

# Proteomic and Transcriptomic Analysis of Extracellular Proteins and mRNA Levels in *Thermobifida fusca* Grown on Cellobiose and Glucose<sup>∇</sup>

Shaolin Chen† and David B. Wilson\*

Department of Molecular Biology and Genetics, 458 Biotechnology Building, Cornell University, Ithaca, New York 14853

Received 16 April 2007/Accepted 19 June 2007

***Thermobifida fusca* secretes proteins that carry out plant cell wall degradation. Using two-dimensional electrophoresis, the extracellular proteome of *T. fusca* grown on cellobiose was compared to that of cells grown on glucose. Extracellular proteins, the expression of which is induced by cellobiose, mainly are cellulases and cellulose-binding proteins. Other major extracellular proteins induced by cellobiose include a xylanase (Xyl10A) and two unknown proteins, the C-terminal regions of which are homologous to a lytic transglycosylase goose egg white lysozyme domain and an NLPC\_P60 domain (which defines a family of cell wall peptidases), respectively. Transcriptional analysis of genes encoding cellobiose-induced proteins suggests that their expression is controlled at the transcriptional level and that their expression also is induced by cellulose. Some other major extracellular proteins produced by *T. fusca* grown on both cellobiose and glucose include Lam81A and three unknown proteins that are homologous to aminopeptidases and xylanases or that contain a putative NLPC\_P60 domain.**

Biodegradation of plant cell walls by bacteria and fungi is accomplished by extracellular proteins, including cellulases and hemicellulases. *Thermobifida fusca* is a filamentous soil bacterium that grows at 55°C in defined medium on cellulose as a carbon source. It appears to be a major degrader of plant cell walls in heated organic material such as compost piles and rotting hay (2). Cellulases produced by *T. fusca* include Cel5A, Cel6A, Cel6B, Cel9A, Cel9B, and Cel48A (23). *T. fusca* also excretes hemicellulases, including two xylanases (*xyl11A* and *xyl10B*), a xyloglucanase (*xg74A*), and a mannanase (*manB*) (3, 23).

In addition to cellulases and hemicellulases, most cellulolytic microorganisms also produce extracellular proteases that partially degrade their cellulases into multiple active isozymes (11). A common modification is the removal of a cellulose-binding domain to produce a catalytic domain species (5, 6, 20). However, little is known about the role of proteolysis in cellulose metabolism. An active thermostable protease excreted from *T. fusca* has been identified that can cleave *T. fusca* cellulases into many isozymes (4). *T. fusca* cannot utilize casein as a nitrogen source, so the protease does not function to scavenge amino acids from proteins in the environment. However, the extracellular protease is not required for growth on cellulose, because a *T. fusca* mutant (ER1) that does not produce the protease was isolated, and this strain grows normally on all carbon sources tested, including cellulose (22). The extracellular protease has been cloned, characterized, and shown to resemble a lytic protease and to have high levels of activity on most proteins (23).

Cellulase biosynthesis in *T. fusca* likely is regulated through the control of transcription, as indicated by a correlation be-

tween the levels of the *cel5A* transcript and the level of the protein product of this gene produced on different carbon sources (15). A 14-bp inverted repeat in the upstream regions of the cloned *cel* genes has been identified as a *cis*-acting regulatory element (13), and a transcription regulator, CelR, that specifically binds to the 14-bp inverted repeat has been isolated (17). Cellobiose causes dissociation of the CelR-DNA complex, as shown by *in vitro* DNA-binding experiments, suggesting that cellobiose acts as an effector for the expression of cellulase genes (17).

In this paper, we describe identification of other extracellular proteins produced by *T. fusca* grown on cellobiose and glucose using a combination of two-dimensional electrophoresis (2DE), mass spectrometry (MS), and genome database searches. DNA microarray and real-time PCR were applied to examine gene expression at the transcription level.

## MATERIALS AND METHODS

**Culture conditions.** *T. fusca* ER1 was grown in Luria-Bertani medium containing 0.2% glucose. This culture was centrifuged at 3,000 × *g* in a desktop centrifuge for 10 min, and the cell pellet was resuspended in Hagerdahl minimal medium (7) supplemented with either 0.5% glucose, 0.5% cellobiose, or 1% bacterial microcrystalline cellulose (BMCC) (Monsanto Cellulon; Monsanto Company, San Diego, CA). The substrate concentrations used in this study were not growth limiting. Cells were cultured at 50°C and 200 rpm and were harvested at the mid-log growth phase by centrifugation (14).

**Protein preparation from *T. fusca* culture supernatants.** Glycerol and phenylmethylsulfonyl fluoride were added to *T. fusca* culture supernatants to 10% and 3 mM, respectively. The supernatants were filtered through a 0.45-μm-pore-size filter and then were concentrated using a membrane filter with a molecular mass cutoff point of 10 kDa. Concentrated supernatants were stored at –70°C. Protein concentrations of concentrated supernatants were determined by using a RC DC protein assay (Bio-Rad), with bovine serum albumin as a standard.

