Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria

Carlos M. G. A. FONTES,*† Geoffrey P. HAZLEWOOD,‡ Ely MORAG,§ Judith HALL,* Barry H. HIRST† and Harry J. GILBERT*||

*Department of Biological and Nutritional Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, U.K., †Department of Physiological Sciences, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, U.K., ‡Department of Cellular Physiology, The Babraham Institute, Babraham, Cambridge CB2 4AT, U.K., and §Department of Biophysics, The Weizmann Institute of Science, Rehovot, Israel

A genomic library of *Clostridium thermocellum* DNA constructed in λ ZAPII was screened for xylanase-expressing clones. Cross-hybridization experiments revealed a new xylanase gene isolated from the gene library, which was designated *xynY*. The encoded enzyme, xylanase Y (XYLY), displayed features characteristic of an endo- β 1,4-xylanase: the enzyme rapidly hydrolysed oat spelt, wheat and rye arabinoxylans and was active against methylumbelliferyl- β -D-cellobioside, but did not hydrolyse any cellulosic substrates. The pH and temperature optima of the enzyme were 6.8 and 75 °C respectively, and the recombinant XYLY, expressed by *Escherichia coli* had a maximum M_r of 116000. The nucleotide sequence of *xynY* contained an open reading frame of 3228 bp encoding a protein of predicted M_r 120 105. The encoded enzyme contained a typical N-terminal 26-residue signal peptide, followed by a 164 amino acid sequence, designated domain A, that was not essential for catalytic activity. Downstream of domain A was a 351-residue xylanase Family F catalytic domain, followed by a 180-residue sequence that exhibited 28 % sequence identity with a thermostable domain of *Thermoanaerobacterium*

saccharolyticum xylanase A. The C-terminal portion of XYLY comprised the 23-residue duplicated docking sequence found in all other *C. thermocellum* plant cell wall hydrolases that are constituents of the bacterium's multienzyme complex, termed the cellulosome, followed by a 286-residue domain which exhibited 32 % sequence identity with the N-terminal region of *C. thermocellum* xylanase Z. The enzyme did not contain linker sequences found in other *C. thermocellum* plant cell wall hydrolases. Analysis of truncated forms of XYLY and hybrid proteins, comprising segments of XYLY fused to the *E. coli* maltose binding domain, confirmed that XYLY contained a central catalytic domain and an adjacent thermostable domain. The C-terminal domain did not bind to cellulose or xylan. Western blot analysis using antiserum raised against XYLY showed that the xylanase was located in the cellulosome and did not appear to be extensively glycosylated. The non-catalytic domains of XYLY are discussed in relation to the general stability of thermophilic xylanases.

INTRODUCTION

Clostridium thermocellum is one of the most active cellulose-degrading microorganisms to have been described. The cellulase activity of this bacterium is vested in a large multienzyme complex that, in the early stages of growth on cellulose, is associated with the cell wall (Lamed and Bayer, 1988). The multienzyme complex contains numerous catalytic subunits which interact with a scaffolding protein, termed CipA, that, in addition to assembling the enzymes into a high molecular mass complex (Tokatlidis et al., 1991), also binds tightly to cellulose (Salamitou et al., 1992). CipA comprises nine almost identical 166-residue domains (Gerngross et al., 1993) and a discrete high-affinity cellulose-binding domain (CBD; Poole et al., 1993). The catalytic subunits, which are part of the cellulosome, each contain a highly conserved non-catalytic 23-residue duplicated docking sequence that binds to one of the CipA 166-residue domains (Tokatlidis et al., 1991). To date, seven endoglucanases and one cellobiohydrolase have been shown to be components of the cellulosome (Beguín and Aubert, 1994). It remains to be elucidated precisely how the assembly of the cellulases into the cellulosome elicits rapid hydrolysis of cellulose.

In addition to cellulose, *C. thermocellum* is capable of rapidly hydrolysing xylan, the major hemicellulose in several plant

species, although the bacterium does not utilize xylose, the final product of xylan hydrolysis, as a carbon source (Thomson, 1993). Although some *C. thermocellum* endoglucanases exhibit xylanase activity (Hall et al., 1988), the microorganism also expresses xylanases that exhibit little or no activity towards cellulose (Hazlewood et al., 1988). The primary structures and molecular architecture of numerous endoglucanases from *C. thermocellum* have been analysed, however, there is a paucity of information on the structure of xylanases from this bacterium. Xylanase Z (XYLZ) from *C. thermocellum* comprises a C-terminal Family F catalytic domain, two central non-catalytic domains, one of which contains the conserved 23-residue duplicated docking sequence (Grepinet et al., 1988), while the other contains a single copy of the duplicated sequence that constitutes a CBD in *Clostridium stercorarium* xylanase A (Sakka et al., 1993). The role of the N-terminal non-catalytic domain is unknown. It is apparent, therefore, that XYLZ is part of the cellulosome. A second *C. thermocellum* xylanase (xylanase X; XYLX), reported in the EMBL database with the accession number M67438, comprises 1120 amino acids, but the enzyme has not been subjected to functional/structural analysis. From the preceding discussion it is apparent that there are several unresolved questions relating to the xylanases of *C. thermocellum*. For example, although XYLZ is clearly part of the cellulosome,

