Extremophiles (1999) 3:103–111

ORIGINAL PAPER

Daniel D. Morris · Moreland D. Gibbs · Michelle Ford
Justin Thomas · Peter L. Bergquist

Family 10 and 11 xylanase genes from *Caldicellulosiruptor* sp. strain Rt69B.1

Received: May 13, 1998 / Accepted: October 22, 1998

Abstract Three family 10 xylanase genes (*xynA*, *xynB*, and *xynC*) and a single family 11 xylanase gene (*xynD*) were identified from the extreme thermophile *Caldicellulosiruptor* strain Rt69B.1 through the use of consensus PCR in conjunction with sequencing and polyacrylamide gel electrophoresis. These genes appear to comprise the complete endoxylanase system of Rt69B.1. The *xynA* gene was found to be homologous to the *xynA* gene of the closely related *Caldicellulosiruptor* strain Rt8B.4, and primers designed previously to amplify the Rt8B.4 *xynA* gene could amplify homologous full-length *xynA* gene fragments from Rt69B.1. The complete nucleotide sequences of the Rt69B.1 *xynB*, *xynC*, and *xynD* genes were obtained using genomic walking PCR. The full-length *xynB* and *xynC* genes are more than 5 kb in length and encode highly modular enzymes that are the largest xylanases reported to date. XynB has an architecture similar to the family 10 xylanases from *Thermoanaerobacterium saccharolyticum* (XynA) and *Clostridium thermocellum* (XynX) and may be cell wall associated, while XynC is a bifunctional enzyme with an architecture similar to the bifunctional β-glycanases from *Caldicellulosiruptor saccharolyticus*. The *xynD* gene encodes a two-domain family 11 xylanase that is identical in architecture to the XynB family 11 xylanase from the unrelated extreme thermophile *Dictyoglomus thermophilum* strain Rt46B.1. The sequence similarities between the Rt69B.1 xylanases with respect to their evolution are discussed.

Key words Xylanases · Genomic walking PCR · Multidomain enzymes · Cellulose-binding domains · Linker sequences · Horizontal gene transfer

Communicated by G. Antranikian

D.D. Morris · M. Ford · J. Thomas
Centre for Gene Technology, School of Biological Sciences, University of Auckland, Auckland, New Zealand

D.D. Morris · M.D. Gibbs · P.L. Bergquist
School of Biological Sciences Macquarie University, Sydney, New South Wales, Australia

P.L. Bergquist
Department of Molecular Medicine, University of Auckland Medical School, Auckland, New Zealand

P.L. Bergquist (✉)
Research Office, Macquarie University, Sydney, New South Wales 2109, Australia
Tel. +61-2-9850-8614; Fax +61-2-9850-8799
e-mail: peter.bergquist@mq.edu.au

Introduction

Caldicellulosiruptor Rt69B.1 is an extremely thermophilic gram-positive bacterium that was isolated from a New Zealand hot spring. It is closely related to *Caldicellulosiruptor* isolate Rt8B.4 and *Caldicellulosiruptor saccharolyticus* (*Cs. saccharolyticus*), as judged by SSU (16S) rRNA sequence comparisons. Rt69B.1 can grow at temperatures up to 70°C in the laboratory and can utilize xylan as a sole carbon source (Rainey et al. 1993, 1994; Dwivedi et al. 1996).

The β-glycanase system of *Cs. saccharolyticus* has been extensively studied, and the genes encoding 10 cellulolytic or hemicellulolytic enzymes have been reported (Love and Streiff 1987; Lüthi and Bergquist 1990; Lüthi et al. 1990a,b,c, 1991; Saul et al. 1990; Gibbs et al. 1992; Morris et al. 1995; Te'o et al. 1996). A striking feature of several of the *Cs. saccharolyticus* β-glycanases is their bifunctional domain architecture, consisting of central family III cellulose-binding domains (CBDs, Tomme et al. 1994), which are bordered by catalytic domains of the cellulase, xylanase, mannanase, or cellobiohydrolase families. The individual domains of these enzymes are partitioned into discrete and functionally independent entities by prominent linker peptides rich in proline, threonine, and serine residues. Additionally, *Cs. saccharolyticus* encodes two unifunctional family 10 xylanases (Henrissat and Bairoch 1995): XynA, a single-domain xylanase, and XynE, a multidomain xylanase (Te'o, unpublished Ph.D thesis, University of Auckland, New Zealand, 1996).

A single xylanase gene (*xynA*) was recently identified from a *Caldicellulosiruptor* Rt8B.4 gene library (Dwivedi et al. 1996). Cloning and sequencing of this gene revealed that the encoded peptide (XynA) was composed of a family 10 xylanase domain with tandem N-terminal 'thermostabilizing' domains (TSDs). The overall architecture of Rt8B.4 XynA xylanase is identical to the XynE family 10 xylanase from *Cs. saccharolyticus*, and these two enzymes share 72% sequence identity over their xylanase domains. Duplicated N-terminal TSDs appear to be exclusive constituents of family 10 xylanases from thermophilic bacteria, and are present in several other thermophilic family 10 xylanases, including *Thermotoga maritima* XynA (Lee et al. 1993), *Thermotoga neopolitana* XynA (Winterhalter et al. 1995), *Thermoanaerobacterium saccharolyticum* XynA (Lui et al. 1996), *Thermoanaerobacterium* sp. XynA (Zverlov et al. 1996), and the XynB and XynC multidomain xylanases from *Thermotoga* sp. FjSS3B.1 (our laboratory, unpublished data). In addition, *Clostridium thermocellum* XynX and XynY possess single N-terminal TSDs (Pack and Jung 1991; Fontes et al. 1995), while a single TSD homolog is positioned at the N-terminus of the family 10 xylanase domain of XynC from the mesophilic bacterium *Cellulomonas fimi* (Clarke et al. 1996). Duplicated family IX CBDs that are positioned at the immediate C-terminus of thermophilic family 10 xylanase domains are also common.

The nucleotide sequences of more than 100 family 10 and family 11 xylanase genes have now been deposited in the Genbank and EMBL databases. These xylanase genes were identified from gene libraries that were screened for either hybridization to labeled gene probes or, more commonly, expression of endoxylanase activity. An alternative approach to the identification of novel family 10 and family 11 xylanase genes is through the polymerase chain reaction (PCR) using broad-specificity oligonucleotide consensus primers. Because this approach is PCR based, it is also highly sensitive and offers an expedient means for the identification of xylanase genes directly from the genomic DNAs of the organisms under consideration. Furthermore, through the use of genomic walking PCR techniques, the identified xylanase gene(s) can be sequenced in a quick and relatively straightforward manner (Morris et al. 1995). We report here the first example of the use of such a two-step PCR approach for the identification and subsequent sequencing of multiple xylanase genes from a single microorganism.