**2DE.** For 2DE, the protein buffer was changed to a rehydration solution, containing 8 M urea, 0.5% 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate, and 0.2% dithiothreitol (DTT), by repeated concentration and dilution with this buffer in a 5K Ultrafree 0.5 centrifugal filter (Millipore). Immobilized pH gradient buffer (Amersham Biosciences) and bromophenol blue (BPB) were added to final concentrations of 0.5% and 0.002%, respectively, to the resulting solution. Protein samples were run in the first dimension using 13-cm immobilized pH gradient strips (pH 3 to 10, nonlinear; Amersham Bio-

\* Corresponding author. Mailing address: Department of Molecular Biology and Genetics, 458 Biotechnology Building, Cornell University, Ithaca, NY 14853. Phone: (607) 255-5706. Fax: (607) 255-2436. E-mail: dbw3@cornell.edu.

† Present address: Department of Plant Biology, Carnegie Institution of Washington, Stanford, CA 94305.

<sup>∇</sup> Published ahead of print on 29 June 2007.

sciences) and a Multiphor II system (Amersham Biosciences). The strips were allowed to rehydrate overnight at room temperature with 250  $\mu$ l of protein samples added to each lane of a rehydration tray. The isoelectric focusing was carried out at 20°C in a gradient mode, increasing the voltage from 300 to 3,500 V over 1.5 h, followed by holding the voltage at 3,500 V for 4.5 h. After completion of the isoelectric focusing run, the strips were equilibrated in 50 mM Tris-HCl buffer, pH 6.8, containing 6 M urea, 30% glycerol, 2% sodium dodecyl sulfate (SDS), 1% DTT, and a trace amount of BPB for 10 min, followed by an additional 10 min of equilibration in the same solution, except that DTT was replaced with 2.5% iodoacetamide. The strips then were transferred and apposed to SDS-10% polyacrylamide vertical gels (14 by 16 cm) that were cast in-house using a Hoefer SE 600 gel cast apparatus. The vertical gels were run using a Hoefer vertical gel running system (Amersham Biosciences) at 20 mA for 30 min, followed by 50 mA for about 3 to 4 h, until the BPB front marker reached the bottom of the gel. The protein spots were visualized by silver staining using Invitrogen's Silver-Quest silver-staining kit according to the manufacturer's instructions. Stained gels were scanned using a Magic-scanning system from Amersham Pharmacia Biotech, and raw scans were processed by Image Master software.

**LC-MS/MS identification of proteins from 2DE.** Selected protein spots were cut into small pieces (1 to 2 mm) using a scalpel. In-gel digestion was performed by following a protocol from Zhang et al. (24). Digested peptides were extracted from gel slices and then analyzed using liquid chromatography-MS/MS (LC-MS/MS) on an LTQ linear ion trap mass spectrometer (Thermo Electron). The peptides were separated on a BioBasic C-18 PicoFrit column (75  $\mu$ m [internal hamester] by 10 cm; New Objective) at a flow rate of 200 nl/min. Bioworks 3.2 software (Thermo Electron) was used to search the *T. fusca* GenBank protein database. Database search parameters used were the following: peptide tolerance, 2.0; fragment ion tolerance, 1.0; and missed cleavage sites, 2.0. A peptide probability of  $10^{-4}$  and an the crossing correlation score (Xcorr) of 2.0 (+2), 2.5 (+3) were used as cutoff filters. All of the hits with a probability of  $10^{-4}$  to  $10^{-6}$  were manually evaluated, and the low-confidence hits were removed from the final list.

**RNA purification and labeling.** Cells grown to mid-log phase were harvested by centrifugation at  $12,000 \times g$  for 5 min at 4°C. The cell pellet was resuspended in RNAlater bacteria reagent (QIAGEN, Valencia, CA) as described by the manufacturer. After incubation at room temperature for 5 min, the cell suspension was centrifuged at  $12,000 \times g$  for 10 min. The cell pellets were resuspended in a one-quarter culture volume of modified Kirby mix (1% [wt/vol] *N*-lauroylsarcosine, 6% [wt/vol] *p*-aminosalicylic acid sodium salt, 6% [vol/vol] phenol, 0.1 M Tris-HCl, pH 8.0) (8). The cell suspension was immediately disrupted using a French press at 3,500 lb/in<sup>2</sup> directly into a polypropylene tube, which contained a one-eighth culture volume of phenol-chloroform-isoamyl alcohol (PCI) (25:24:1). The tube was vortexed for 2 to 3 min and then was centrifuged at  $6,000 \times g$  for 10 min at 4°C. The upper phase was transferred to a fresh polypropylene tube, which contained a one-eighth culture volume of PCI. After the phenol-chloroform extraction step was repeated two times, the aqueous phase was transferred to a fresh polypropylene tube that contained a one-eighth culture volume of chloroform. The tube was vortexed for about 30 s and centrifuged as described above. The clear supernatant was transferred into a fresh tube. Total nucleic acid was precipitated, washed, and treated with RQ1 RNase-free DNase as described by Kieser et al. (8). The DNA-free total nucleic acid sample was further purified using RNeasy MIDI columns (QIAGEN, Valencia, CA) as described by the manufacturer. RNA labeling was carried out according to a protocol from The Institute For Genomic Research ([http://pga.tigr.org/sop/M004\\_1a.pdf](http://pga.tigr.org/sop/M004_1a.pdf)). Briefly, RNA was labeled with aminoallyl-labeled nucleotides via first-strand cDNA synthesis, followed by a coupling of the aminoallyl groups to either cyanine 3 (Cy3) or Cy5 fluorescent dyes.