it is unclear whether other xylanases from this bacterium are components of the multienzyme complex. Studies by Morag et al. (1990) have shown that there are multiple polypeptides within the cellulosome that display xylanase activity, but it is unclear how many of these enzymes are broad specificity cellulases and/or truncated forms of XYLZ. Several xylanases from other organisms contain discrete CBDs (Gilbert and Hazlewood, 1993). However, it remains to be established whether *C. thermocellum* xylanases, other than XYLZ, contain CBD homologues. Recently, a non-catalytic domain from *Thermoanaerobacterium saccharolyticum* xylanase A was shown to confer enhanced thermostability on the enzyme (Lee et al., 1993). How prevalent these thermostable domains are among xylanases from thermophilic microorganisms also remains to be elucidated.

To develop our understanding of xylan hydrolysis by thermophilic bacteria, in general, and by *C. thermocellum* in particular, we have analysed the structure/function relationships of a third xylanase from *C. thermocellum* defined as xylanase Y (XYLY). This Family F enzyme, which is a component of the cellulosome, contains the highly conserved *C. thermocellum* 23-residue duplicated docking sequence, in addition to a non-catalytic thermostable domain. The structure of XYLY is discussed in the general context of the molecular architecture of thermostable xylanases.

MATERIALS AND METHODS

Bacterial strains, vectors and culture media

C. thermocellum strain YS was cultured at 60 °C in medium CM3 containing cellobiose (Romaniec et al., 1987). *Escherichia coli* strains JM83 and XL1-Blue were cultured in Luria broth at 37 °C. Phages were propagated in *E. coli* grown on NZY medium. Transformants and recombinants of *E. coli* were selected by the inclusion of ampicillin (100 µg/ml) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (2 µg/ml) respectively. Detection of xylanase-expressing clones on solid media was as described by Black et al. (1994). To induce expression of genes under control of *lacZp*, *E. coli* strains were grown to mid-log phase, isopropyl β-D-thiogalactopyranoside was added to the medium at a final concentration of 100 µM and the bacteria were cultured for a further 3 h. The vectors used in this study were λZAPII, pBluescriptSK⁻ (Stratagene), pMAL-c2 (New England Biolabs), pCRTMII (Invitrogen Corporation), pMTL22p, pMTL23p, pMTL20p and pMTL21p (Chambers et al., 1988).

Screening of *C. thermocellum* genomic library

To screen a genomic library of *C. thermocellum*, DNA constructed in λZAPII as detailed previously (Poole et al., 1992), *E. coli* XL1-Blue, infected with recombinant phage (1.4×10^4 plaque-forming units) was plated onto NZY in a soft agar overlay. After 16 h incubation at 37 °C, the plaques which had formed were overlaid with molten agar (1.5% w/v) in 50 mM sodium phosphate/12 mM citric acid buffer, pH 6.5 (PC buffer) containing 0.25% soluble xylan. After 4 h incubation at 60 °C, plates were stained with Congo Red and destained with 1 M NaCl to reveal xylanase-positive clones which were excised and purified by replating on *E. coli* XL1-Blue.

Recombinant DNA methodology

Agarose gel electrophoresis, transformations, Southern hybridizations and the use of restriction enzymes and other nucleic acid modifying enzymes were as described by Sambrook et al. (1989). Plasmid DNA was isolated using Qiagen columns (Hybaid) and following the protocol recommended by the manufacturer. To sequence *xynY*, denatured plasmid DNA was

purified as described by Murphy and Kavanagh (1988) and subjected to dideoxy chain-termination sequencing (Sanger et al., 1980) using the Sequenase sequencing kit (USB). Overlapping sequences were generated by creating nested deletions of the *Clostridium* insert in pCF1 using exonuclease III/S1-nuclease. Sequences were compiled using the computer programs of Staden (1980).

To construct a hybrid protein containing the catalytic region of XYLY fused to the *E. coli* maltose-binding protein (MBP), the PCR was used to amplify the region of *xynY* encoding the catalytic domain of the xylanase. The primers 5'-CGCGAAG-CTTACCATAACCCGTAAGGAATGGC-3' and 5'-CGCG-GAGCTCTTAGGACTGAGGAATAATGGATGC-3' were employed to amplify the sequence between map positions 526 and 1626 (as defined in Figure 3). The reaction conditions for 30 cycles were as follows: annealing, 56 °C; elongation, 2 min at 72 °C; denaturation, 1 min at 96 °C. The amplified DNA was ligated into pCRTMII, excised on an *EcoRI* restriction fragment and cloned into the *EcoRI* site of pMAL-c2.

