Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biotechnology 118 (2005) 17–34
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

The *Phanerochaete chrysosporium* secretome: Database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium

Amber Vanden Wymelenberg [a], Grzegorz Sabat [b], Diego Martinez [c], Alex S. Rajangam [d], Tuula T. Teeri [d], Jill Gaskell [e], Philip J. Kersten [e], Dan Cullen [e,*]

[a] *Department of Bacteriology, University of Wisconsin, Madison, WI 53706, USA*
[b] *Genetics and Biotechnology Center, University of Wisconsin, Madison, WI 53706, USA*
[c] *Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[d] *AlbaNova University Center, Royal Institute of Technology, 10691 Stockholm, Sweden*
[e] *USDA Forest Products Laboratory, One Gifford Pinchot Dr., Madison, WI 53726, USA*

Received 26 November 2004; received in revised form 3 March 2005; accepted 26 March 2005

Abstract

The white rot basidiomycete, *Phanerochaete chrysosporium*, employs an array of extracellular enzymes to completely degrade the major polymers of wood: cellulose, hemicellulose and lignin. Towards the identification of participating enzymes, 268 likely secreted proteins were predicted using SignalP and TargetP algorithms. To assess the reliability of secretome predictions and to evaluate the usefulness of the current database, we performed shotgun LC–MS/MS on cultures grown on standard cellulose-containing medium. A total of 182 unique peptide sequences were matched to 50 specific genes, of which 24 were among the secretome subset. Underscoring the rich genetic diversity of *P. chrysosporium*, identifications included 32 glycosyl hydrolases. Functionally interconnected enzyme groups were recognized. For example, the multiple endoglucanases and processive exocellobiohydrolases observed quite probably attack cellulose in a synergistic manner. In addition, a hemicellulolytic system included endoxylanases, α-galactosidase, acetyl xylan esterase, and α-L-arabinofuranosidase. Glucose and cellobiose metabolism likely involves cellobiose dehydrogenase, glucose oxidase, and various inverting glycoside hydrolases, all perhaps enhanced by an epimerase. To evaluate the completeness of the current database, mass spectroscopy analysis was performed on a larger and more inclusive dataset containing all possible ORFs. This allowed identification of a previously undetected hypothetical protein and a putative acid phosphatase. The expression of several genes was supported by RT-PCR amplification of their cDNAs.
Published by Elsevier B.V.

Keywords: *Phanerochaete chrysosporium*; Genome; Proteome; Secretome; Glycosyl hydrolase; Cellulase

* Corresponding author: Tel.: +1 608 231 9468; fax: +1 608 231 9262.
E-mail address: dcullen@facstaff.wisc.edu (D. Cullen).

0168-1656/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.jbiotec.2005.03.010

1. Introduction

The model wood decay fungus, *Phanerochaete chrysosporium*, completely degrades all major components of plant cell walls, including cellulose, hemicellulose, and lignin (for a review, see Eriksson et al., 1990). The ligninolytic system has been intensively studied and is generally believed to involve extracellular oxidative enzyme systems (for review, see Cullen and Kersten, 2004). Degradation of cellulose, hemicellulose and pectin has received relatively less attention, although the recently completed genome reveals impressive diversity among genes believed to encode carbohydrate-active enzymes (Martinez et al., 2004). More specifically, at least 166 sequences are predicted to encode glycoside hydrolases.

Several cellulases have been purified from defined basal media with crystalline cellulose as carbon source. Eriksson and Pettersson (1975a, 1975b) characterized multiple endoglucanase (EG) and exocellobiohydrolase (CBH) isozymes. Two CBHI isozymes, designated CBH62 and CBH58, and a single CBHII, designated CBH50, were subsequently purified (Uzcategui et al., 1991c). Genetic analysis identified the corresponding genes, and in accord with glycosyl hydrolase nomenclature of Henrissat (1991) (http://afmb.cnrs-mrs.fr/CAZY/index.html), these have been named *cel7C*, *cel7D*, and *cel6A*. Four additional *CBH1* genes have been sequenced, and all the six *cel7s* have been the subject of molecular modeling (Munoz et al., 2001). Transcripts of the *cel7* genes are present in cellulose-containing media (Covert et al., 1992b; Vanden Wymelenberg et al., 1993), but protein identification to date has been limited to CEL7C and CEL7D (Uzcategui et al., 1991b, 1991c). Five endoglucanase isozymes were partially resolved by Eriksson and Pettersson (1975a) and more recently, experimentally determined peptide sequence has been reported for two glycosyl hydrolase family 5 (GH5) isozymes (Uzcategui et al., 1991a) and a single GH28 isozyme (Henriksson et al., 1999). To date, only one endoglucanase gene, *cel61A*, has been characterized (Vanden Wymelenberg et al., 2002), although the corresponding protein has not yet been identified.