Materials and methods

Bacterial strains

Escherichia coli strain JM101 {Δ(*lac-proAB*), *thi-l*, *supE44 F′* [*traD36* proAB$^+$ *lacZΔM15*]} was used as the bacterial host for all DNA cloning and expression studies.

Polymerase chain reactions

PCRs were performed in 50-μl reaction mixtures containing 30 pmoles of forward primer and 30 pmoles of reverse primer, 0.25 mM deoxynucleoside triphosphates (dNTPs), 1 U *Taq* DNA polymerase, 2.5 mM $MgCl_2$, and 1/10th volume of *Taq* DNA polymerase PCR buffer (Perkin-Elmer, Scoresby, Victoria, Australia). Approximately 10 ng of DNA was used as template. PCRs were cycled in a Perkin-Elmer GeneAmp 2400.

Consensus PCRs

Family 10 xylanase consensus fragments (FXCFs, approximately 150 bp) were amplified using the newFF (5′-CAY ACN CTS GTT TGG CA-3′; degenerate bases are assigned according to the standard nucleotide degeneracy codes) and newFR (5′-TGG GAY GTK GTR AAY GA-3′) consensus primers. Family 11 xylanase consensus fragments (GXCFs) of approximately 300 bp were amplified using the xynGF (5′-TAT NTG RST NTM TAT GGW TGG-3′) and xynGR (5′-GAA GGN TAC CAA AGN AGC GG-3′) consensus primers. Smaller GXCFs, of approximately 150 bp, were amplified using the newGF (5′-GAR TAY TAY RTY GTY GAM MGY TGG GG-3′) and newGR (5′-ACY TTY NMS CAG TAC TGG AGY GTY CG-3′) consensus primers. Consensus PCRs were performed for 35 cycles using the following profile: 94°C for 30 s; either 37°C (for the xynGF/xynGR primers) or 55°C (for the newGF/newGR and newFF/newFR primers) for 30 s; and 72°C for 30 s.

Genomic walking PCRs

The rationale behind the design and implementation of the genomic walking PCR (GWPCR) technique has been described previously (Morris et al. 1995). Rt69B.1 restriction fragments were generated using the *Nco*I, *Dra*I, *Eco*RV, *Hin*cII, *Hpa*I, *Pvu*II, *Ssp*I, *Hin*dIII, *Pst*I, *Eco*RI, *Sac*I, *Xba*I, *Kpn*I, *Bam*HI, and *Sph*I restriction endonucleases. Synthetic DNA linkers, which were prepared by annealing an upper oligonucleotide (5′-CAT GGC GCA GGA AAC AGC TAT GAC CGG T-3′) and a lower oligonucleotide (DS43, 5′-CGC GTC CTT TGT CGA TAC TGG CCA-3′), were ligated onto the termini of the Rt69B.1 restriction fragments. GWPCRs were performed for 35 cycles, using 30-s denaturation, primer-annealing, and primer-extension steps, and annealing temperatures of 60° to 68°C.

DNA sequencing

DNA sequencing was carried out on an Applied Biosystems model 373A (stretch) automated DNA sequencer, using dye primer or dye terminator chemistries. The M13mp18 bacteriophage vector and the pUC18 plasmid vector were used for sequencing the Rt69B.1 PCR fragments. PCR products were purified using the High Pure PCR product purification system (Boehringer Mannheim, Auckland, New Zealand) and then treated for 30 min at 37°C with the Klenow fragment of *Escherichia coli* DNA polymerase I (1 U; BRL Life Technologies) and T4 polynucleotide kinase (1 U; Boehringer Mannheim, Auckland, New Zealand) in the presence of 0.25 mM dNTPs and 1/10th volume of T4

DNA ligase buffer (Boehringer Mannheim) to make the termini of the fragments blunt. These fragments were purified from 1% low melting temperature agarose using the BresaClean DNA purification kit (Bresatec, Adelaide, South Australia). The final Rt69B.1 DNA fragments were cloned non-directionally into the phosphatased *Sma*I site of either M13mp18 RF DNA or pUC18 plasmid DNA. Single-stranded bacteriophage DNAs were prepared using the methods described by Sambrook et al. (1989), and sequencing-grade pUC18 plasmid DNAs were prepared using the High Pure plasmid purification system (Boehringer Mannheim). The sequence analysis software of the Genetics Computer Group (Devereux et al. 1984) installed on a Silicon Graphics Indigo workstation was used for analysis of the Rt69B.1 sequence data.

Ribosomal small subunit RNA (SSU) sequence of *Caldicellulosiruptor* Rt69B.1

The sequence of the SSU rDNA gene and the phylogenetic position of the isolate was determined as previously described (Dwivedi et al. 1996).

Construction of recombinant pJLA602 plasmids

Recombinant pJLA602 plasmids (Schauder et al. 1987) that contained gene fragments encoding the family 10 xylanase domains of Rt69B.1 XynA, XynB, and XynC, and the family 11 xylanase domain of XynD were prepared. Rt69B.1 *xynA* gene fragments were amplified using the PP12 (5′-CCT ATT CCA TGG ACC TTT ATT CAA TCT CAG ATG-3′) and PP11 (5′-ATA ATT GGA TCC ATT ACT CTT TTG GGA CAA CTG-3′) primers designed previously (Dwivedi et al. 1996); Rt69B.1 *xynB* and gene fragments were amplified using the XynBN (5′-CGG TTA GCA TGC CTG CTG CGG CAA AAG AGC C-3′) and XynBC (5′-GCA ACA GGA TCC ACC GAA GGT TCT ACA ATA GC-3′) primers; Rt69B.1 *xynC* gene fragments were amplified using the XynCN (5′-GCG CTC CCA TGG TAT TAC TTT ATG TTG AGG CG-3′) and XynCC (5′-GAC CGG ATC CCG GCA CCA CAG TTG GCT C-3′) primers; Rt69B.1 *xynD* and gene fragments were amplified using the XynDN (5′-GGC AGG CAT GCC ATT TAC CTC TAA TGC-3′) and XynDC (5′-CCA TTA CTG GAT CCA CTA CTC GAA CCA CC-3′) primers. The *xynA*, *xynB*, *xynC*, and *xynD* forward primers incorporated either a unique *Nco*I restriction site (PP10 and XynCN) or an *Sph*I restriction site (XynBN and XynDN); *Bam*HI sites were present in the reverse primers. Consequently, directional in-frame ligation into *Nco*I–*Bam*HI-cut or *Sph*I–*Bam*HI-cut pJLA602 plasmid DNA was possible.

