aubert, J. P. (1992) 3-Dimensional structure of a thermostable bacterial cellulase. Nature (London) **357**, 89–91

14 Ding, S. Y., Bayer, E. A., Steiner, D., Shoham, Y. and Lamed, R. (1999) A novel cellulosomal scaffoldin from *Acetivibrio cellulolyticus* that contains a family 9 glycosyl hydrolase. J. Bacteriol. **181**, 6720–6729

15 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

16 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

17 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

18 Li, X. L., Chen, H. and Ljungdahl, L. G. (1997) Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. Appl. Environ. Microbiol. **63**, 628–635

19 Durand, R., Fischer, M., Rascle, C. and Fevre, M. (1995) *Neocallimastix frontalis* enolase gene, *enol*: First report of an intron in an anaerobic fungus. Microbiology **141**, 1301–1308

20 Teunissen, M. J., Op den Camp, H. J. M., Orpin, C. G., Huis in 't Veld, J. H. J. and Vogels, G. D. (1991) Comparison of growth characteristics of anaerobic fungi isolated from ruminant and non-ruminant herbivores during cultivation in a defined medium. J. Gen. Microbiol. **137**, 1401–1408

21 Dubray, G. and Bezard, G. (1982) A highly sensitive periodic acid–silver stain for 1,2-diol groups of glycoproteins and polysaccharides in polyacrylamide gels. Anal. Biochem. **119**, 325–329

22 Gerwig, G. J., de Waard, P., Kamerling, J. P., Vliegenthart, J. F., Morgenstern, E., Lamed, R. and Bayer, E. A. (1989) Novel O-linked carbohydrate chains in the cellulase complex (cellulosome) of *Clostridium thermocellum*. 3-*O*-Methyl-*N*-acetylglucosamine as a constituent of a glycoprotein. J. Biol. Chem. **264**, 1027–1035

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

24 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, New York

25 Ohyama, T., Tagashira, H. and Tsujibayashi, H. (1997) Easy screening of recombinant plasmids. Tech. Tips Online **1**, 8:T40073

26 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

27 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

28 Karplus, K., Sjolander, K., Barrett, C., Cline, M., Haussler, D., Hughey, R., Holm, L. and Sander, C. (1997) Predicting protein structure using hidden Markov models. Proteins (Suppl.), 134–139

29 Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16**, 10881–10890

30 MacCallum, R. M., Kelley, L. A. and Sternberg, M. J. (2000) SAWTED: structure assignment with text description-enhanced detection of remote homologues with automated SWISS-PROT annotation comparisons. Bioinformatics **16**, 125–129

31 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

32 Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. **A47**, 110–119

33 Kleywegt, G. J. (1996) Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr. **D52**, 842–857

34 Sakon, J., Irwin, D., Wilson, D. B. and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nat. Struct. Biol. **4**, 810–818

35 Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. and Jones, T. A. (2001) Around O. In International Tables for Crystallography, Volume F, Crystallography of Biological Macromolecules (Rossmann, M. G. and Arnold, E., eds), chapter 17.1, pp. 353–367, Kluwer Academic Publishers, Dordrecht

36 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

37 Kraulis, P. (1991) MOLSCRIPT – A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24**, 946–950

38 Harris, M. and Jones, T. A. (2001) Molray – a web interface between O and the POV-Ray ray tracer. Acta Crystallogr. **D57**, 1201–1203

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

40 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

41 Choi, S. K. and Ljungdahl, L. G. (1996) Dissociation of the cellulosome of *Clostridium thermocellum* in the presence of ethylenediaminetetraacetic acid occurs with the formation of truncated polypeptides. Biochemistry **35**, 4897–4905

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

43 Byrne, K. A., Lehnert, S. A., Johnson, S. E. and Moore, S. S. (1999) Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. Gene **239**, 317–324

44 Kriventseva, E. V. and Gelfand, M. S. (1999) Statistical analysis of the exon–intron structure of higher and lower eukaryote genes. J. Biomol. Struct. Dyn. **17**, 281–288
45 Gilbert, H. J., Hazlewood, G. P., Laurie, J. I., Orpin, C. G. and Xue, G. P. (1992) Homologous catalytic domains in a rumen fungal xylanase – evidence for gene duplication and prokaryotic origin. Mol. Microbiol. **6**, 2065–2072
46 Chothia, C. and Lesk, A. M. (1986) The relation between the divergence of sequence and structure in proteins. EMBO J. **5**, 823–826
47 Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A. and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. J. Bacteriol. **180**, 1709–1714
48 Riedel, K. and Bronnenmeier, K. (1998) Intramolecular synergism in an engineered exo-endo-1,4-β-glucanase fusion protein. Mol. Microbiol. **28**, 767–775
49 Nicol, F., His, I., Jauneau, A., Vernhettes, S., Canut, H. and Hofte, H. (1998) A plasma membrane-bound putative endo-1,4-β-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. EMBO J. **17**, 5563–5576
50 Watanabe, H. and Tokuda, G. (2001) Animal cellulases. Cell. Mol. Life Sci. **58**, 1167–1178
51 Giorda, R., Ohmachi, T., Shaw, D. R. and Ennis, H. L. (1990) A shared internal threonine-glutamic acid-threonine-proline repeat defines a family of *Dictyostelium discoideum* spore germination specific proteins. Biochemistry **29**, 7264–7269

Received 20 December 2001/20 March 2002; accepted 16 April 2002