Following the synergistic activities of CBHs and EGs, cellobiose and related oligosaccharides are converted to glucose by β-glucosidases. Several isozymes have been purified from cellulose-grown *P. chrysosporium* cultures (Deshpande et al., 1978; Igarashi et al., 2003; Smith and Gold, 1979) and a single gene, *bgl1*, identified (Li and Renganathan, 1998).

Relatively little is known about the enzymes involved in hemicellulose and pectin degradation in *P. chrysosporium* (Castanares et al., 1995; Copa-Patino et al., 1993). Two cDNAs believed to encode endoxylanases have been sequenced (GenBank accessions AAG44993, AAG44995), and both were recently expressed in *Aspergillus niger* (Decelle et al., 2004).

In addition to the abovementioned hydrolytic enzymes, oxidative enzymes have been purified from cellulose-containing media (Eriksson et al., 1974). In this context, cellobiose dehydrogenase has received considerable recent attention (for review, see Cameron and Aust, 2001; Henriksson et al., 2000a, 2000b). In addition to its direct action on cellobiose, CDH has been implicated in cellulose degradation via Fenton chemistry (Mansfield et al., 1997). In brief, Fenton systems involve the spontaneous reaction of peroxide and reduced iron to generate highly reactive hydroxyl radicals. Enzymes capable of iron reduction or peroxide production may be involved in hydroxyl radical generation, and CDH is capable of generating both reactants (Fe(II) and H_2O_2). Peroxide-generating enzymes of *P. chrysosporium* also include the copper radical oxidase, glyoxal oxidase, and the FAD-dependent pyranose 2-oxidase and glucose oxidase. Recent analysis of the *P. chrysosporium* genome reveals a complex array of genes encoding six additional copper radical oxidases and five FAD-dependent oxidases (Martinez et al., 2004). With the exception of glyoxal oxidase (Kersten and Cullen, 1993) and pyranose 2-oxidase (de Koker et al., 2004), nothing is known about the expression of these oxidase genes.

As a first step toward clarifying the role of individual genes in cellulose degradation, we performed liquid chromatography–tandem mass spectroscopy (LC–MS/MS) analysis on filtrates from *P. chrysosporium* strain RP-78 grown in standard cellulose-containing medium. Alternative databases and sample treatments were evaluated, and a total of 50 specific gene products were identified.

2. Materials and methods

2.1. Culture conditions

As previously described (Covert et al., 1992b; Vanden Wymelenberg et al., 1993, 2002), 200 ml defined basal salts (Eriksson and Hamp, 1978) were supplemented with wood-derived crystalline cellulose (0.4% Avicel PH-101, Fluka Chemie AG, Buchs, CH) in 2 l flasks and inoculated with *P. chrysosporium* RP78 to a final concentration of 5×10^4 spores per ml. Duplicate cultures were grown at 37 °C, 250 rpm, and harvested after 6 days.

2.2. Protein analysis

Following filtration through nylon screen, culture fluids were combined and briefly centrifuged. Avicel was removed via vacuum filtration through GF/A glass filters (Whatman International, Maidstone, UK). The filtrate was concentrated in an Amicon ultrafiltration device using a 5000 MWCO polyethersulfone membrane (Millipore Corp., Bedford, MA). Ten- and 100-fold concentrations were achieved for CBIND (Novagen Inc., Madison, WI) and non-CBIND preparations, respectively. The non-CBIND preparation was concentrated an additional 10-fold with a Microsep 3K Omega centrifugal device (Pall Life Sciences, Ann Arbor, MI).

Concentrated soluble protein was either loaded directly onto SDS-PAGE gels or first fractionated by cellulose binding. The entire concentrate (∼30 ml) from ultrafiltration was applied to a CBIND 900 cartridge, according to the manufacturer's instructions. Elution was in 3 ml ethylene glycol, and the eluate was diluted with 3 ml H_2O and placed in a Microsep 3K device for further concentration to a final volume of 100 μl. Eighty-five microlitres was acetone precipitated and re-suspended in 15 μl H_2O for SDS-PAGE. An equal volume of Laemmli buffer (Bio-Rad Laboratories Inc., Hercules, CA) was added to samples before loading onto a 12.5% SDS-PAGE gel (Criterion Tris–HCl Ready Gel, Bio-Rad Laboratories). Gels were stained with Coomassie Blue R-250 (Bio-Rad Laboratories) to estimate the protein abundance and MW distribution. Total protein fraction resolved on a gel was manually fractioned with a surgical blade into 10, 2 mm long and 5 mm wide strips. These gel strips were further cut into ∼1 mm pieces and placed in individual siliconized 1.5 ml microcentrifuge tubes (Fisher Scientific) for subsequent enzymatic digestion.