Assays and purification of the fusion proteins and XYLY

All cellulase and xylanase assays were performed at 60 °C unless otherwise stated. The buffers and substrates used to perform the cellulase and xylanase assays were as described by Hazlewood et al. (1990) and Kellett et al. (1990) unless otherwise stated. The source of enzyme was either purified fusion protein or cell-free extracts of selected *E. coli* strains, prepared by sonic disruption of the bacterial cells followed by removal of cell debris and insoluble material by centrifugation at 15000 *g* for 30 min at 4 °C. HPLC analysis of hydrolysis products was as described by Black et al. (1994). Quantification of 4-methylumbelliferyl β-D-cellobiosidase (MUCase) activity was as described by Hall et al. (1992) using a Shimadzu RF-1501 spectrofluorophotometer to detect the release of methylumbelliferone. Protein was measured as described by Sedmak and Grossberg (1977). SDS/PAGE was performed using the method of Laemmli (1970) and zymogram analysis of polypeptides in polyacrylamide gels was carried out as described by Malburg and Forsberg (1993). Antiserum raised against the catalytic domain of XYLY was used in Western-blot analysis of *C. thermocellum* cellulosomal proteins as described by Hall et al. (1993) using the enhanced chemiluminescence system (Amersham) to detect immunoreactive polypeptides.

The MBP-XYLY fusion proteins were purified as described by Black et al. (1994). To purify the catalytic domain of XYLY from the fusion protein, the purified hybrid polypeptide was treated with Factor Xa for 8 h at 37 °C and dialysed against 2×1000 vol. of 50 mM Tris/HCl, pH 8.0, for 24 h. The cleaved protein was then re-applied to the amylose affinity column and pure XYLY was recovered in the eluent. Cellulose-binding assays were performed as described by Kellett et al. (1990) using Avicel PH105 purchased from Serva.

Thermostability of MBP-XYLY

The thermostability studies of MBP-XYLY fusion proteins were conducted as follows: the purified proteins were incubated for 20 min at a range of temperatures, immediately placed on ice and then assayed for residual activity at 60 °C. In a second experiment the hybrid proteins were incubated at 70 °C for 60 min and at regular intervals aliquots were removed and assayed for xylanase activity at 60 °C. In the final experiment XYLY was assayed for xylanase activity at various temperatures. MUC was used as the substrate in the enzyme assays (Hall et al., 1992) and the reactions were allowed to proceed for 10 min before quantification of the product. In the thermostability experiments

described above, parallel experiments were performed: in one set of experiments the actual enzyme units were the same for each fusion protein, while in the second set of experiments the protein concentration was the same for each hybrid polypeptide. Both sets of experiments gave very similar results.

RESULTS

Isolation of *xynY*

Approximately 14000 recombinants of a genomic library of *C. thermocellum* DNA, constructed in λ ZAPII, were screened for xylanase-expressing clones. The clostridial inserts from the 41 positive clones were excised into pBluescript SK⁻ and the degree of similarity between recombinant plasmid DNA from each of the clones and the previously described *C. thermocellum* xylanase gene, *xynZ*, was determined. Clone pCF1 expressed an active xylanase which was clearly different to the *xynZ* product, XYLZ; the xylanase gene contained in pCF1 and its encoded product were designated *xynY* and XYLY respectively. A restriction map of the clostridial insert in pCF1 is shown in Figure 1. To evaluate whether *xynY* was reiterated throughout the clostridial genome, *C. thermocellum* genomic DNA was subjected to Southern hybridization using the clostridial insert from pCF2 as the probe. The data (Figure 2) revealed that *xynY* was present as a single copy within the *Clostridium* genome.

Properties of XYLY

The properties of XYLY were evaluated using cell-free extract of *E. coli* harbouring pCF1 as the source of the enzyme. Analysis of the substrate specificity of XYLY revealed that the enzyme hydrolysed arabinoxylans from oat spelt (10.8 i.u./mg protein), wheat (14 i.u./mg protein) and rye (16.5 i.u./mg protein), releasing primarily xylose and xylobiose. The initial stages of the hydrolysis of these substrates was associated with a rapid decrease in the viscosity of the polysaccharides (results not shown),

Figure 1 Restriction map of plasmids containing *xynY*

The positions of the cleavage sites for the restriction endonucleases ClaI (C), EcoRI (E), HindIII (H), PstI (P), SalI (Sa) and SmaI (S) are shown. The regions of *xynY* that encode the signal peptide (■), domain A (■), domain B (□), domain C (▨), domain D (▩) and domain E (▧) are indicated. Plasmids containing 5' truncated forms of *xynY* were cloned in-phase with the translational start codon of *lacZ'* in appropriate pMTL vectors (pCF2, pCF6, pCF7 and pCF8) or *malE* in pMAL-c2 (pCF3, pCF4 and pCF5) respectively. The *Clostridium* insert in pCF2 and pCF3 was generated by PCR as described in the Materials and methods section. The locations of the 5' and 3' nucleotides of truncated forms of *xynY*, encoded by pCF2-8, are shown in Figure 3. The capacity of the plasmids to direct expression of a functional xylanase is indicated by +/−.