**Preparation of DNA array.** One hundred twenty-seven open reading frames (ORFs) present in the published genome of *T. fusca* were selected for the microarray study, including those that encode cellulases, hemicellulases, cellulose-binding proteins, and secreted proteases. Given the high G+C content and low sequence complexity of the *T. fusca* genes, 25mer oligonucleotides were used for the oligonucleotide-based microarray. The oligonucleotides, including a C-6 amino linker, were arrayed on glass slides (Quantifoil/Schott epoxy slides) using an OminGrid robot (GeneMachines, San Carlos, CA) with Chipmaker pins (TeleChem International, Inc., Sunnyvale, CA) according to the manufacturer's instructions. The oligonucleotide probe for each gene was spotted three times on each slide. Microarray slides were washed and blocked with QMT blocking solution (Quantifoil) at 50°C according to the manufacturer's instructions.

**Microarray hybridization and data analysis.** Hybridization was carried out using SlidHyb buffer 1 (Ambion, Inc., Austin, TX) and LifterSlips cover glasses (Erie Scientific Company, Portsmouth, NH) according to the manufacturers'

instructions. Briefly, Cy3 and Cy5 cDNA was resuspended in 4  $\mu$ l of 10 mM EDTA and then was denatured at 95°C for 10 min. SlidHyb buffer 1 was added to the denatured cDNA to a final volume of 28  $\mu$ l. The resulting mixture then was slowly pipetted onto a microarray slide, and a LifterSlip cover glass (22 by 25 mm) was applied. The slide was placed in a hybridization chamber (Corning, Inc., Corning, NY) and hybridized for 14 to 20 h at 42°C in a water bath. The slide was washed at room temperature in  $2\times$  SSC plus 0.2% SDS for 10 min ( $1\times$  SSC is 0.15 M NaCl plus 0.015 M sodium citrate),  $2\times$  SSC for 10 min, and  $0.2\times$  SSC for 10 min. After it was dried by centrifugation, the hybridized slide was scanned with a Gene Pix4000B scanner (Molecular Devices Corporation, Union City, CA) using the Gene Pix Pro 4.0 software package (Molecular Devices). The internal normalization by rank-invariant method was used to select genes that were nondifferentially expressed in each slide (21). This procedure performs slide-dependent nonlinear normalization of the log ratios of the Cy3 and Cy5 channels and determines the confidence intervals of gene expression ratios (21).

**Real-time PCR.** For quantitative reverse transcription-PCR cDNA synthesis, 10  $\mu$ g of total RNA and 6  $\mu$ g of random hexamer primers (Life Technologies) were heat denatured for 10 min at 70°C and chilled on ice before adding the remaining components for the reverse transcription reaction: 0.5 mM (each) dATP, dCTP, dGTP, and dTTP, 10 mM dithiothreitol, 10 U of SuperScript II reverse transcriptase (Life Technologies), and reaction buffer to a final concentration of  $1\times$ . Reactions proceeded for 10 min at 25°C and then for 3 to 12 h at 42°C, were heat inactivated at 70°C for 15 min, and then were chilled on ice. After adding 1 U of RNase H, the reaction mixture was kept at 37°C for 1 h. The resulting first-strand cDNA was stored at -20°C until use for real-time PCR.

Real-time PCR was performed on an Applied Biosystems 7900HT sequence detection system according to the manufacturer's instructions. PCR components were optimized by using the FailSafe real-time PCR PreMix selection kit (Epicenter Technologies, Madison, WI). FailSafe PCR enzyme mix from Epicenter Technologies was used to generate and detect double-stranded DNA according to the manufacturer's protocol. Cycling conditions were as follows: an initial denaturation at 95°C for 1.5 min, and then 50 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 30 s. The real-time PCR results were analyzed with SDS 2.0 software (Applied Biosystems).

## RESULTS AND DISCUSSION

**Proteomic analysis.** Extracellular proteins of bacterial cultures usually are recovered by precipitation of filtered culture supernatants with trichloroacetic acid. Trichloroacetic acid precipitation procedures give a poor yield for *T. fusca* supernatants, and solubilized precipitates were difficult to resolve by SDS-polyacrylamide gel electrophoresis (data not shown). We therefore used filter-concentrated protein samples, which resulted in high yields and well-resolved 2DE patterns for samples grown on glucose and cellobiose (Fig. 1). Amounts of extracellular proteins separated on 2DE gels were equivalent to those of 1-ml *T. fusca* cultures, i.e., 38 and 59  $\mu$ g of total extracellular proteins for cultures grown on glucose and cellobiose, respectively.

Glucose is a repressive, noninducing substrate for cellulase expression, so the expression profile for glucose-minimal-medium-grown cells was used as the reference to which the expression profile of cells grown on cellobiose were compared. The results of 2DE analysis suggest that cellobiose induced the appearance of over 20 major spots, the intensities of which were undetectable in the glucose culture (Fig. 1). These spots were further analyzed, and 15 proteins were identified by tandem MS of tryptic digests and by searching the *T. fusca* protein database (Table 1). The results suggest that cellulases are among the major proteins induced by cellobiose. Consistent with a previous Western blot analysis of cellulase levels (18), some cellulases, such as Cel6B and Cel48A, are more abundant than others, such as Cel5A and Cel9B, in cultures grown on cellobiose.