'In gel' digestion and mass spectrometric analysis was performed as described (www.biotech.wisc.edu/ServicesResearch/MassSpec/ingel.htm.) In short, gel pieces were de-stained completely in MeOH/H_2O/NH_4HCO_3 (50%:50%:100 mM), dehydrated for 10 min in acetonitrile/H_2O/NH_4HCO_3 (50%:50%:25 mM) and then once more for 1 min in 100% acetonitrile. The samples were dried in a Speed-Vac for 5 min, reduced in 25 mM DTT (dithiothreitol in 25 mM NH_4HCO_3) for 30 min at 56 °C, alkylated with 55 mM IAA (iodoacetamide in 25 mM NH_4HCO_3) in darkness at room temperature for 30 min, washed in H_2O for 20 min, equilibrated in 25 mM NH_4HCO_3 for 10 min, dehydrated for 10 min in acetonitrile/H_2O/NH_4HCO_3 (50%:50%:25 mM) and then once more for 1 min in 100% acetonitrile. Following drying, samples were rehydrated with 25 μl of trypsin solution (20 ng/μl trypsin (*Sequence Grade Modified*, PROMEGA Inc., Madison, WI) in 25 mM NH_4HCO_3) or with 25 μl Asp-N solution (8 ng/μl endoproteinase Asp-N, Roche Biochemicals) in 50 mM Na_2HPO_4, 5 mM Tris–HCl (pH 8.0). Additional buffer overlay (∼15 μl) was provided to keep gel the fragments immersed. The digestions were conducted overnight (18 h) at 37 °C, then terminated by acidification with 0.1% TFA (Trifluoroacetic acid). Peptides generated from digestions were extracted in two subsequent steps, first with an equal volume of 0.1% TFA (∼50 μl) and vigorous vortexing for 15 min, then with the same volume of acetonitrile/H_2O/TFA (70%:25%:5%) and vortexing. The collected peptide solution was dried completely in a Speed-Vac, re-suspended in 50 μl of 0.1% TFA, and solid phase extracted (*C18 SPEC-PLUS*™*-PT* pipette tips, Varian Inc., Lake Forest, CA). Peptides were eluted off the C18 column with acetonitrile/H_2O/TFA (70%:25%:0.2%), dried in a Speed-Vac and finally reconstituted with 26 μl of 0.1% formic acid.

Peptide fractions were individually analyzed by nanoLC–MS/MS using 1100 series LC/MSD Trap SL spectrometer (Agilent, Palo Alto, CA). Chromatography of peptides prior to mass spectral analysis was accomplished using C18 reverse phase HPLC trap column (Zorbax 300SB-C18, 5 μM, 5 mm × 0.3 mm, Agilent) and separation column (Zorbax 300SB-C18,

3.5 μm, 0.075 mm × 150 mm, Agilent) onto which 22 μl of each extracted peptide fraction was automatically loaded. An Agilent 1100 series HPLC delivered solvents A: 0.1% (v/v) formic acid in water; and B: 95% (v/v) acetonitrile, 0.1% (v/v) formic acid, at either 10 μl/min to load the sample, or at 0.28 μl/min to elute peptides directly into the nano-electrospray. The elution was for 60 min in a gradient from 20 to 80% (v/v) solvent B. Peptides eluting from the HPLC column/electrospray source were trapped in an ion cell and sequential MS/MS spectra spanning from 300 to 2200 *m/z* were generated for the 10 most abundant ions present at each switching event. Redundancy was limited by dynamic exclusion. MS/MS data were converted to matrix generic format (mfg) files using Data Analysis Software (Agilent).

Resulting mgf files were used to search an amino acid sequence database translated from the 11,722 publicly available *P. chrysosporium* gene models (www.jgi.doe.gov/whiterot). Briefly, these genes were predicted using a protein homology-based program, Genewise (Birney and Durbin, 2000), followed by GrailEXP (Xu and Uberbacher, 1997), an ab initio and mRNA homology gene finder incorporating expressed sequences. The mgf files for each sample were also searched against a larger database containing translations of all potential ORFs greater then 20 amino acids. To reduce redundancy, the coding regions within the existing 11,722 models were subtracted from this ORF database based on genomic position. (Both databases and the secretome compilation (below) are freely available as compressed fasta files and Excel worksheet, respectively, upon request.) Spectrum Mill MS Proteomics Workbench (Agilent) and an in-house licensed Mascot search engine (Matrix Science, London, UK) were used to identify the peptides. Throughout, protein similarity scores are based on the Smith–Waterman algorithm (Smith and Waterman, 1981), using the BLOSUM62 matrix.