Figure 2 Southern hybridization of *xynY* with *C. thermocellum* genomic DNA

C. thermocellum DNA restricted with EcoRI and SmaI (lane 1), SmaI (lane 2), EcoRI (lane 3) and HindIII (lane 4) was subjected to Southern hybridization using the *Clostridium* insert from pCF2 as the probe.

supporting the view that the enzyme has an endo-mode of action. XYLY exhibited no significant activity against either soluble or insoluble forms of cellulose, cellulo-oligosaccharides or other plant cell wall polysaccharides such as galactomannan, arabinan or galactan. However, the enzyme did display some activity against MUC (1.6 i.u./mg protein). These data suggest that the catalytic domain of XYLY displays the catalytic properties of a typical endo- β 1,4-xylanase and belongs to β -glycanase Family F (defined by Gilkes et al., 1991), as all enzymes in this group have the capacity to hydrolyse MUC. SDS/PAGE and zymogram analysis of cell-free extract derived from *E. coli* containing pCF1 revealed several polypeptides that exhibited xylanase activity, the largest of which had an M_r of 116000 (results not shown)

Nucleotide sequence of *xynY*

Deletion and subcloning of various restriction fragments from pCF1 located the region of *xynY* that encoded a functional xylanase between map positions 2.5 and 3.6 (Figure 1). To obtain the full sequence of *xynY*, the nucleotide sequence of the clostridial insert of pCF1, extending from map position 1.8 to 5.3, was determined in both strands. The data revealed a single, open reading frame of 3234 bp coding for a protein of M_r 120 105 (Figure 3). The proposed ATG translational start codon was preceded (6 bp) by the sequence AGGAGGT, which exhibits strong similarity to the ribosome-binding sequence most frequently found for *C. thermocellum* genes. The deduced primary sequence of the N-terminal region of XYLY conformed to a typical signal peptide comprising three lysine residues followed by a stretch of 22 predominantly small hydrophobic residues. Codon usage for *xynY* was similar to that of other *C. thermocellum* genes characterized to date.

Inspection of the primary structure of XYLY revealed several interesting features. The region designated domain B, extending from residue 249 to 514 (Figure 4), exhibited extensive sequence identity with the catalytic domain of other Family F xylanases (Figure 4). Strongest identity was observed with xylanase A from *Ruminococcus flavefaciens* (Zhang and Flint, 1992), xylanase 4 from *Caldocellum saccharolyticum* (Luthi et al., 1990), and xylanase B from *Butyrivibrio fibrisolvens* (Lin and Thomson,