*T. fusca* cellulases are composed of multiple reactive domains: the cellulose-binding domain is responsible for the high

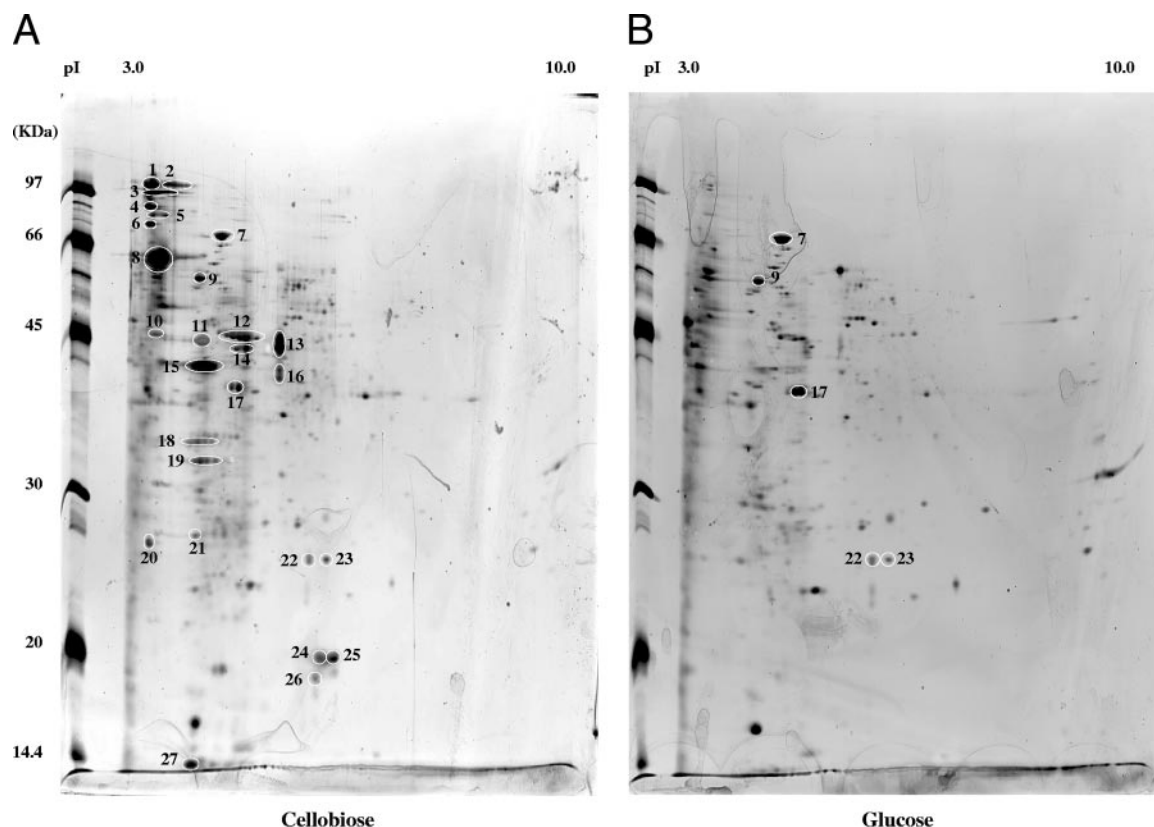


FIG. 1. Two-dimensional maps of the extracellular proteomes of *T. fusca* grown on cellobiose (A) and glucose (B). Proteins were separated in a pH gradient from 3 to 10 and in the SDS-10% polyacrylamide gel electrophoresis that were visualized by silver stain. Theoretical  $M_r$  and pI scales are indicated. Spot numbering refers to those given in Table 1, which reports protein identification by LC-MS/MS.

level of cellulase binding to insoluble cellulose, and the catalytic domain is responsible for hydrolyzing the substrate. The results of this study suggest that putative carbohydrate-binding proteins E7 and E8, which do not contain a catalytic domain, are two other major proteins induced by cellobiose. E7 has 222 amino acid residues and starts with a secretion signal, followed by a family 33 chitin-binding domain (positions 22 to 208). E8 is 438 amino acids long and contains both a family 33 chitin-binding domain (positions 17 to 222) and a family II cellulose-binding domain (positions 335 to 435). Purified E7 was found to enhance the initial rates of hydrolysis of cellulases toward insoluble cellulose but not those of soluble cellulose (data not shown). It is likely that the cellulose-binding proteins function in the hydrolysis of cellulose, particularly insoluble cellulose, through interactions with the substrate.