2.3. Reverse transcription polymerase chain reaction

P. chrysosporium RP78 RNA was extracted from cellulose-grown cultures using the RNeasy Maxi kit (Qiagen Inc., Valencia, CA) according to the manufacturer's instructions, with modifications. Specifically, mycelia were harvested by filtration through Miracloth (CN Biosciences, La Jolla, CA), snap frozen in liquid N_2, and ground to a fine powder with a prechilled mortar and pestle. Ground mycelia were transferred to 50 ml conical tubes, and 15 ml RLC buffer (RNeasy Plant Mini Kit, Qiagen) were added. Tubes were shaken vigorously and the contents were forced several times through sterile syringes fitted with 18-gauge needles. The lysates were centrifuged 5000 × *g* for 10 min in a swinging bucket rotor. One volume 70% ethanol was added to the supernatants and shaken vigorously. The lysates were loaded onto RNeasy maxi columns in 15 ml increments and spun for 5–10 min at 5000 × *g* until total sample volumes were loaded onto the columns. Fifteen millilitres of RW1 buffer were added to the columns and spun 5 min as above, followed by 10 ml RPE buffer and a 2 min spin. Ten millilitres of additional RPE buffer were added and the columns were centrifuged for 10 min at 5000 × *g* to dry the membranes. To elute the RNA, the columns were transferred to conical collection tubes, 0.8 ml RNase-free H_2O were loaded directly onto the membranes, incubated 1 min room temperature, and centrifuged for 3 min at 5000 × *g*. The elution was repeated with an additional 0.8 ml H_2O. Magnetic capture of poly(A) RNA with oligo-$(dT)_{25}$ beads followed the manufacturer's recommendations (Dynal, Great Neck, NY). mRNA was also isolated from colonized wood chips (Akhtar, 1997), as previously described (Janse et al., 1998; Vallim et al., 1998).

Polyadenylated RNA was reverse transcribed with oligo dT_{15} primer and cDNA was PCR amplified, as previously described (Janse et al., 1998; Vallim et al., 1998; Vanden Wymelenberg et al., 1993, 2002) using gene-specific primers (Table 1). Primers flanked introns to allow the differentiation of genomic and cDNA amplicons. The cDNAs from cellulose-grown cultures were directly sequenced to confirm the specificity of amplification.

3. Results

3.1. Genome mining

A database encompassing translations of the current 11,722 gene models was submitted to SignalP v.2, TargetP v.1 and TMHMM v.2 (all available at http://www.cbs.dtu.dk/services/) for the identification of secreted proteins. Our screening excluded sequences

Table 1
Primers used for RT-PCR amplification of select cDNAs[a]

Model[b] (gene)	Primer sequence	Length (nt)[c] (gene/cDNA)
pc.143.21.1 (*xyn10D*)	CTCACTGCAGTACCCTCGTCA AACGCCGGCTTCTTCAC	824/606
pc.163.7.1 (*cel12A*)	GCCGACCGCGTGGAACT GCCGGACCGTAAATACATCATCAG	670/390
pc.11.202.1 (*aga27*)	GTCCGTGTACGTCCGTGATAGATT TACCGCCTTCCATACCTTTGTTT	640/240
pc.69.19.1 (*pho1*)	ACAAGACAGTCGCTGACGATG TGACGCCTTCAGTCAAAGCAAC	560/510
gx.47.9.1 (*scp1*)	ATGCCGGCAGTCAACTTCGTCAAA CGCGCCCATGCCGTCTAC	828/620
pc.19.17.1 (*ogt1*)	GGATGGGACGCGTTGGATTA CCGACTGCATGAGGTGAGACA	923/411

[a] Genes where identifying peptides gave Spectrum Mill scores <13.
[b] Automated gene model predictions at www.jgi.doe.gov/whiterot.
[c] Processed transcripts (mRNAs) yield RT-PCR products that are shorter than genomic DNA products and thereby help differentiate cDNA amplicons from those arising from contaminating DNA.

with signal peptides greater than 40 amino acids of the N-termini, one or more transmembrane domains, and mitochondrial targeting signals. Final analysis predicted 268 soluble, putative secreted proteins. This computational approach, no doubt, greatly underestimates the secreted proteins, mostly because of the common problem of inaccurate and incomplete gene models.

In silico analysis of the *P. chrysosporium* genome database revealed several new genes of possible importance to cellulose and hemicellulose degradation. Endoglucanase isozymes EG28, EG44, and EG38/36 purified from *P. chrysosporium* strain K3 (Henriksson et al., 1999; Uzcategui et al., 1991a) were matched to strain RP78 gene models pc.163.7.1 (*cel12A*), pc.6.6.1 (*cel5B*), and pc.26.11.1 (*cel5A*), respectively. Careful inspection of these gene models suggested that certain intron/exon junctions were inaccurately assigned and these were corrected by RT-PCR amplification and sequencing of full length cDNA for *cel5A* and *cel12A* and partial cDNA for *cel5B*. The cDNA sequences of *cel5A* and *cel12A* have been deposited (GenBank accessions AY682743, AY682744, respectively).

3.2. Protein identification

Several approaches were attempted to maximize the number of extracellular proteins detected. Using conservative cut-off scores (e.g. Spectrum Mill MS peptide scores >13; Mascot scores >40) the number of reliable identifications was highest in the sample digested with trypsin without fractionation by cellulose binding. Under these conditions, a total of 109 peptide sequences were unambiguously assigned to 35 specific gene models (Tables 2–4). Digestion of the same protein sample with AspN yielded 55 significant peptide sequences, matching 23 gene models. Protein eluted from cellulose cartridges and digested with AspN gave 36 peptides corresponding to 12 genes. When digested with trypsin, the same cellulose-bound protein sample yielded 16 significant peptide sequences belonging to 8 gene models. Thus, analysis of total protein concentrated by ultrafiltration was found to be most informative. However, we also observed that several peptides were only detected after concentration with the cellulose cartridges (Fig. 1), e.g. CBHI isozyme (Cel7E), CBHII, putative xylanase Xyn10B, putative endoglucanase Cel61B. While these examples correspond to gene models with clear cellulose binding domains, other peptides eluted from cellulose columns do not have conserved or obvious binding modules, i.e. lignin peroxidase isozyme H2, putative xylanase *xyn10D*, and putative endoglucanase *cel45A*.