-200
TAAGAACTTTAAACACCCCTTTATAAAAAACAAAGAATTACAGGCAATTATAGTGTA
-100
TGTGGATTTTAACTAAAAAGGAGGAATGTAATTCGTAATAGATATTATGATATAAT
TTGTTTAGAGCATGCTTAAGTTTATTTAAATTTAATTTATAAATTAAATTTAAATTTAA
+1
ATTTAAAGGAGGTGCTTTGAAAAACAAGAGAGTTTGGCAAAATAACGGCTCTTG
M K N K R V L A K I T A L V
100
GTATTGCTGGAGGTGTTTTTGTATTACCGTCAACATAAGTCAGCTATATGCTGATTAT
V L L G V F F V L P S N I S Q L Y A D Y
A
5'pCF6
GAAGTGGTTCATGACACTTTTGAAGTTAACTTTGACGGATGGTGAACCTGGGAGTCGAC
E V V H D T F E V N F D G W C N L G V D
200
ACATATTTAAACCGCAGTTGAAAATGAAGGAAACACGGTACAAGAGGTATGATGGTAAT
T Y L T A V E N E G N N G T R G M M V I
AATCGCTCCAGTGGAGTGCAGGTGCGTATTCGGAAGAGTTTCTATCTCGACGGTGGT
N R S S A S D G A Y S E K G F Y L D G G
300
GTAGAATACAAGTACAGTGTTTTGTAAAACACACGGGACCGGACCGGAACTTTCAAA
V E Y K Y S V F V K H N G T G T E T F K
400
CTTCTGTGTCCTATTGCGATTTCGGAACAGAAAGAAATAAGGAAGTAATTGCAACA
L S V S Y L D S E T E E E N K E V I A T
5'pCF7
AAGGATGTTGTGCGCGGAGAAATGGACTGAGATTTTCGGAATAACAAAGCACCAAACT
K D V V A G E W T E I S A K Y K A P K T
500
GCAGTGAATATTACTTTGTCAATTACACCGACAGCACTGTAGATTTTATTTTGACGAT
A V N I T L S I T T D S T V D F I F D D
5'pCF2-5
GTAAACATAACCCGTAAGGAATGGCTGAGGCAACACAGTATATGACGAAACCGTGTG
V T I T R K G M A E A N T V Y A A N A V
600
CTGAAAGATATGATGCAAACTATTTCAGAGTTGGTTCGGTACTTAACTCCGGAACGGTA
L K D M Y A N Y F R V G S V L N S G T V
B
700
AACAAATTCATCAATAAAGGCCTTGATTTTAAAGAGAGTTTAAAGTATTACCTGTGAAAT
N N S S I K A L I L R E F N S I T C E N
GAAATGAAGCCTGATGCCACACTGGTTCAATCAGGATCAACCAATACAAATATCAGGGTT
E M K P D A T L V Q S G S T N T N I R V
800
TCTCTTAATCGTGCAGCAAGTATTTTAACTCTGTGCACAAAATAATATAGCCGTCAGA
S L N R A A S I L N F C A Q N N I A V R
GGTCATACACTGGTTTGGCAGCCAGACACCTCAATGGTTTTTCAAAGACAATTTCCAG
G H T L V W H S Q T P Q W F F K D N F Q
900
GACAACGGAACTGGGTTTCCCAATCAGTTATGGACAGCGTTTGGAAAGCTACATAAAA
D N G N W V S Q S V M D Q R L E S Y I K
1000
AATATGTTTGTGAAATCCAAAGACAGTATCCGCTTTGAATCTTTATGCCTATGACGTT
N M F A E I Q R Q Y P S L N L Y A Y D V
GTAAATGAGGCAAGTATGATGCAAAACAGGACAGATATTATGGCGGGCGAGGGAA
V N E A V S D D A N R T R Y Y G G A R E
1100
CCTGGATACGGAAATGGTAGATCTCCATGGGTTTACAGATCTACGGAGACAATAATTTAT
P G Y G N G R S P W V Q I Y G D N K F I
5'pCF8
GAGAAAGCATTTACATATGCAAGAAAATATGCTCCGGCAATTTGAAGCTTTACTACAAC
E K A F T Y A R K Y A P A N C K L Y Y N
1200
GATTACAACGAATATTGGGATCATAAGAGAGACTGTATTGCCTCAATTTGTGCAAACTTG
D Y N E Y W D H K R D C I A S I C A N L
1300
TACAACAAGGGCTTGTCTTACGGGTGTGGGAATGCAGTCCCATATTATGCGGATATGAAT
Y N K G L L D G V G M Q S H I N A D M N
GGATTCTCAGGTATACAAAATTATAAGCAGCTTTGCAGAAATATATAAATATCGGTTGT
G F S G I Q N Y K A A L Q K Y I N I G C
1400
GATGTCCAATACCGAGCTTGATATTAGTACAGAAAACGGCAAAATTTAGCTTACAGCAG
D V Q I T E L D I S T E N G K F S L Q Q
CAGGCTGATAAATATAAGAGCTTTTCCAGGCAGCTGTGATATAACAGAACTCCAGC
Q A D K Y K A V F Q A A V D I N R T S S
1500
AAAGGAAAGGTTACGGCTGTCTGTGTATGGGACCTAATGACGCCAATCTTGGCTCGGT
K G K V T A V C V W G P N D A N T W L G
1600
TCACAAAATGCACCTCTTTTGTAAACGCAACATCAACGGAACCGGCATACAATGCG
S Q N A P L L F N A N N Q P K P A Y N A
3'pCF2-3
GTTGCATCCATTATTCCTCAGTCCGAATGGGGCGACGTAACATCCGGCCGGCGCGGA
V A S I I P Q S E W G D G N N P A G G G
1700
GGAGGAGGCAACCGGAAGAGCCGATGCAACCGGATATTATCATGACACTTTTGAA
G G G K P E E P D A N G Y Y Y H D T F E
C
GGAAGCGTAGGACAGTGGACAGCCAGAGGACCTGCGGAAGTTCTGCTTAGCGGAAGAACG
G S V G Q W T A R G P A E V L L S G R T
1800
GCTTACAAGGTTTCAAGTCACTCTTGGTAAGGAACCGTACGGCAGCATGGAACGGAGCA
A Y K G S E S L L V R N R T A A W N G A
1900
CAACGGGCGCTGAATCCAGAACGTTTGTCCCGGAACACATATTGTTTCAGCGTAGTG
Q R A L N P R T F V P G N T Y C F S V V
GCATCGTTTATTGAAGGTGCGTCTTCCACAACATTCTGCATGAAGCTGCAATACGTAGAC
A S F I E G A S S T T F C M K L Q Y V D
2000
GGAAGCGGCACTCAACGGTATGATACCATAGATATGAAAATCTGGGTCCAAATCAGTGG
G S G T Q R Y D T I D M K T V G P N Q W
GTTCACTGTACAATCCGCAATACAGAATCTTCCGATGCAACAGATATGTATGTTTAT
V H L Y N P Q Y R I P S D A T D M Y V Y
2100
GTGAAACAGCGGATGACACCATTAACCTTCTAGATAGTGAAGCAATCGGAGCGGTTGCC
V E T A D D T I N F Y I D E A I G A V A
2200
GGAACGTGAATCGAAGGACCTGCTCCACAGCCTACACAGCCTCCGGTACTGCTTGGCGAT
G T V I E G P A P Q P T Q P P V L L G D
D
GTAAACGGTATGGAACCATTAACCTCAACTGACTTGCAATGTTAAAGAGAAGCGTGTG
V N G D G T I N S T D L T M L K R S V L
2300
AGGGCAATCACCTTACCGACGATGCAAGGCTAGAGCAGACGTTGCAAGAATGGATCG
R A I T L T D D A K A R A D V D K N G S
3'pCF4
ATAAACAGCACTGATGTTTACTTCTTTCACGCTACCTTTTAAAGTAATCGACAAATTT
I N S T D V L L L S R Y L L R V I D K F
E
2400
CCTGTAGCAGAAAATCCTTCTTCTTTTAAATATGATCGCGCGGCAATATCGGCCG
P V A E N P S S S F K Y E S A V Q Y R P
2500
GCTCCTGATTCTTATTAAACCTTGTCCGACGGCGGGAAGAATTTGTAAGGAAACATAT
A P D S Y L N P C P Q A G R I V K E T Y

Figure 3 For continuation of Figure and for legend see facing page.