In addition to cellulolytic proteins, major extracellular proteins induced by cellobiose include a xylanase, Xyl10A, a putative proteolytic enzyme, E9, and a putative lysozyme, E10 (Fig. 1 and Table 1). E9 has 332 amino acids. Its C-terminal sequence, 229 to 330, is homologous to the NLPC\_p60 domain. NLPC\_p60 is a domain conserved through a variety of bacterial lineages. Although the function of this domain is not completely understood, some members of this family have been shown to act as murein hydrolases that cleave peptide linkages within peptidoglycan (1). E10 is a 216-amino-acid protein and contains a putative lytic transglycosylase and goose egg white lysozyme domain between amino acids 141 and 209. Cellobiose

probably induces high-level expression of Xyl10A, E9, and E10, because cellulose microfibrils in plant primary cell walls are embedded in a matrix of polysaccharides such as hemicellulose and pectin. Xyl10A could function to help remove plant cell wall materials, such as xylan, to allow *T. fusca* cellulases better access to embedded cellulose microfibrils. E9 and E10 are not similar to any known plant cell wall polymer hydrolases and could function to attack competing bacteria or possibly to attack minor components in plant cell walls.

Degradation of cellulases into cellulose-binding domain and catalytic domain species has been observed with most cellulolytic microorganisms (5, 6, 20). *T. fusca* YX has been shown to excrete a serine protease, TfpA, that is able to degrade cellulases (9–11). The strain used in this study, ER1, does not produce active TfpA (22), but degradation was still observed with some cellulases (Fig. 1 and Table 1), indicating that ER1 is capable of producing other extracellular proteases. For instance, Cel6B (molecular mass, 59.6 kDa) contains 596 amino acids, and its cellulose-binding domain and catalytic domain are in residues 41 to 139 and 190 to 551, respectively. Spot 11 (molecular mass, ~45 kDa) was identified as a fragment of Cel6B, and all 24 peptides found in spot 11 are in residues 186 to 584, corresponding to the catalytic domain of Cel6B (Table 2). Similarly, Cel48A consists of 984 amino acids (molecular mass, 104 kDa) with cellulose-binding and catalytic domains in residues 38 to 138 and 344 to 979, respectively. All 18 Cel48A peptides identified in spot 4 (molecular mass, ~72 kDa) are in

TABLE 1. Identified extracellular proteins of *T. fusca* grown on cellobiose

Spot no.	Protein name (designation)	Accession no.	Theoretical $M_r$ /pI	Exp $M_r$ /pI	Random probability <sup>a</sup>	No. of peptides identified/% coverage	Locus <sup>b</sup>
1	$\beta$ -1,4-Exocellulase (Cel48A)	YP_290015	104/4.2	110/3.3	$5.03^{-12}$	24/13	Tfu_1959
3				110/3.4	$4.97^{-8}$	7/8.0	
4				90/3.3	$1.04^{-10}$	18/12	
6				80/3.4	$3.22^{-8}$	7/6.5	
21	$\beta$ -1,4-Endocellulase (Cel9B)	YP_289685	101/4.0	27/4.2	$1.33^{-9}$	2/2.7	Tfu_1627
3				110/3.4	$3.27^{-13}$	8/9.8	
4				90/3.3	$1.85^{-9}$	2/1.7	
5				82/3.4	$7.23^{-11}$	4/6.6	
10	$\beta$ -1,4-Endo/exocellulase (Cel9A)	YP_290232	90/4.4	45/3.4	$1.45^{-11}$	6/7.4	Tfu_2176
2				110/3.7	$4.8^{-12}$	11/16	
18	$\beta$ -1,4-Exocellulase (Cel6B)	YP_288681	60/4.1	34/4.2	$7.16^{-7}$	2/2.3	Tfu_0620
8				65/3.4	$1.1^{-16}$	30/25	
11	$\beta$ -1,4-Endoxylanase (Xyl10A)	YP_290979	49/4.9	44/4.1	$1.02^{-10}$	24/30	Tfu_2923
12				45/4.7	$2.74^{-14}$	23/51	
15				42/4.1	$1.16^{-11}$	3/7.7	
14	$\beta$ -1,4-Endoglucanase (Cel5A)	YP_288962	50/5.2	45/4.6	$4.06^{-6}$	4/8.4	Tfu_0901
15	Putative cellulose-binding protein (E8)	YP_289723	47/4.4	42/4.1	$4.98^{-10}$	11/19	Tfu_1665
13	$\beta$ -1,4-Endoglucanase (Cel6A)	YP_289135	46/5.9	45/5.3	$1.67^{-12}$	9/20	Tfu_1074
16	Putative secreted proteolytic enzyme containing NLPC_P60 domain (E9)	YP_289090	36/4.5	40/5.3	$1.67^{-9}$	1/2.7	Tfu_1029
19				35/4.2	$2.35^{-8}$	2/7.0	
24				Putative secreted cellulose-binding protein (E7)	YP_289329	25/7.5	
25	Putative secreted protein containing lytic transglycosylase and goose egg white lysozyme domain (E10)	YP_288519	23/4.2	20/6.1	$1.00^{-30}$	13/65	Tfu_0458
26				17/5.8	$7.90^{-12}$	2/12	
27				14/3.8	$1.91^{-5}$	1/3.6	
20				26/3.3	$6.72^{-12}$	2/11	
7 <sup>c</sup>	$\beta$ -1,3-Glucanase (Lam81A)	YP_290186	82/5.0	70/4.4	$1.0^{-30}$	52/33	Tfu_2130
9 <sup>c</sup>	Predicted secreted aminopeptidase (E11)	YP_290459	49/4.5	57/4.0	$1.46^{-9}$	28/27	Tfu_2403
17 <sup>c</sup>	Putative secreted proteolytic enzyme containing NLPC_P60 domain (E12)	YP_290318	42/4.9	40/4.7	$6.31^{-11}$	8/22	Tfu_2262
22 <sup>c</sup>	Putative secreted xylanase or chitin deacetylase (E13)	YP_290845	27/6.7	27/5.8	$4.25^{-6}$	4/9.8	Tfu_2789
23 <sup>c</sup>				27/6.1	$1.13^{-6}$	6/18	