Searches of the larger database containing all potential ORFs greater than 20 codons proved useful for extending the carboxy- and *N*-terminal regions

Table 2
Peptide identifications matching previously known *P. chrysosporium* proteins and/or genes[a]

Model[b]	Gene[c]	Function	Peptide sequence[d]	Sample[e] (number of peptides)	Protein	cDNA[f]	Comment[g]
gx.66.3.1	*gly3B*	β-glycosidase	ELLHQMEQETIVLLENR	1(1)	Abbas et al. (2004)	–	CBM1
pc.26.11.1**	*cel5A*	Endoglucanase	VIFGIMNEPHDIPSISTWVNSVQGAVNAIR WNGGIIAQGGPTDAQFQSIWTQLAQK YDQTVQAALNSGPNVGVIIDLHNYAR	1(4), 2(2), 4(1)	Uzcategui et al. (1991a)	*AY682743*	CBM1; EG38/36
pc.6.6.1	*cel5B*	Endoglucanase	LQAATQWLQQNNLK DPNNNVAIEMHQYL DQTYLSGLQTIVSYITGKGGYAL	1(1), 2(1), 4(2)	Uzcategui et al. (1991a)	–	CBM1; EG44
pc.3.82.1*	*cel6*	CBHII	DEIHYINALAPLLQQAGW DPTLSSKAASVANIPTFTWL	4(2)	Uzcategui et al. (1991c)	Tempelaars et al. (1994)	CBM1; CBH50
pc.116.24.1*	*cel7E*	CBHI	DAAAFTPHPCTTDGQTRCSG	4(1)	–	Munoz et al. (2001)	CBM1
pc.139.26.1*	*cel7C*	CBHI	GLTVDTSKPPTVVTQFITNDGTSAGTL-TEIR IPGIDPVNSITDNFCSQQK DTSKPFTVVTQFITN	1(3), 2(2), 4(10)	Uzcategui et al. (1991c)	Sims et al. (1988)	CBM1, CBH62
pc.33.51.1*	*cel7D*	CBHI	VYLMADDTHQLLK NTGLCDGDGCGFNSFR TAFGCTNWFAQK	1(7), 2(4), 4(8)	Uzcategui et al. (1991c)	Vanden Wymelenberg et al. (1993)	CBM1, CBH58
pc.95.47.1*	*cel7F*	CBHI	YGTGYCDSQCPK	1(1)	–	Munoz et al. (2001)	CBM1
gw.74.21.1*	*xyn10A*	Xylanase A	LYINEYNIEYAGAK MTLPSTPALLAQQK DDNTMFGQITPANSMKW	1(2), 4(1)	Decelle et al. (2004)	AF301903	CBM1
gx.42.28.1*	*xyn10B*	Xylanase	LYINEYNIEFAGAK GQVYAWDVVNEPFNDDGTWRT DIRMTLPSTPALLAQQQT	1(1), 2(2)	Decelle et al. (2004)	–	CBM1
pc.143.21.1	*xyn10D*	Xylanase	mtlpstpallqaqk (score: 12.45)	3(1)	Abbas et al. (2004)	–	
pc.120.11.1*	*xyn11A*	Xylanase B	DFGTYNPAVSLTHKGTLTS	2(1)	Decelle et al. (2004)	AF301905	CBM1
pc.163.7.1*	*cel12A*	Endo-glucanase	advsydiwfgk (score: 10.19)	1(1)	Henriksson et al. (1999)	*AY682744*	EG28
pc.11.202.1*	*aga27A*	α-Galacto-sidase	vyiqtqpyevfdlwqk (score: 11.2)	1(1)	Brumer et al. (1999)	Hart et al. (2000)	
pc.67.67.1*	*exg55A*	Exo-1,3-β-glucanase	WSGASSGHLQGSLVLNNIQLTNVPVAVGK VSSPLVVLYQTQLIGDAK GDGNTDDTAAIQAAINAGGR	1(5), 2(7)	Abbas et al. (2004)	–	

pc.5.157.1*	*cdh1*	Cellobiose dehydro-genase	YLEQSFNVVSQLLK VFGTNNLFIVDAGIIPHLPTGNPQGTLMS-AAEQAAAK GTVLTPPWLVNPVDK	1(7), 2(4), 4(1)	Henriksson et al. (2000a)	Raices et al. (1995)	CBM
pc.19.174.1*	*lipD*	LiP isozyme H2	DSIAISPKLQSQGKFGGGGA DQVLARML	4(2)	Tuisel et al. (1990)	de Boer et al. (1987)	
pc.111.22.1*		Protease	GWDPVTGLGTPNFAALK ATQSSNTLGVSGFIDQFANQADL-TTFLNR (manually curated model) SLANNLCNAYAQLGAR (manually curated model)	1(3)	Datta (1992)	–	

[a] Six-day-old Avicel-containing media, as described in Section 2.