Figure 3 Nucleotide sequence of *xynY*

The nucleotide sequence of *xynY* and the deduced primary structure of XYLY are shown. The locations of the first residues of domains A, B, C, D and E are indicated with the corresponding letters. The positions of the two primers used to amplify the region of *xynY* coding for the catalytic domain of the xylanase (pCF2/3) are indicated by overlining. The 5' and 3' nucleotides of truncated forms of *xynY* are indicated by a downward arrow and the plasmids that encode the derivatives of the xylanase gene. The nucleotide sequence has been submitted to the EMBL database under the accession number X83269.

1991) and *T. saccharolyticum* (Lee et al., 1993). Among the residues of XYLY which exhibited identity with other xylanases was Glu-462, which has been shown to be the active-site nucleophile that stabilizes the transition-state oxocarboxonium ion that is generated during hydrolysis of the glycosidic bond (Tull et al., 1991). The 176-residue sequence immediately downstream of the putative catalytic domain, designated domain C (Figure 1), exhibited 28% sequence identity with residues 200–353 of xylanase A from *T. saccharolyticum* (Figure 4), which apparently confer increased thermostability on the *T. saccharolyticum* enzyme (Lee et al., 1993). Domain C was followed by a 23-residue duplicated sequence, designated domain D, which is highly conserved among cellulases and XYLZ from *C. thermocellum*; that are components of the multienzyme complex termed the 'cellulosome'. The duplicated sequence plays a pivotal role in cellulosome assembly by docking the various catalytic subunits of the complex onto a non-catalytic scaffolding protein termed CipA (Béguin and Aubert, 1994). The C-terminal 283 residues of XYLY, designated domain E, exhibited 24% homology with the N-terminal domain of XYLZ from *C. thermocellum* (Figure 4).

The sequence analysis described above suggests that XYLY is composed of five distinct domains. The domains of other plant cell wall hydrolases are invariably delineated by linker sequences comprising a number of protein motifs, the most common being serine- or threonine/proline-rich sequences. Inspection of the interdomain sequences of XYLY failed to reveal any obvious linker sequences, except for the region between domains B and C which contained six consecutive glycine residues. Linker sequences containing this amino acid have been observed in xylanase D from *Cellulomonas fimi* (Millward-Sadler et al., 1994).

Structural/functional analysis of XYLY

To evaluate the roles of the domains of XYLY, which had been identified by comparing the sequence of XYLY with the known

primary sequences of other xylanases, truncated forms of *xynY* were constructed and the biochemical properties of the encoded polypeptides were evaluated. PCR amplification of the region of *xynY* encoding residues 178–542 of XYLY generated a restriction fragment that encoded a functional xylanase when cloned in-phase with the *lacZ'* translational start of pMTL22p (pCF2) or the *malE* gene of pMAL-c2 (pCF3). Truncated forms of *xynY* that contained only part of this region did not encode a functional xylanase (Figure 1). In conjunction with the observation that domain B exhibited sequence identity with Family F xylanases, these data show that domain B is the catalytic domain of XYLY.

To assess whether domain C conferred increased thermostability on the catalytic activity of XYLY, three derivatives of *xynY* were fused in-phase with *malE* (encodes the *E. coli* MBP) such that the encoded hybrid proteins contained the following domains from XYLY: MBP–domain B (pCF3); MBP–domain B–domain C–domain D' (pCF4); MBP–domain B–domain C–domain D–domain E (pCF5). SDS/PAGE of each hybrid protein, purified by amylose affinity chromatography, revealed that MBP–domain B, MBP–domain B–domain C–domain D' and MBP–domain B–domain C–domain D–domain E had M_s of 78 000, 122 000 and 148 000 respectively. The thermostability and thermophilicity of the three enzymes, presented in Figure 5, show that the temperature optimum of XYLY' containing domain D was 75 °C, while forms of the enzyme lacking this domain had a temperature optimum of 60 °C. Similarly, derivatives of the xylanase containing domain D exhibited increased thermostability; XYLY' encoded by pCF4 or pCF5 retained 90% of its original catalytic activity after incubation at 70 °C for 15 min, while XYLY' expressed by pCF3, which lacked domain D, retained only 10% of its original activity under the same conditions (Figure 5). These data suggest that domain D confers increased thermostability on the Family F catalytic domain of XYLY.

To evaluate whether the C-terminal region of XYLY, defined as domain E, exhibits cellulose- or xylan-binding capacity, purified MBP–XYLY' hybrids encoded by pCF4 and pCF5 were

Figure 4 Homology of the domains of XYLY with other xylanases

Comparison of domain C (a) with the thermostable domain of *T. saccharolyticum* XylA (A; Lee et al., 1993), domain D (b) with the N-terminal domain of *C. thermocellum* XYLZ (Z; Grépinet et al., 1988) and domain B (c) with the catalytic domains of xylanase A from *R. flaveliaciens* (R; Zhang and Flint, 1992), xylanase 4 from *C. saccharolyticum* (C; Luthi et al., 1990), xylanase B from *B. fibrisolvens* (B; Lin and Thomson, 1991) and xylanase A from *T. saccharolyticum*

Figure 5 Thermostability and thermophilicity of XYLY

Purified MBP-XYLY' encoded by pCF3 (▲), pCF4 (▼) and pCF5 (■) were assayed for MUCase activity at various temperatures (a). The three enzymes were also assayed at 60 °C for MUCase activity after incubation for 20 min at the temperatures indicated (c) or after incubation for up to 60 min at 70 °C (b).

incubated with crystalline cellulose (Avicel PH105) or insoluble xylan, and the binding of the fusion proteins to the polysaccharide was assessed. The results (not shown) revealed no significant binding to either cellulose or xylan.