<sup>a</sup> The probability, assigned by BioWorks 3.2 software, of finding the identified peptides randomly.

<sup>b</sup> The loci listed are ORF numbers in the *T. fusca* YX genome sequence NC\_007333, from NCBI (<http://www.ncbi.nlm.nih.gov/Genomes/>).

<sup>c</sup> Protein spots with comparable intensities on both cellobiose- and glucose-derived 2DE gels.

the sequence from amino acids 226 to 893, matching the catalytic domain of Cel48A.

Putative extracellular proteases identified in this study include E9, E11, and E12. While E9 is present only in cultures

grown on cellobiose, E11 and E12 are present in cultures grown on both cellobiose and glucose (Fig. 1 and Table 1). E11 is 512 amino acids long, and its C-terminal sequence, residues 244 to 487, is homologous to a family 28 peptidase domain.

TABLE 2. Peptides identified in spots 4 and 11 and their positions in Cel6B and Cel48A<sup>a</sup>

Spot no.	Protein	Peptides
4	Cel48A	226TVSAPIAIR234, 239AAVIASPTVR249, 250VPQGGTADFEVR261, 262LSNQPSGNTVTVAR276, 277TSGSSDLTVSSGSQIQFTSSNWNQPK303, 355IKDPASGYFR364, 561YTNAPDADAR570, 571AVQVMFWAHEWAK583, 584EQGKENEIAGLMDK597, 601MGDYLR606, 607YAMFDKYFKK616, 616KIGNCVGATSCPGGQK632, 767VAQLYYVTGDAR778, 779AEAILDK785, 879DSIGIATPEQPSWDR893
11	Cel6B	186AAAEPGGSANESTAVWLDR206, 207IGAIEGNDSPPTGSMGLR224, 225DHLEEA VR232, 233QSGGDPLTIQVVIYNLPGR251, 285FADYENLR292, 323QNGGYVNGVGYALRK337, 394SATVEPYLDVNGTVNGQLIR413, 444SDIGMLIDTSR454, 455NGWGGPNRPTGPSSTDLNTYVDESR480, 485IHPGNWCNQAGAGLGERPTVNPAPGVDA512, 575ELLANAYPPL584

<sup>a</sup> Redundant peptides that were identified by MS analysis are not shown. The numbers before and after each peptide represent its position in the sequence.

TABLE 3. Fold changes of *T. fusca* gene expression in BMCC and cellobiose compared to expression in glucose

Locus <sup>a</sup>	Gene product (protein designation)	Transcript ratio <sup>b</sup> in:	
		Cellobiose	BMCC
Tfu_1627	$\beta$ -1,4-Endoglucanase (Cel9B)	9.4–15	18–25
Tfu_1074	$\beta$ -1,4-Endoglucanase (Cel6A)	4.1–8.3	5.0–7.4
Tfu_0620	Nonreducing end-directed $\beta$ -1,4-exocellulase (Cel6B)	12–28	28–40
Tfu_2176	$\beta$ -1,4-Endoglucanase (Cel9A)	3.2–5.4	4.3–6.2
Tfu_0901	Endo-1,4- $\beta$ -glucanase (Cel5A)	1.3–2.7	2.1–3.7
Tfu_1959	Reducing end-directed $\beta$ -1,4-exocellulase (Cel48A)	2.5–4.9	2.1–4.3
Tfu_1268	Predicted cellulose-binding protein (E7)	3.9–11	4.8–6.6
Tfu_1665	Predicted cellulose-binding protein (E8)	6.3–15	8.8–13
Tfu_0934	Predicted ABC-type sugar transport system, periplasmic component for cellobiose/ celotriose	24–40	15–22
Tfu_0937	$\beta$ -Glucosidase (BglC)	7.1–16	5.1–8.7
Tfu_0938	CelR	3–13	1.3–1.9
Tfu_2923	$\beta$ -1,4-Endoxylanase (Xyl10A)	4.7–5.6	4.3–6.0
Tfu_2788	Putative $\beta$ -1,4-endoxylanase	0.64–1.4	0.25–0.37
Tfu_1213	$\beta$ -1,4-Xylosidase (TFX11A/Xyl11A)	0.35–0.58	0.27–0.38
Tfu_2791	$\beta$ -1,4-Endoxylanase (Xyl10B)	0.29–1.5	0.74–1.0
Tfu_1612	Putative secreted xyloglucanase	0.69–1.5	0.82–1.5
Tfu_2990	Putative secreted xylanase	0.82–1.7	0.43–0.65
Tfu_0900	Putative secreted $\beta$ -mannanase	0.75–1.6	1.1–1.5

<sup>a</sup> The loci listed are ORF numbers in the *T. fusca* YX draft assembly genome sequence NC\_007333, from NCBI (<http://www.ncbi.nlm.nih.gov/genomes/>).