[b] Automated gene model predictions at www.jgi.doe.gov/whiterot. Models with asterisk were computationally predicted as part of the 'secretome'. Those with two asterisks revealed secretion signals following manual identification of the translational start codon. Models without asterisks were *N*-terminally incomplete, containing multiple short exons and uncertain translational starts.

[c] Glycosyl hydrolase families, as described by Henrissat and co-workers (http://afmb.cnrs-mrs.fr/CAZY).

[d] Peptide sequences shown had highest Spectrum Mill MS score. Less confidence is assigned to lower case peptides with scores <13.

[e] Soluble protein concentrated by ultrafiltration (sample1 and 2) or by cellulose binding (3 and 4), followed by SDS-PAGE fractionation and digestion with trypsin (1 and 3) or with Endoproteinase Asp-N (2 and 4). The number of significant peptides are shown parenthetically and varied with method of concentration and enzymatic digestion.

[f] GenBank accession in script represent full length cDNAs sequenced as part of this study.

[g] Predicted carbohydrate-binding module (CBM), followed by protein designation, if any.

Table 3
Peptides corresponding to previously uncharacterized glycosyl hydrolases of *P. chrysosporium*[a]

Model[b]	Family[c]	Function	Peptide sequence[d]	Sample (number of peptides)[e]	Comment[f]
gx.102.2.1**	GH5	Putative β-mannanase	YVDEPTILAWELANEPR AYQDIVNAGSTVVR DIVNAGSTVVRTLGFN	1(2), 2(1)	CBM1; *man5C*
pc.30.50.1**	GH5	Putative β-mannanase	YVNEPTIMAWELSNEPR AVQDIANAGSTVVRT SANKPVIMEEFGVTSDQTDVYK	1(4), 2(2)	CBM1; *man5D*
pc.5.251.1	GH10	Putative xylanase	DSANAKLQAVVNLVKQVNNGGTKLI DSAVTIAQFGQLTPENSMKW DPNAKLYIN	4(4)	CBM1, *xyn10C*
pc.5.247.1*	GH12	Putative xyloglucanase/ endoglucanase	STFNIQLNQGINK TTWSWSGGSGGVK**	1(2)	*cel12B*
pc.180.9.1*	GH15	Putative glucoamylase	SIYGVNSGIASNAAVATGR VIVDQUAAGQDTSTR AGVVIASPSTTNPDYLYTWVRD	1(5), 2(2)	CBM20, gla15A
pc.22.125.1*	GH16	Putative β-1,3-glucanase	VNYVNQATAVAK AYTTHVAIIDVR NAYWDIAAVR	1(3)	*gly16A*
pc.11.89.1	GH18	Putative chitinase	VYIGAPASSSAAGTGYQSISTLQNIATQMR	1(1)	*chi18A*
pc.63.54.1*	GH18	Putative chitinase	VFIGAPASSTAAGSGYVDIGTLSNILQQTK DNWAKTTSPNKNVKVFIGAPASSTAAGSGV	1(1), 2(1)	*chi18B*
gx.1.81.1	GH28	Putative polygalacturonase	ttltvnsgaer (score: 12.5) qfgvlidqsypdtlgtptgvtisdinftgak (score: 12.3) DATVSGVIIQGNTITNS	1(2), 2(1)	*epg28A*
gx.32.31.1	GH28	Putative rhamnogalacturonase	DIYNVGGNQMFMIK GASQWAFQLDGVISR NTICNVLDYGGSIGSSDIGPAIQSAFNNCVLK	1(5); 2(2)	CBM1, *rhg28C*
pc.71.18.1*	GH43	Putative β-glycosidase	WFSASSLAGPWSAQQDIAPSATR** qndalspiagtmistsnvverpk (score: 10.1) diapsatrtwysqnaf (score:11.8)		*glyl43A*
pc.55.40.1	GH45	Putative endoglucanase	AASGFTAAINQLAFGSAPGLGAGDACGR dacgrcfaltgnh (score11.1)	3(1); 4(1)	*cel45A*
pc.50.85.1	GH51	Putative α-L-arabinofuranosidase	LAFPTMQSSSGEAAFMTGLER TIDYTAPPFSVSVIR DSSTGPGALRASLGHPEPFTLHYVEIGNE	1(2), 2(1)	*arb51A*
gw.48.27.1**	GH61	Putative endoglucanase	DTLQWFKI	4(1)	CBM1, *cel61B*
pc.129.11.1*	GH74	Putative xyloglucanase/ endoglucanase	FNPVWTTNVDLDTAGNVPSTIVR VTSLPDSGTYAPDPTDSSGINSDK INDAAHGFGSISSNVLTADPR	1(9), 2(9)	*gly74A*
pc.19.88.1*	GH74	Putative xyloglucanase/ endoglucanase	DGGKTFAKQATLGTSSSVSTVIVHPAVTG	2(1)	CBM1, *gly74B*