The Rt69B.1 PCR fragments were purified from the reaction mixtures using the High Pure PCR purification system (Boehringer Mannheim), and 1 μg of purified DNA was then digested with *Nco*I–*Bam*HI or *Sph*I–*Bam*HI, as appropriate. The digested DNAs were then purified from 1% low melting temperature agarose gels, and ligated into pJLA602 plasmid DNA linearized with either *Nco*I and *Bam*HI or *Sph*I and *Bam*HI, as appropriate.

Production of xylanase samples

A total of 1200 ml of LB medium was supplemented with the sodium salt of ampicillin to a final concentration of 100 μg/ml, and inoculated with 12 ml of an overnight culture (grown at 32°C) of an *Escherichia coli* strain harboring the desired recombinant pJLA602 plasmid. The culture was incubated in a 5-1 flask with rapid shaking for approximately 3 h at 32°C, at which point the flask was transferred to a 42°C shaking water bath for 3 h of induction. The cells were then harvested by centrifugation, resuspended in milli-Q water, and lysed by passage through a French pressure cell. The whole cell extracts were then heat treated at 70°C for 30 min, and centrifuged at high speed to remove the denatured host proteins. The final extracts were stored at 4°C for immediate use or at −20°C for extended storage.

Assays for xylanase activity

Qualitative endo-1, 4-β-D-xylanase assays of transformant *E. coli* colonies harboring recombinant pJLA602 plasmids were performed using the Congo Red procedure of Teather and Wood (1982). The generation of reducing sugars from 0.25% oat spelts xylan solutions was measured quantitatively using the *p*-hydroxybenzoic acid hydrazide (PHBAH) colorimetric assay (Lever 1973). Sodium acetate was used to buffer solutions between pH 4.0 and pH 6.0, bistris propane (BTP; 1,3-bis[tris (hydroxymethyl)-methylamino] propane) for pH 6.0–9.0 and CAPS (3-[cyclohexylamino]-1-propanesulfonic acid) for pH 9.0–11.0.

GenBank accession numbers

The accession numbers for the xylanase gene sequences reported in this paper are *xynB*, AF036923; *xynC*, AF036924; and *xynD*, AF036925.

Results

Phylogenetic analysis of Rt69B.1

The SSU rRNA gene was isolated by PCR using oligonucleotide primers designed to amplify rRNA gene fragments from all known prokaryotic species. A PCR fragment of approximately 1800 bp was obtained, cloned into M13 mp10 in the forward and reverse orientation, and sequenced. Rt69B.1 was found to lie among cluster D of thermophilic *Clostridia* (Rainey et al. 1993) and is closely related to other *Caldicellulosiruptor* strains from which we have cloned β-glycanases previously (Dwivedi et al. 1996).

Consensus PCR analysis of the Rt69B.1 family 10 xylanase genes

A total of 12 family 10 xylanase consensus fragments (FXCFs), which were amplified from Rt69B.1, were

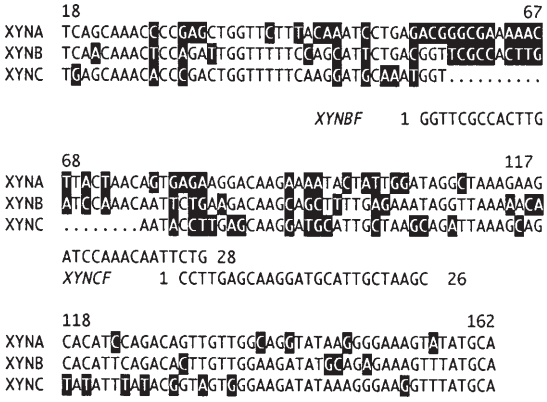

Fig. 1. Multiple sequence alignment of the Rt69B.1 *xynA*, *xynB*, and *xynC* family 10 xylanase consensus fragments (FXCFs). The consensus primer sequences have been removed. Variant bases are highlighted in reverse font (*dark*). The positions of the xynBF and xynCF downstream walking primers are shown (note that the xynBR and xynCR upstream walking primers are complementary to xynBF and xynCF, respectively)

sequenced. These sequences could be divided into three distinct sequence families, which were named *xynA*, *xynB*, and *xynC* (Fig. 1). The *xynA* and *xynB* sequences were 168 bp in length and shared 64% sequence identity, while the *xynC* sequences were slightly shorter at 150 bp and shared 65% and 71% sequence identity, respectively, to the *xynA* and *xynB* fragments.

Following electrophoresis through 6% polyacrylamide gel (PAGE), the Rt69B.1 *xynA*, *xynB*, and *xynC* FXCFs could be resolved as individual bands. It should be noted that the *xynA* and *xynB* FXCFs could not be resolved beyond single bands on 4% agarose gels; thus, PAGE was clearly separating the FXCFs according to both molecular weight and sequence-dependent conformational differences (data not shown). The Rt69B.1 *xynA* FXCF sequence shared 98% sequence identity to the FXCF region of the *xynA* gene from the closely related *Caldicellulosiruptor* isolate Rt8B.4 (Dwivedi et al. 1996). On the assumption that these two genes were mostly identical over their entire sequence, the Rt8B.4 *xynA* PCR primers were used to amplify the full-length *xynA* gene from Rt69B.1. Sequencing of the Rt69B.1 *xynA* PCR products showed that the Rt8B.4 and Rt69B.1 *xynA* genes shared 98% sequence identity.

Consensus PCR analysis of the Rt69B.1 family 11 xylanase genes

A total of six family 11 xylanase consensus fragments (GXCFs), which were amplified from Rt69B.1 using the xynGF/xynGR consensus primers, were sequenced. These sequences were identical. Subsequent analysis using the new GF/newGR consensus primers failed to identify any additional GXCF species. Following PAGE, the newGF/newGR GXCFs migrated as a single band, and direct sequencing of these fragments yielded a single sequence species. Therefore, it was concluded that Rt69B.1 possessed only a single family 11 xylanase gene, which was named *xynD*.

Genomic walking of the Rt69B.1 *xynB* and *xynC* family 10 xylanase genes

Genomic walking of the Rt69B.1 *xynB* and *xynC* genes involved the amplification of specific gene fragments upstream and downstream of the *xynB* and *xynC* FXCF regions. Consequently, it was necessary to design specific *xynB* and *xynC* genomic walking primers that bound at unique positions within the respective *xynB* and *xynC* FXCF regions. Analysis of the aligned Rt69B.1 *xynA*, *xynB*, and *xynC* FXCF sequences revealed a central region of relatively high sequence variability where unique *xynB* and *xynC* genomic walking primers could be designed (Fig. 1). To cover the 5′- and 3′-ends of the genes encoding single-domain family 10 xylanases, it was necessary to 'walk' at least 350 bp upstream and 650 bp downstream of the typical FXCF region.