Production of XYLY by *C. thermocellum*

To determine whether XYLY is a component of the *C. thermocellum* cellulosome, the MBP-XYLY' fusion protein, purified by amylose affinity chromatography, was cleaved within the linker sequence between MBP and XYLY and the MBP was removed

(A; Lee et al., 1993) respectively. The XYLY sequence is indicated by Y. The positions of the first and last residues, within the full-length sequence of the appropriate enzyme, are shown for each domain. Sequences which are either identical or similar for the sequences compared are indicated by an asterisk.

Figure 6 Western blot analysis of the cellulosome of *C. thermocellum*

C. thermocellum cellulosomes (lanes 1 and 2), purified by the method of Morag et al. (1992), and high molecular mass standards (Sigma) (lane 3) were fractionated by SDS/PAGE using a 7.5% (w/v) polyacrylamide gel and were either subjected to Western blot analysis using antiserum raised against the catalytic domain of XYLY (lane 1) or stained with Coomassie Blue (lanes 2 and 3). The M_r s ($\times 10^{-3}$) of the immunoreactive polypeptides are shown.

by rebinding to amylose. The purified catalytic domain of XYLY was used to raise antiserum against the xylanase. When anti-XYLY (catalytic domain) antiserum was employed in Western blot analysis of *C. thermocellum* cellulosome proteins, two polypeptides of M_r s 126 000 and 77 000 respectively, cross-reacted with the antiserum (Figure 6). The larger polypeptide was similar in size to the predicted size of full-length XYLY. These data indicate that *xynY* is expressed by *C. thermocellum* and that XYLY is a component of the cellulosome. The smaller polypeptide that cross-reacted with the probe could either be a processed form of XYLY or a second xylanase that cross-reacted with antibodies raised against the catalytic domain of XYLY.

DISCUSSION

It is well established that the true cellulase activity of *C. thermocellum* is vested in a large multienzyme complex termed the cellulosome. Although numerous endoglucanases and a cellobiohydrolase have been shown to be components of the complex, only one xylanase (XYLZ) has previously been identified within the cellulosome, although numerous polypeptides in the complex exhibit xylanase activity (Morag et al., 1990). Data presented in this report indicate that at least one other xylanase (XYLY) is a component of the cellulosome, based on Western blot analysis. Although it could be argued that the anti-XYLY antiserum was cross-reacting with a Family F xylanase that was not XYLY, the location of the 23-residue duplicated sequence, which plays an integral role in the binding of catalytic subunits to the cellulosome, supports the view that XYLY is a component of the cellulosome. It is apparent, therefore, that the cellulosome plays an important role in the hydrolysis, by *C. thermocellum*, of not only cellulose but of other components of the plant cell wall.

In a previous study Lee et al. (1993) showed that deletion of the N-terminal non-catalytic region of XYLA from *T. saccharolyticum* diminishes the thermostability of the catalytic domain. Data presented in this study revealed that XYLY contained a homologue of this domain that also appears to confer thermostability. It is apparent that these thermostabilizing domains can function when located N- or C-terminal of the xylanase catalytic domain. These data also infer that domains

which confer increased thermostability may be a common phenomenon among Family F xylanases from thermophilic microorganisms. It is interesting to note that such domains are only located in xylanases and have not been observed in the numerous thermophilic endoglucanases characterized to date. It remains to be elucidated whether Family G xylanases also contain non-catalytic domains that confer thermostability.

It is apparent that XYLY from *C. thermocellum*, in common with xylanases from several thermophilic bacteria including *Caldocellum saccharolyticum* (Luthi et al., 1990), *T. saccharolyticum* (Lee et al., 1993) and *C. thermocellum* XYLZ (Grépinet et al., 1988) and XYLX (accession number M67438), contains a Family F catalytic domain. This observation could reflect the close evolutionary relationship between thermophilic bacteria. Thus the evolution of xylanase activity may have arisen through lateral transfer of a single ancestral gene between thermophilic bacteria. This scenario appears unlikely in view of the fact that endoglucanases from *C. thermocellum* are derived from several distinct glycosyl hydrolase families, suggesting that the cellulase system of this bacterium arose through extensive lateral gene transfer. An alternative possibility could relate to the inherent thermostability of Family F and G xylanases. It is possible that the folding of the polypeptide chain in Family F enzymes is such that it has been relatively easy for the enzyme to evolve into a thermophilic enzyme, while Family G xylanases are, in general, inherently more thermolabile. Future studies will focus on the role of the N- and C-terminal non-catalytic domains of XYLY, the structural features of the enzyme that confer thermostability on the catalytic domain and the mechanism by which domain C enhances this thermostability.