Table 3 (*Continued*)

Model[b]	Family[c]	Function	Peptide sequence[d]	Sample (number of peptides)[e]	Comment[f]
gx.24.50.1*	GH88	Putative delta 4,5 unsaturated glucuronyl hydrolase	ILATAQQTQNPPK	1(1)	*gly88A*

[a] Six-day-old avicel-containing media, as described in Section 2.
[b] Automated gene model predictions at www.jgi.doe.gov/whiterot. Asterisks refer to *N*-terminal completeness as in Table 2.
[c] Glycosyl hydrolase families, as described by Henrissat and co-workers (http://afmb.cnrs-mrs.fr/CAZY).
[d] Peptide sequences shown had highest Spectrum Mill MS score. Less confidence is assigned to lower case peptides with scores <13. In two instances (marked **), Mascot analyses of same spectra revealed an additional sequence. The GH51 sequence in bold represents a peptide assigned to an ORF located 3′ of model pc.50.85.1. Manual inspection of model showed this corresponds to the COOH terminus of *arb51A*.
[e] Soluble protein concentrated by ultrafiltration (sample1&2) or by cellulose binding (3 and 4), followed by SDS-PAGE fractionation and digestion with trypsin (1 and 3) or with endoproteinase Asp-N (2 and 4). The number of significant peptides varied with method of concentration and enzymatic digestion. Most peptides were detected by trypsin digests of concentrated protein, but certain proteins predicted to contain carbohydrate-binding modules were only detected by cellulose binding, i.e. putative xylanases pc.143.21 and pc.251.1, putative endoglucanases pc.55.40.1 and gw.48.27.1.
[f] Predicted carbohydrate-binding module (CBM), if any. Italicized gene designations adhere to family designations. Those with '*gly*' prefix belong to functionally heterogeneous families.

of incomplete gene models pc.50.85.1 (putative α-L-arabinofuranosidase) and pc.69.19.1 (putative acid phosphatase), respectively. In the case of the acid phosphatase, no other peptides were detected and the Spectrum Mill score was relatively low (Table 4). However, RT-PCR identified the cDNA corresponding to model pc.69.19.1, thereby supporting the mass spectroscopy identification (Fig. 2). Another ORF database match was model pc.168.16.1, a hypothetical protein conserved in many fungal species. Thus, of the 50 expressed genes, 2 (pc.69.19.1, pc.186.16.1) were identified exclusively with the expanded ORF database.

The 50 identifications are categorized as those 18 genes for which nucleotide or partial protein sequence was previously reported (Table 2), peptides corresponding to 17 previously unknown glycosyl hydrolases (Table 3), and peptides corresponding to another 15 previously unknown genes of diverse function (Table 4). Results for specific protein families follow.

3.3. Glycosyl hydrolase families GH6 and GH7 (cellobiohydrolases)

Consistent with previous studies of *P. chrysosporium* in cellulose-containing media (Uzcategui et al., 1991c), CBH1 isozymes corresponding to CBH62 (*cel7C*) and CBH58 (*cel7D*) were identified (Table 2).

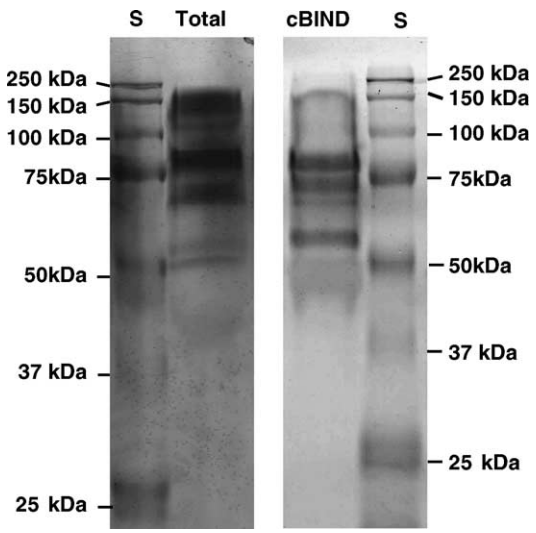

Fig. 1. SDS-PAGE of total extracellular protein and CBIND-treated protein. The "total" protein lane represents approximately 1.5 ml culture fluid whereas the "CBIND" lane represents, approximately, 225 ml culture fluid, as estimated by the nominal-fold concentration of samples after ultrafiltration (see Section 2). Comparisons by visual inspection of SDS-PAGE gels with serial dilutions indicate less than 1% of total protein is recovered from the CBIND cartridge, providing a sample highly enriched in cellulose-binding proteins. Molecular weight standards (S) are indicated.