Rt69B.1 *xynB* GWPCR

Complete nucleotide sequencing of the Rt69B.1 *xynB* gene was achieved through an upstream walk of 1864 bp (via the xynBR and xynBR2 walking primers; Fig. 2) and a downstream walk of 3100 bp (via the xynBF, xynBF2, and xynBF3 walking primers). The final *xynB* sequence was 5177 bp in length, and contained a complete *xynB* open reading frame of 4788 bp, which encoded a XynB peptide of 1595 amino acid residues. Unexpectedly, the xynBR2 walking primer bound at two different positions along the *xynB* gene because of the chance placement of this primer within an intragenic repeat region. As a consequence, two distinct products of consistent molecular weight difference were amplified following PCR with xynBR2 and the appropriate linker primer. However, these products could be readily distinguished according to their upstream sequences and thus were included in the *xynB* sequence assembly project.

Rt69B.1 *xynC* GWPCR

Complete sequencing of the Rt69B.1 *xynC* gene was achieved by an upstream walk of 1334 bp (via the xynCR, xynCR2, and xynCR3 walking primers; Fig. 2), and a downstream walk of 3964 bp (via the xynCF, xynCFD, xynCFD2, and xynCFD3 walking primers). The final Rt69B.1 *xynC* sequence was 5437 bp in length and contained a complete *xynC* open reading frame of 5340 bp that encoded a putative XynC peptide of 1779 residues.

Analysis of the Rt69B.1 XynB and XynC peptide sequences

The Rt69B.1 XynB and XynC xylanases are both highly modular enzymes, and are composed of at least 10 and 8

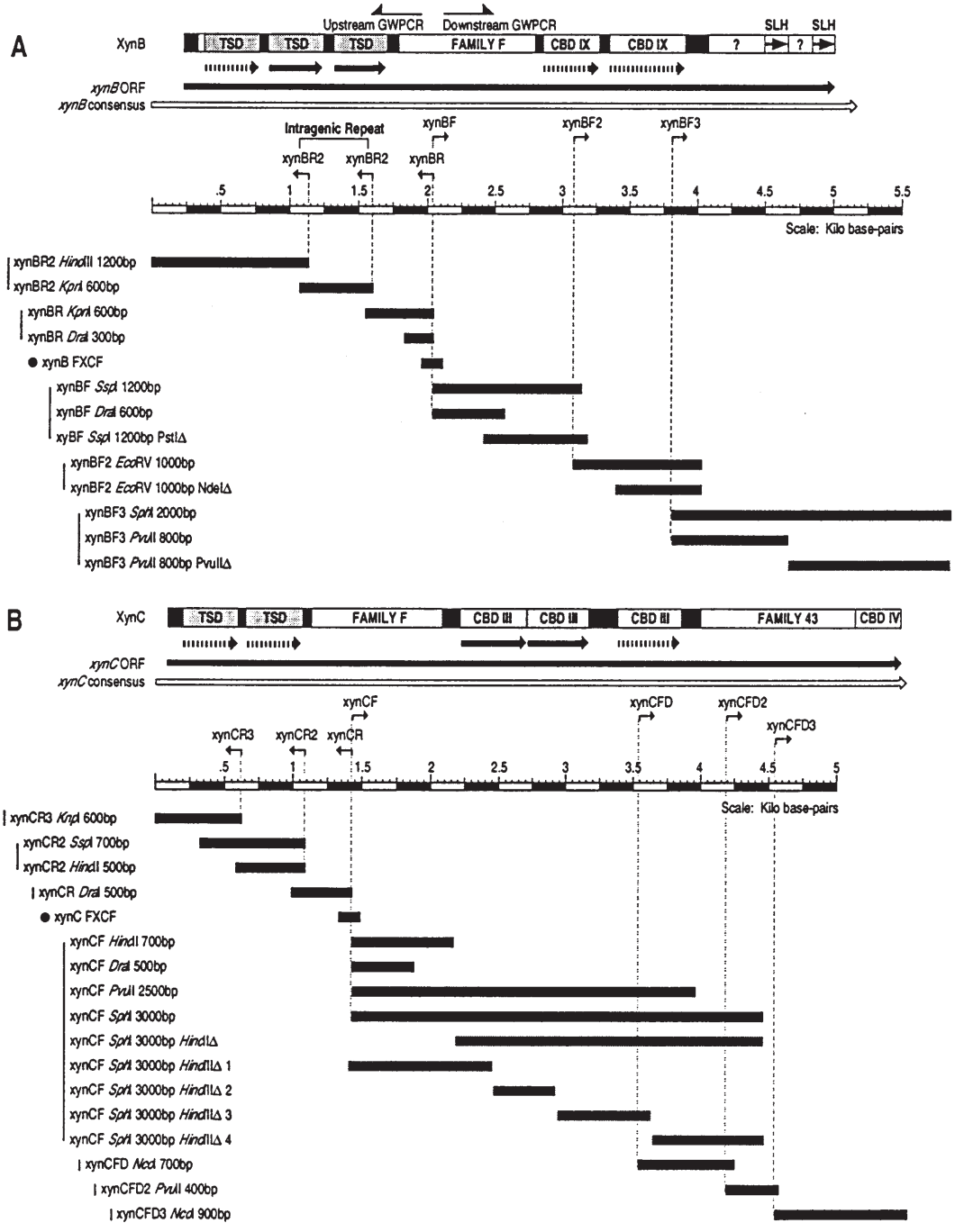

Fig. 2A,B. Overview of the genomic walking PCR of the Rt69B.1 *xynB* and *xynC* family 10 xylanase genes. **A** GWPCR of Rt69B.1 *xynB*. The following features are indicated, from the *top* to the *bottom* of the panel. XynB domain architecture (key: TSD, "thermostabilizing domain; CBD IX, family IX cellulose-binding domain; SLH, S-layer homology domain"; ?, domain of unknown function); intragenic repeat regions (low homology repeats, *dashed*; high homology repeats, *solid*); *xynB* open reading frame (*black arrow*); *xynB* sequence generated from the *xynB* GWPCR products (*white arrow*); *scale bar*, in divisions of 50 bp, labeled in kilobase pairs, and showing the respective positions of the *xynB* genomic walking primers; positions of the upstream and downstream *xynB* GWPCR products and the *xynB* FXCF (*boxed*). Individual fragments are labeled according to the walking primer and Rt69B.1 restriction fragment linker library used, with their size in base pairs. Subcloned fragments prepared from the larger GWPCR products are shown *shaded* and are labeled according to the restriction endonuclease used for their preparation. **B** GWPCR of Rt69B.1 *xynC*

discrete domains, respectively (Fig. 2). Furthermore, at 1595 and 1779 residues in length, XynB and XynC are the largest xylanases reported so far. The XynB xylanase is composed of a central family 10 xylanase domain, bordered by three N-terminal 'thermostabilizing' domains and tandem C-terminal family IX cellulose-binding domains. Additionally, a pair of S-layer homology domains (also called S-layer repeats, or SLRs; Liu et al. 1996) separating two unrelated domains of unknown function are present at the C-terminus of XynB.