<sup>b</sup> Analysis of the microarray data with 95% confidence intervals using a Markov chain Monte Carlo (MCMC) method described in Materials and Methods showed that the expected transcript level changes in BMCC and cellobiose were between upper and lower bounds.

E12 has 388 amino acids, and like E9, the C-terminal region of E12 (amino acids 278 to 374) contains a putative NLPC\_p60 domain.

Other extracellular proteins produced by cells grown on both cellobiose and glucose include Lam81A and E13. Recently we purified and characterized Lam81A, and it is specific for  $\beta$ -1,3-linked glucose polysaccharides and is endohydrolytic (16). The natural substrate of Lam81A is likely to be the plant cell wall polysaccharide, callose, a  $\beta$ -1,3-linked glucose polymer. The callose content of a typical living plant is low, but there is evidence for callose accumulation after exposure of plants to a range of abiotic and biotic stresses, including wounding, desiccation, metal toxicity, and microbial attack (19). It also is possible that Lam81 functions to hydrolyze the cell walls of competing fungi, like some plant  $\beta$ -1,3 glucanases (12). E13 is a 758-amino-acid protein that is homologous to a xylan or chitin deacetylase. It may function in helping to remove hemicellulose to expose cellulose microfibrils for degradation by *T. fusca*.

**Transcriptional analysis.** Our previous analysis showed a close correlation between the levels of the *cel5A* transcript and the level of the protein product of this gene produced on different carbon sources (15). In this study, we examined transcriptional expression of all six *T. fusca* cellulase genes and some other genes in cells grown on cellobiose and cellulose (BMCC) by using a DNA microarray and quantitative real-

TABLE 4. Real-time PCR analysis of the expression of *celR*, *cel6A*, *cel9B*, *cel5A*, and *xyl11A* in BMCC and cellobiose compared to that in glucose

Locus (gene designation)	Transcript ratio in:	
	Cellobiose	BMCC
Tfu_0938 ( <i>celR</i> )	8.8 $\pm$ 2.0	2.2 $\pm$ 0.2
Tfu_1074 ( <i>cel6A</i> )	17 $\pm$ 4	6.9 $\pm$ 0.7
Tfu_1213 ( <i>xyl11A</i> )	0.29 $\pm$ 0.07	0.33 $\pm$ 0.03
Tfu_1627 ( <i>cel9B</i> )	8.4 $\pm$ 1.9	32 $\pm$ 3
Tfu_0901 ( <i>cel5A</i> )	4.6 $\pm$ 1.1	2.8 $\pm$ 0.3

time PCR analysis. Again, the expression profile for glucose-minimal-medium-grown cells was used as the reference to compare the expression profiles for cells grown on cellobiose and cellulose. For DNA microarray analysis, we applied a small-scale microarray as described by Tseng et al. (21).

The results suggest that cellobiose enhances the transcription of genes that relate to many aspects of cellulolysis, including cellulose binding and degradation (*cel5A*, *cel6A*, *cel6B*, *cel9A*, *cel9B*, *cel48A*, the E7 gene, and the E8 gene), uptake of degradation products (Tfu\_0934), cellobiose catabolism (*bglC*), and gene regulation (*celR*) (Tables 3 and 4). All of the results support the idea that cellobiose acts as an inducer of cellulolytic genes (17).

To confirm the role of cellobiose in gene expression, we further examined transcriptional expression of genes in cells grown on BMCC. BMCC was selected because it is synthesized by bacteria, which do not produce hemicellulose. The results suggest that BMCC induces the transcriptional expression of cellulolytic genes but not hemicellulase genes, except for the one encoding Xyl10A (Tables 3 and 4). Sequence analysis further indicates that the upstream regions of all BMCC-induced genes and the proteins induced by cellobiose have at least one CelR-binding site (data not shown).

**Conclusions.** The results of this study support the model that cellobiose acts as an inducer for cellulose biodegradation. Cellobiose induces the expression of genes that relate to many aspects of cellulolysis, including cellulose binding and degradation, uptake of degradation products, and cellobiose catabolism. In addition to genes that are directly involved in cellulolysis, cellobiose also induces the expression of xylanase Xyl10A and other proteins, such as E9 and E10, which may work as enzymes to cleave glycosyl and peptide linkages in plant cell wall materials or to attack competing bacteria, respectively. Some or all of the cellobiose-induced enzymes probably function in removing plant cell wall materials to allow cellulases better access to the cellulose. *T. fusca* also produces extracellular proteins in cultures grown on glucose, including a putative callose hydrolase, Lam81A, a putative proteolytic enzyme, E12, and a xylanase or chitin deacetylase, E13. It will be interesting to see whether these enzymes function in initiating plant cell wall degradation, particularly disassembly of the surface structure of plant cell walls, producing signaling molecules and inducing expression of enzymes for further degradation of plant cell walls, since a product of callose hydrolysis by Lam81, laminaribiose, was shown to be a good inducer of *T. fusca* cellulase synthesis (16).