We thank the Biotechnology and Biological Sciences Research Council for supporting this work (LE13/138) and Junta Nacional de Investigação e Tecnológica (JNICT) for supporting C. M. G. A. F.

REFERENCES

- Beguin, P. and Aubert, J.-P. (1994) FEMS Microbiol. Rev. **13**, 25–58
- Black, G. W., Hazlewood, G. P., Xue, G.-P., Orpin, C. G. and Gilbert, H. J. (1994) Biochem. J. **299**, 381–387
- Chambers, S. P., Prior, S. E., Barstow, D. A. and Minton, N. P. (1988) Gene **68**, 139–149
- Gerngross, U. T., Romaniec, M. P. M., Huskisson, N. S. and Demain, A. L. (1993) Mol. Microbiol. **8**, 325–334
- Gilbert, H. J. and Hazlewood, G. P. (1993) J. Gen. Microbiol. **139**, 187–194
- Gilkes, N. R., Henricsson, B., Kilburn, D. G., Miller, Jr., R. C. and Warren, R. A. J. (1991) Microbiol. Rev. **55**, 303–315
- Grépinet, O., Chebrou, M.-C. and Béguin, P. (1988) J. Bacteriol. **170**, 4582–4588
- Hall, J., Ali, S., Surani, M. A., Hazlewood, G. P., Clark, A. J., Simons, J. P., Hirst, B. H. and Gilbert, H. J. (1993) Bio/Technology **11**, 376–379
- Hall, J., Hirst, B. H., Hazlewood, G. P. and Gilbert, H. J. (1992) Biochim. Biophys. Acta **1130**, 259–266
- Hall, J., Hazlewood, G. P., Barker, P. J. and Gilbert, H. J. (1988) Gene **69**, 29–38
- Hazlewood, G. P., Romaniec, M. P. M., Davidson, K., Grépinet, O., Béguin, P., Millet, J., Raynaud, O. and Aubert, J.-P. (1988) FEMS Microbiol. Lett. **51**, 231–236
- Hazlewood, G. P., Davidson, K., Laurie, J. I., Romaniec, M. P. M. and Gilbert, H. J. (1990) J. Gen. Microbiol. **136**, 2089–2097
- Kellett, L. E., Poole, D. M., Ferreira, L. M. A., Durrant, A. J., Hazlewood, G. P. and Gilbert, H. J. (1990) Biochem. J. **272**, 369–376
- Laemmli, U. K. (1970) Nature (London) **227**, 680–685
- Lamed, R. and Bayer, E. A. (1988) Adv. Appl. Microbiol. **33**, 1–46
- Lee, Y.-E., Lowe, S. E., Henricsson, B. and Zeikus, J. G. (1993) J. Bacteriol. **175**, 5890–5898
- Lin, L.-L. and Thomson, J. A. (1991) Mol. Gen. Genet. **228**, 55–61
- Luthi, E., Love, D. R., McAnulty, J., Wallace, C., Caughey, P. A., Saul, D. and Bergquist, P. L. (1990) Appl. Environ. Microbiol. **56**, 1017–1024
- Malburg, L. M. and Forsberg, C. W. (1993) Can. J. Microbiol. **39**, 882–891
- Millward-Sadler, S. J., Poole, D. M., Henricsson, B., Hazlewood, G. P., Clarke, J. H. and Gilbert, H. J. (1994) Mol. Microbiol. **11**, 375–382
- Morag, E., Bayer, E. A. and Lamed, R. (1990) J. Bacteriol. **172**, 6098–6105
- Morag, E., Bayer, E. A. and Lamed, R. (1992) Enzyme Microb. Technol. **14**, 289–292

- Murphy, G. and Kavanagh, T. (1989) *Nucleic Acids Res.* **16**, 5198
- Poole, D. M., Morag, E., Lamed, R., Bayer, E. A., Hazlewood, G. P. and Gilbert, H. J. (1992) *FEMS Microbiol. Lett.* **106**, 77–84
- Romaniec, M. P. M., Clarke, N. G. and Hazlewood, G. P. (1987) *J. Gen. Microbiol.* **133**, 1297–1307
- Sakka, K., Kojima, Y., Kondo, T., Karita, S., Ohmiya, K. and Shimada, K. (1993) *Biosci. Biotech. Biochem.* **57**, 273–277
- Salamitou, S., Tokatlidis, K., Béguin, P. and Aubert, J.-P. (1992) *FEBS Lett.* **304**, 89–92
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
- Sedmak, J. J. and Grossberg, J. R. (1977) *Anal. Biochem.* **79**, 544–552
- Staden, R. (1980) *Nucleic Acids Res.* **16**, 3673–3694
- Thomson, J. A. (1993) *FEMS Microbiol. Rev.* **104**, 65–82
- Tokatlidis, K., Salamiou, S., Béguin, P., Dhurjati, P. and Aubert, J.-P. (1991) *FEBS Lett.* **291**, 185–188
- Tull, D., Withers, S. G., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J. and Aebersold, R. (1991) *J. Biol. Chem.* **266**, 15621–15625
- Zhang, J.-X. and Flint, H. J. (1992) *Mol. Microbiol.* **6**, 1013–1023

Received 5 August 1994/18 November 1994; accepted 25 November 1994