The second and third XynB TSDs are 82% identical, and collectively share 32% sequence identity to the N-terminal-most TSD. Likewise, the tandem XynB CBD-IXs share 31% identity. The XynB SLH domains, which share 35% sequence identity, are so named according to their homology to the conserved domains of approximately 50 to 70 residues present in S-layer and S-layer-associated proteins of thermophiles (Liu et al. 1996). Directly repeated SLH domains are also present at the C-termini of the family 10 xylanases from *Clostridium thermocellum* (XynX, 3 copies; Pack and Jung 1991), *Thermoanaerobacterium saccharolyticum* B6A-RI (XynA, 2 copies; Lee et al. 1993), and *Thermoanaerobacterium* strain JW/SL-YS 485 (XynA, 3 copies; Liu et al. 1996). The SLH domains from the latter xylanase appear to be responsible for the adherence of enzyme to the *Thermoanaerobacterium* cell wall.

Similar to XynB, the N-terminal region of XynC is composed of a central family 10 xylanase domain with tandem N-terminal TSDs (Fig. 2). However, the C-terminal region of XynC is unrelated in architecture to XynB, and is composed of two directly repeated family III CBDs (CBD-III1 and CBD-III2), followed by a prominent proline–threonine-rich linker peptide, a third family III CBD (CBD-III3), a family 43 β-glycanase domain, and a C-terminal family IV CBD (CBD-IV). Hence, Rt69B.1 XynC appears to be a bifunctional enzyme that is similar in architecture to the multidomain–bifunctional β-glycanases from *Cs. saccharolyticus* (CelA, CelB, CelC, and ManA).

The XynC TSDs show 27% sequence identity and XynC CBD-III1 and CBD-III2 (97% identical) show 49% sequence identity to XynC CBD-III3. The XynC family 10 xylanase domain and downstream CBD-III share 72% sequence homology with the respective region from *Cs. saccharolyticus* CelB (Saul et al. 1990). Furthermore, the C-terminal region of XynC, encompassing the family 43 β-glycanase domain and CBD-IV, is 89% identical in sequence to the C-terminal domains of *Cs. saccharolyticus* XynF (Fig. 3). Last, the C-terminal region of XynC is also homologous to the two N-terminal domains of the *Bacillus polymyxa* XynD β-glycanase, which is composed of an N-terminal family 43 β-glycanase domain, a central CBD-IV, and an additional C-terminal domain (Gosables et al. 1991).

Comparison of the Rt69B.1 XynA, XynB, and XynC family 10 xylanases

The Rt69B.1 XynA, XynB, and XynC xylanases are members of a subfamily of enzymes, described here as the TSD-IX subfamily, which are related in structure to the XynA family 10 xylanase from *Thermotoga maritima*, composed of a family 10 xylanase domain bordered by tandem N-terminal TSDs and tandem C-terminal CBD-IX domains (Winterhalter et al. 1995). Multiple sequence analysis of the catalytic domains of the TSD-IX subfamily xylanases separates these enzymes into two predominant clusters. One of these main TSD-IX clusters is composed of the family 10 xylanases from *T. maritima* (XynA), *Thermotoga neopolitana* (XynA), *Cs. saccharolyticus* (XynE), and the three Rt69B.1 family 10 xylanases; the second main cluster consists of the family 10 xylanases from a *Thermoanaerobacterium* species (XynA), *Thermoanaerobacterium saccharolyticum* (XynA), and *Clostridium thermocellum* (XynX). These latter group of enzymes

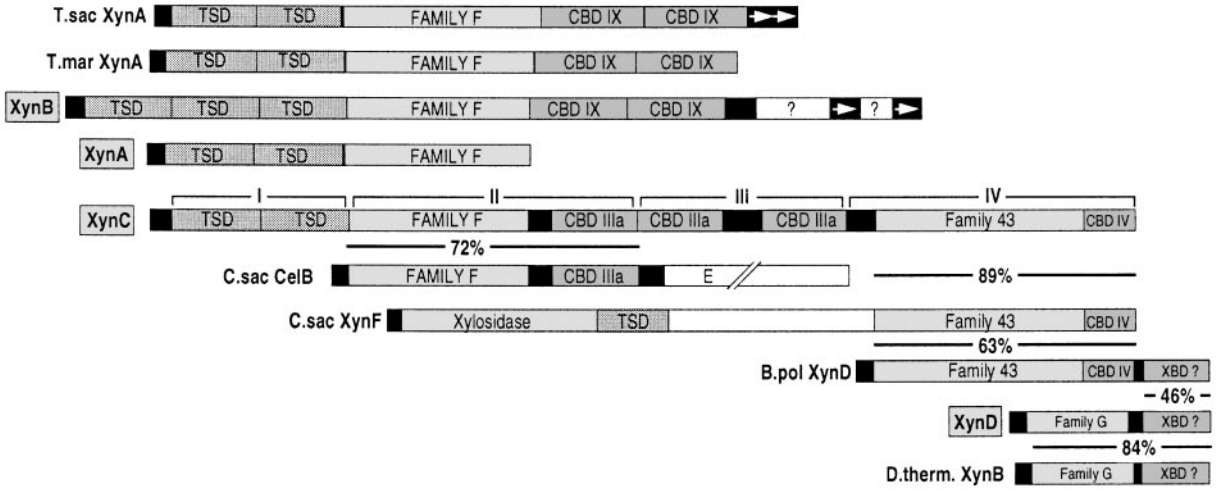

Fig. 3. Architectural and sequence homologies between the Rt69B.1 family 10 (XynA, XynB, and XynC) and family 11 (XynD) xylanases. From *top* to *bottom*, showing *Thermoanaerobacterium saccharolyticum* XynA, *Thermotoga maritima* XynA, Rt69B.1 XynB, Rt69B.1 XynA, Rt69B.1 XynC, *Caldicellulosiruptor saccharolyticus* CelB, *Cs. saccharolyticus* XynF, *Bacillus polymyxa* XynD, Rt69B.1 XynD, and *Dictyoglomus thermophilum* XynB. Key: TSD, thermostabilizing domain; CBD IX, family IX cellulose-binding domain; ?, domain of unknown function; E, endoglucanase domain (truncated in figure by size constraints); family 43, family 43 β-glycanase domain (reported xylosidase/arabinofuranosidase activities); CBDIV, family VI CBD; XBD?, possible xylan-binding domain. Repeated SLH (S-layer homology) domains are indicated by *white arrowheads*; the interdomain linker peptides are indicated by *black boxes*. Percent homologies between related domains are indicated (relevant to the respective Rt69B.1 xylanase). The four putative Rt69B.1 *xynC* gene segments fused by evolutionary domain-shuffling mechanisms are indicated

appear to be of common origin and possess directly repeated C-terminal SLH domains (Fig. 3).