## ACKNOWLEDGMENTS

We thank Diana Irwin for technical assistance and support.

This work was supported by grant DE-FG02-01ER63150 from the Department of Energy Basic Energy Research Program.

## REFERENCES

1. Anantharaman, V., and L. Aravind. 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol.* **4**:R11.
2. Bachmann, S. L., and A. J. McCarthy. 1991. Purification and cooperative activity of enzymes constituting the xylan-degrading system of *Thermomonospora fusca*. *Appl. Environ. Microbiol.* **57**:2121–2130.
3. Béki, E., I. Nagy, J. Vanderleyden, S. Jager, L. Kiss, L. Fulop, L. Hornok, and J. Kukolya. 2003. Cloning and heterologous expression of a  $\beta$ -D-mannosidase (EC 3.2.1.25)-encoding gene from *Thermobifida fusca* TM51. *Appl. Environ. Microbiol.* **69**:1944–1952.
4. Calza, R. E., D. C. Irwin, and D. B. Wilson. 1985. Purification and characterization of two  $\beta$ -1,4-endoglucanases from *Thermomonospora fusca*. *Biochemistry* **24**:7797–7804.
5. Ghangas, G. S., and D. B. Wilson. 1988. Cloning of the *Thermomonospora fusca* endoglucanase E2 gene in *Streptomyces lividans*: affinity purification and functional domains of the cloned gene product. *Appl. Environ. Microbiol.* **54**:2521–2526.
6. Gilkes, N. R., R. A. J. Warren, R. C. Miller, and D. G. Kilburn. 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *J. Biol. Chem.* **263**:10401–10407.
7. Hägerdahl, B. G. R., J. D. Ferchak, and E. K. Pye. 1978. Cellulolytic enzyme system of *Thermomonospora* sp. grown on microcrystalline cellulose. *Appl. Environ. Microbiol.* **36**:606–612.
8. Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood (ed.). 2000. *Practical Streptomyces genetics*. The John Innes Foundation, Norwich, United Kingdom.
9. Kim, T., and X. G. Lei. 2005. Expression and characterization of a thermotable serine protease (TfpA) from *Thermomonospora fusca* YX in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* **68**:355–359.
10. Kristjansson, M. M., and J. E. Kinsella. 1990. Heat stable proteinase from *Thermomonospora fusca*. *Int. J. Peptide Protein Res.* **36**:201–207.
11. Lao, G., and D. B. Wilson. 1996. Cloning, sequencing, and expression of a *Thermomonospora fusca* protease gene in *Streptomyces lividans*. *Appl. Environ. Microbiol.* **62**:4256–4259.
12. Leubner-Metzger, G., and F. J. Meins. 1999. Function and regulation of plant  $\beta$ -1,3-glucanases (PR-2), p. 49–76. In S. Muthukrishnan (ed.), *Pathogenesis related proteins in plants*. CRC Press, Boca Raton, FL.
13. Lin, E., and D. B. Wilson. 1988. Identification of a *celE* binding protein and its potential role in induction of the *celE* gene in *Thermomonospora fusca*. *J. Bacteriol.* **170**:3843–3846.
14. Lin, E., and D. B. Wilson. 1987. Regulation of  $\beta$ -1, 4-endoglucanase synthesis in *Thermomonospora fusca*. *J. Bacteriol.* **53**:1352–1357.
15. Lin, E., and D. B. Wilson. 1988. Transcription of the *celE* gene in *Thermomonospora fusca*. *J. Bacteriol.* **170**:3838–3842.
16. McGrath, C. E., and D. B. Wilson. 2006. Characterization of a *Thermobifida fusca* beta-1,3-glucanase (Lam81A) with a potential role in plant biomass degradation. *Biochemistry* **45**:14094–14100.
17. Spiridonov, N. A., and D. B. Wilson. 1999. Characterization and cloning of CelR, a transcriptional regulator of cellulase genes from *Thermomonospora fusca*. *J. Biol. Chem.* **274**:13127–13132.
18. Spiridonov, N. A., and D. B. Wilson. 1998. Regulation of biosynthesis of individual cellulases in *Thermomonospora fusca*. *J. Bacteriol.* **180**:3529–3532.
19. Stone, B. A., and A. E. Clarke. 1992. *Chemistry and Biology of (1,3)- $\beta$ -D-Glucans*. La Trobe University Press, Victoria, Australia.
20. Tomme, P., H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeysens. 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.* **170**:575–581.
21. Tseng, G. C., M. K. Oh, L. Rohlin, J. C. Liao, and W. H. Wong. 2001. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.* **29**:2549–2557.
22. Wilson, D. B. 1992. Biochemistry and genetics of actinomycete cellulases. *Crit. Rev. Biotechnol.* **12**:45–63.
23. Wilson, D. B. 2004. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *Chem. Rec.* **4**:72–82.
24. Zhang, S., C. K. Van Pelt, and J. D. Henion. 2003. Automated chip-based nanoelectrospray-mass spectrometry for rapid identification of proteins separated by two-dimensional gel electrophoresis. *Electrophoresis* **24**:3620–3632.