The N-terminal regions of Rt69B.1 XynA, XynB, and XynC are architecturally identical and are also related in sequence. The N-terminal-most and C-terminal-most TSDs from each xylanase can be segregated into distinct subfamilies, with the exception of the second TSD from Rt69B.1 XynA. This observed TSD distribution pattern appears to be consistent throughout the TSD-IX subfamily, and nearly all the TSDs from the enzymes listed here can be segregated into one of two subfamilies comprising the first and second TSDs from each enzyme.

The xylanase domains of Rt69B.1 XynA, XynB, and XynC are closely related in sequence, with XynA–XynB showing 60% identity, XynA–XynC 63% identity, and XynB–XynC 60% identity. All three xylanase domains start at the conserved bacterial family 10 xylanase motif PSL and end on the consensus W(AS)(IL)(IV)(ED)P(ST)V(LV)P, which is also present with varying degrees of homology at the C-termini of other bacterial family 10 xylanase domains. The XynA and XynC family 10 xylanase domains are 329 residues in length, and the XynB domain is slightly longer at 340 residues. The length variations within the Rt69B.1 XynA, XynB, and XynC xylanase domains can be mapped to several of the variable loop regions that partition the alternating beta-strand and alpha-helix motifs (Derewenda et al. 1994; Harris et al. 1994; White et al. 1994).

Genomic walking of the Rt69B.1 *xynD* family 11 xylanase gene

The genomic walking of the Rt69B.1 family 11 xylanase gene (*xynD*) was simplified to a large extent by the presence of only a single family 11 xylanase gene in Rt69b.1. As a consequence, the *xynD* genomic walking primers could be placed freely within the 300-bp *xynD* GXCF consensus region, which allowed maximal novel sequence information to be obtained from the *xynD* GWPCR products. Furthermore, because family 11 xylanase domains are exclusively located at the N-terminus of multidomain enzymes, only a short upstream walk of approximately 300 bp was required to cover the 5′-ends of family 11 xylanase genes; a downstream walk of only 50 bp was sufficient to cover the 3′-end of single-domain family 11 xylanase genes.

The genomic walking of Rt69B.1 *xynD* was achieved in a single upstream walk of 335 bp via the xynDR walking primer and a single downstream walk of 621 bp via the xynDF walking primer (Fig. 4). The resulting 1272-bp *xynD* sequence contained a complete *xynD* open reading frame of 1086 bp, which encoded a XynD peptide of 361 residues.

Analysis of the Rt69B.1 XynD peptide sequence

The Rt69B.1 XynD family 11 xylanase is composed of an N-terminal family 11 xylanase domain, a central 20-residue serine- and glycine-rich interdomain linker peptide, and a C-terminal domain of 150 residues that shows high homology to the C-terminal domain of the XynY family 11 xylanase from a species of *Bacillus* (Yu et al. 1993) and, remarkably, the C-terminal domain from *B. polymyxa* XynD. This latter homology is intriguing, given that the remaining N-terminal region of *B. polymyxa* XynD is homologous in sequence and architecture to the C-terminal end of Rt69B.1 XynC (see Fig. 3).

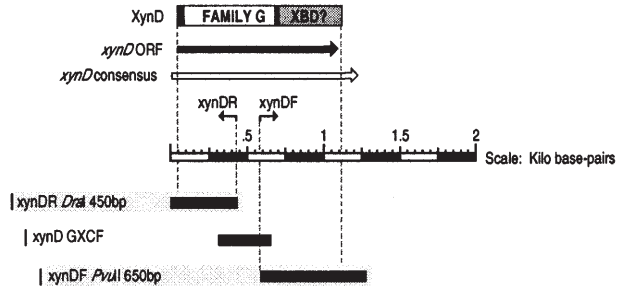

Fig. 4. Genomic walking polymerase chain reaction (GWPCR) of the Rt69B.1 *xynD* family 11 xylanase gene. Indicated (from *top* to *bottom*) are the XynD peptide structure, showing N-terminal family 11 xylanase domain, linker peptide, and possible C-terminal binding domain; *scale bar*, showing positions of the *xynD* GWPCR primers; and forward and reverse *xynD* GWPCR products

The XynD xylanase is identical in architecture to the XynB family 11 xylanase from *Dictyoglomus thermophilum* strain Rt46B.1 (Morris et al. 1998), and these two enzymes show 85% sequence identity. The most significant difference between the Rt69B.1 XynD and Rt46B.1 XynB sequences is in the interdomain linker region, which is 10 residues shorter in XynB.

Retrospective analysis of the *Caldicellulosiruptor* strain Rt8B.4 xylanase gene repertoire

Given the close phylogenetic relationship between Rt69B.1 and Rt8B.4 and the extremely high homology between the *xynA* genes from these strains, it was likely that Rt8B.4 would possess homologs of the Rt69B.1 *xynB*, *xynC*, and *xynD* genes. Following PAGE through 6% acrylamide gels, the FXCFs amplified from Rt8B.4 could be resolved into three distinct bands of similar sizes to the Rt69B.1 FXCFs (data not shown). Subsequently, four Rt8B.4 FXCFs were sequenced, which segregated with high homology with the Rt69B.1 *xynB* and *xynC* sequences following multiple sequence analysis. Therefore, it was concluded that Rt8B.4, like Rt69B.1, possessed three family 10 xylanase genes. In support for this conclusion, gene fragments could be amplified from Rt8B.4 using the Rt69B.1 *xynB* (xynBN and xynBC) and *xynC* (xynCN and XynCC) PCR primers.

Curiously, no GXCFs could be amplified from Rt8B.4 using either the xynGF–xynGR or the newGF–newGR consensus primers. Similarly, no fragments could be amplified from Rt8B.4 using the Rt69B.1 xynD PCR primers (xynDN and xynDC). Hence, Rt8B.4 did not appear to posses any family 11 xylanase genes. The type strain *Caldicellulosiruptor saccharolyticus* also does not contain any family 11 xylanase genes (unpublished data, our laboratory).

Preliminary characterization of the Rt69B.1 XynA, XynB, XynC, and XynD xylanase domains

Recombinant pJLA602 expression plasmids, which encoded the catalytic domains of XynA, XynB, XynC, and XynD, were prepared to compare the basic characteristics of these enzymes. The XynA, XynB, and XynC catalytic domains showed optimal activity over the pH range of 6.0–6.5. At optimal pH, XynA, XynB, and XynC released reducing sugars from oat spelts xylan at the highest rate at 65°C, 70°C, and 65°C, respectively (during a 10-min incubation). Under similar assay conditions the XynD catalytic domain showed optimal activity at pH 5.5 and 70°C.

Discussion

The endoxylanase genes of cellulolytic microorganisms are typically identified using traditional procedures that involve the screening of genomic DNA libraries for the expression of endoxylanase activity. Having identified the appropriate recombinant(s), an extensive subcloning regimen is then required to pinpoint and sequence the gene(s) of interest. We have presented here an alternative approach to the identification and sequencing of family 10 and 11 xylanase genes that involves consensus PCR (step one, xylanase gene identification), followed by genomic walking PCR (step two, xylanase gene sequencing). In theory, through the use of such two-step PCR techniques, it is possible to identify every individual xylanase gene in an organism under investigation and, subsequently, to specifically sequence the gene(s) of interest. The data presented here show that the theoretical potential of this two-step PCR technique is attainable, at least with respect to the identification and sequencing of multiple family 10 xylanase genes.

An essential requirement of this two-step PCR method is the presence of sequence and length heterogeneity within the fragments amplified by the xylanase consensus primers. Such heterogeneity not only allows for accurate enumeration of the xylanase gene repertoires via sequencing and PAGE but also enables selective GWPCRs to be performed. The FXCFs and GXCFs amplified by the consensus PCR primers described here both incorporate regions that can vary in both sequence and length; however, the variation is considerably higher within the FXCFs. In some cases, the higher degree of sequence conservation within the GXCFs may make it difficult to design specific walking primers for genomes possessing multiple family 11 xylanase genes.

As a result of the high degree of sequence variation between the Rt69B.1 *xynA*, *xynB*, and *xynC* FXCFs, it was relatively straightforward to design specific *xynB* and *xynC* genomic walking primers (and *xynA* primers should they have been required). It is noteworthy that continuing specific GWPCRs could be performed from the subsequent *xynB* and *xynC* walking primers, despite the homologies that existed both within and between these genes. The possibility of designing a walking primer within an intragenic or intergenic repeat is a potential problem with β-glycanase genes because of the abundance of duplicated domains within the encoded peptides. However, a genomic walking primer inadvertently so designed would be unlikely to lead to sequencing errors, simply because in most instances these would consistently yield multiple products following GWPCR and consequently prompt a redesign of the genomic walking strategy being implemented. Alternatively, the GWPCR products amplified from repeated genomic walking primers may be differentiated according to their immediate upstream or downstream sequences, as seen in this present study with the xynBR2 GWPCR products.

The results from the Rt69B.1 *xynB* and *xynC* GWPCRs highlight some important considerations that should be addressed during the genomic walking of family 10 and family 11 xylanase genes. First, the largest possible GWPCR products should be sequenced, both to reduce the number of genomic walking steps required and to increase the chance of amplifying fragments that encompass any upstream or downstream repeat sequences. Second, before the design of new walking primers, the sequences of the GWPCR products should be carefully examined for the presence of typically repetitive sequences, such as those encoding interdomain linker peptides, or the conserved regions of catalytic domains, CBDs, or TSDs. Last, where possible, the GWPCRs of multiple xylanase genes should be performed concurrently to allow direct sequence comparisons to be made between each gene and therefore reduce the chance of designing walking primers within intergenic repeats.

In view of the high degree of sequence homology between the family 10 xylanase domains of Rt69B.1 XynA, XynB, and XynC and the similarities in the N-terminal architectures of these enzymes, it appears that the *xynA*, *xynB*, and *xynC* genes have arisen through the duplication of an ancestral xylanase gene that encoded in its minimal form a family 10 xylanase domain with duplicated N-terminal TSDs. The differences over the C-terminal regions of XynA, XynB, and XynC are presumably the result of the 'domain-shuffling' mechanisms central to the evolution of microbial glycosyl hydrolases (Gilkes et al. 1991). The sequence and architectural homologies observed between the xylanases from Rt69B.1 and assorted β–glycanases from other cellulolytic–hemicellulolytic bacterial strains provide a remarkable example of these domain-shuffling processes. At least four distinct gene segments can be identified within Rt69B.1 *xynC* on the basis of isolated homologies between XynC and the CelB and XynF enzymes from *Cs. saccharolyticus* (Fig. 4). Similarly, the two C-terminal domains of XynC can combine with the C-terminal domain of XynD to form an enzyme of architecture identical to *Bacillus polymyxa* XynD. It is noteworthy that the joint in the *B. polymyxa* XynD peptide sequence signifying the end of homology to XynC and the commencement of homology to XynD is continuous, which suggests that the C-terminal domains of XynC and XynD arose through the splitting of an ancestral *B. polymyxa xynD*-like gene.

Acknowledgments This work was supported in part by grants from the University of Auckland Research Committee, and the Foundation for Research, Science and Technology, New Zealand.

References

Clarke J, Davidson K, Gilbert HJ, Fontes CMGA, Hazlewood GP (1996) A modular xylanase from mesophilic *Cellulomonas fimi* contains the same cellulose-binding domain and thermostabilising domain as xylanases from thermophilic bacteria. FEMS Microbiol Lett 139:27–35

Derewenda U, Swenson L, Green R, Morosoli R, Shareck F, Kluepfel D, Derewenda ZS (1994) Crystal structure, at 2.6-Å resolution, of the *Streptomyces lividans* xylanase A, a member of the F family of beta-1,4-D-glycanases. J Biol Chem 269:20811–20814

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dwivedi PP, Gibbs MD, Saul DJ, Bergquist PL (1996) Cloning, sequencing and overexpression in *Escherichia coli* of a xylanase gene, *xynA* from the thermophilic bacterium *Caldicellulosiruptor* Rt8B.4. Appl Microbiol Biotechnol 45:86–93

Fontes CM, Hazlewood GP, Morag E, Hall J, Hirst BH, Gilbert HJ (1995) Evidence for a general role for non-catalytic thermostabilising domains in a xylanase from thermophilic bacteria. Biochem J 307:151–158

Gibbs MD, Saul DJ, Lüthi E, Bergquist PL (1992) The β-mannanase from "*Caldocellum saccharolyticum*" is part of a multidomain enzyme. Appl Environ Microbiol 58:3864–3867

Gilkes NR, Henrissat B, Kilburn DG, Miller RCJ, Warren RAJ (1991) Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol Rev 55:303–315

Gosables MJ, Perez-Gonzalez JA, Gonzales R, Navarros R (1991) Two beta-glycanase genes are clustered in *Bacillus polymyxa*: molecular cloning, expression and sequence analysis of genes encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase. J Bacteriol 173:7705–7710

Harris GW, Jenkins JA, Connerton I, Cummings N, Leggio LL, Scott HJ, Hazlewood GP, Laurie JI, Gilbert HJ, Pickersgill RW (1994) Structure of the catalytic core of the family F xylanase from *Pseudomonas fluorescens* and identification of the xylopentaose-binding sites. Structure 2:1107–1116

Henrissat B, Bairoch A (1995) New families in the classification of gylcosyl hydrolases based on amino acid sequence similarities. Biochem J 293:781–788

Lee Y, Lowe SE, Zeikus GJ (1993) Gene cloning, sequencing and biochemical characterisation of endoxylanase from *Thermoanaerobacterium saccharolyticum* B6A-RI. Appl Environ Microbiol 59:3134–3137

Lever M (1973) Colorimetric and fluorometric carbohydrate determination with *p*-hydroxybenzoic acid hydrazide. Biochem Med 7:274–281

Liu S, Gherardini FC, Matuschek M, Bahl H, Weigel J (1996) Cloning, sequencing and expression of the gene encoding a large S-layer-associated endoxylanase from *Thermoanaerobacterium* sp. strain JW/SL-YS 485 in *Escherichia coli*. J Bacteriol 178:1539–1547

Love DR, Streiff MB (1987) Molecular cloning of a β-glucosidase gene from an extremely thermophilic anaerobe in *E. coli* and *B. subtilis*. Bio/Technology 5:384–387

Lüthi E, Bergquist PL (1990a) A β-xylosidase from the thermophile *Caldocellum saccharolyticum* expressed in *Escherichia coli*. FEMS Microbiol Lett 67:291–294

Lüthi E, Bhana NJ, Bergquist PL (1990b) Overproduction of an acetylxylan esterase from the extreme thermophile "*Caldocellum saccharolyticum*" in *Escherichia coli*. Appl Microbiol Biotechnol 34:214–219

Lüthi E, Bhana NJ, Bergquist PL (1990c) Xylanase from the extremely thermophilic bacterium "*Caldocellum saccharolyticum*": overexpression of the gene in *Escherichia coli* and characterisation of the gene product. Appl Environ Microbiol 56:2677–2683

Lüthi E, Bhana NJ, Grayling RA, Love DR, Bergquist PL (1991) Cloning, sequence analysis and expression in *Escherichia coli* of a gene coding for a β-mannanase from the extremely thermophilic bacterium "*Caldocellum saccharolyticum*". Appl Environ Microbiol 57:694–700

Morris DD, Reeves RA, Gibbs MD, Saul DJ, Bergquist PL (1995) Correction of the β-mannanase domain of the *celC* pseudogene from *Caldicellulosiruptor saccharolyticus* and activity of the gene product on kraft pulp. Appl Environ Microbiol 61:2262–2269

Morris DD, Gibbs MD, Chin CW, Koh MH, Wong KKY, Allison RW, Nelson PJ, Bergquist PL (1998) Cloning and expression of the *xynB* gene from *Dictyoglomus thermophilum* strain Rt46B.1, and characterization of the gene-pruduct on kraft pulp. Appl Environ Microbiol 64:1759–1764

Pack MY, Jung KH (1991) Nucleotide sequence of exo-beta-1,4-glucanase gene of *Clostridium thermocellum*. Unpublished GenBank accession, M67438

Rainey FA, Ward NL, Morgan HW, Toalster R, Stackebrandt E (1993) Phylogenetic analysis of anaerobic thermophilic bacteria: aid for their reclassification. J Bacteriol 175:4772–4779

Rainey FA, Donnison AM, Janssen PH, Saul DJ, Rodigo AG, Bergquist PL, Daniel RM, Stackebrandt E, Morgan HW (1994) Description of *Caldicellulosiruptor saccharolyticus* gen. nov. sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. FEMS Microbiol Lett 120:263–266

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, vol 3, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Saul DJ, Williams LC, Grayling RA, Chamley LW, Love DR, Bergquist PL (1990) *celB*, a gene coding for a bifunctional cellulase from the extreme thermophile "*Caldocellum saccharolyticum*." Appl Environ Microbiol 56:3117–3124

Schauder B, Blöcker H, Frank R, McCarthy JEG (1987) Inducible expression vectors incorporating the *Escherichia coli atpE* transcriptional initiation region. Gene 52:279–283

Te'o VSJ, Saul DJ, Bergquist PL (1996) *celA*, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. Appl Microbiol Biotechnol 43:291–296

Teather RM, Wood PJ (1987) Use of Congo Red polysaccharide interaction in enumeration and characterisation of cellulolytic bacteria from bovine rumen. Appl Environ Microbiol 43:777–780

Tomme P, Warren RAJ, Miller RC, Kilburn DG, Gilkes NR (1994) Cellulose binding domains: classification and properties. In: Saddler JN, Penner MH (eds) Enzymatic degradation of insoluble carbohydrates. American Chemical Society Symposium Series 618. American Chemical Society Press, San Diego, pp 142–161

White A, Withers SG, Gilkes NR, Rose DR (1994) Crystal structure of the catalytic domain of the β-1,4-glycanase Cex from *Cellulomonas fimi*. Biochemistry 33:12546–12552

Winterhalter C, Heinrich P, Candussio A, Wich G, Liebl W (1995) Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*. Mol Microbiol 15:431–444

Yu JH, Park YS, Yum DY, Kim JM, Kong IS, Bai DH (1993) Nucleotide sequence and analysis of a xylanase gene (*xynS*) from alkalitolerant *Bacillus* sp. YA-14 and comparison with other xylanases. J Microbiol Biotechnol 3:139–145

Zverlov V, Piotukh K, Dakhova O, Velikodvorskaya G, Borriss R (1996) The multidomain xylanase A of the hyperthermophilic bacterium *Thermotoga neapolitana* is extremely thermoresistant. Appl Microbiol Biotechnol 45:245–247