Table 4
Peptides corresponding to uncharacterized gene models of *P. chrysosporium*[a]

Model[b]	Gene[c]	Putative function	Peptide sequence[d]	Sample (number of peptides)[e]	Similarity[f]	Comment
pc.100.54.1**	*axe1*	Acetylxylan esterase	FAISNWGVDPNR MQVWHGTADTTLYPQNFFEEIK VAAAYPGYTGPYPK	1(3)	*Aspergillus ficuum* (AF331757); 752	CBM1; carbohydrate esterase family 1
gx.91.6.1**	*gox1*	Glucose oxidase	LTQGADVNDGHPTGVFVTPLDPAR WAASQAAAGGGVSADALQEIMR GQDVQYDSWNALLEPEEK	1(4)	*Aspergillus niger* (AJ294936); 480	FAD-dependent peroxide generation
pc.69.19.1*	*pho1*	Acid phosphatase (possible phosphorous scavenger)	**davrfgiqtlspkffggpp** (score 12.0)	2(1)	*Yarrowia lipolytica* (P30887); 154	Peptide lies in ORF upstream of model
pc.12.138.1	*asp1*	Aspartyl protease A1	ATGATLDNNTGLLR FYSVFDTTNKR LGLATTPFTTATTN	1(3); 2(3)	*Polyporus tulipiferae* (P17576); 1294	
pc.127.30.1*	*asp2*	Aspartyl protease A1	TPNTGFETIIDSGTTIMYGPPSAVK LASSGSTLFLGGTDTSK YTGSIEYHAIDTSTGFWQAK	1(3); 2(1)	*Amanita muscaria* (AJ295271); 1170	
gx.47.9.1	*scp1*	Serine carboxypeptidase	lwnqifnaavqvnpafniyr (score 10.83)	1(1)	*A. niger* (S78072); 984	
pc.12.135.1*	*rnt1*	RNase	GAEAVAFFNTVVSLFK GSLLDGQFVPIDAPEK TLPTYDWLSSAGITPSTSK	1(3)	*Irpex lacteus* (AAB35888); 1021	RNase T2 family
gx.38.23.1	*gta1*	Glutaminase	SAQPLGDWYETTDGSVEGFR YLLLPLFEYQATGQYPNK DLFISAVHSYAADGK	1(12); 2(9)	*A. nidulans* (AB029553); 1457	
pc.19.17.1	*ogt1*	O-linked GlcNAc transferase	tmsppvvsllpfntfiyplsprtir (score 10.6)	1(1)	*N. crassa* (AL513464); 1529	Glycosyl transferase family 41
gx.10.71.1**	*ale1*	Aldose-1-epimerase	LDITTNNPATQVYTATYWLNTPR VVEVDGDAIPTGNFIDVTGTPW-DFR DAFQGSENILNHHLHV	1(2); 2(1)	*Vibrio cholerae* (NP_231234); 312	

pc.97.36.1	*thn1*	Thaumatin-like	GPFDSTGFPVGCK IDNDKGCGVASCPVDLGPNCPA-PLK	1(2)	*A. thaliana* (Q9C9Q0); 350	
pc.79.55.1	*thn2*	Thaumatin-like	gcgvascpvdlgpncpaplk (score 11.3) DLGPNCPAPLKGPF	1(1); 2(1);	*A. thaliana* (Q9SZP5); 339	
pc.140.25.1		Hypothetical protein	FAVSGLPANVFGMAGK	1(1)	*Paramecium bursaria* protein (NC_000852); 68	Model hits *P. chrysosporium* EST
pc.168.14.1		Hypothetical protein (possible peptidoglycan hydrolase)	IFTYGGVGPTGSCTPGDLSTVK DANGNPYGSFAIQTYR	1(2)	*Streptococcus mutans* (smu:SMU.22) protein; 61	Model hits *P. chrysosporium* EST
pc.168.16.1**		Hypothetical protein	**TGIVDAPVDPADNIQWVSLSALD-KPDGTPWGK** **faqqvyr** (score 11.1)	1(2)	*N. crassa* (NCU06806); 56	Model not in original database. Peptide identified in ORF database

[a] Six-day-old Avicel-containing media, as described in Section 2.

[b] Automated gene model predictions at www.jgi.doe.gov/whiterot. Asterisks refer to *N*-terminal completeness as in Table 2.

[c] Gene designations only used for those sequence with Smith–Waterman similarity scores >150.

[d] Peptide sequences shown had highest Spectrum Mill MS score. Less confidence is assigned to lower case peptides with scores <13. Sequences in bold represents peptides assigned to ORFs excluded from the current 11,722-protein models database. These were identified by searches of the larger ORF database.

[e] Soluble protein concentrated by ultrafiltration (sample1 and 2) or by cellulose binding (3 and 4), followed by SDS-PAGE fractionation and digestion with trypsin (1 and 3) or with endoproteinase Asp-N (2 and 4).

[f] Best NCBI blastP hit to the model. Species (accession); Smith–Waterman score using BLOSUM 62 matrix.

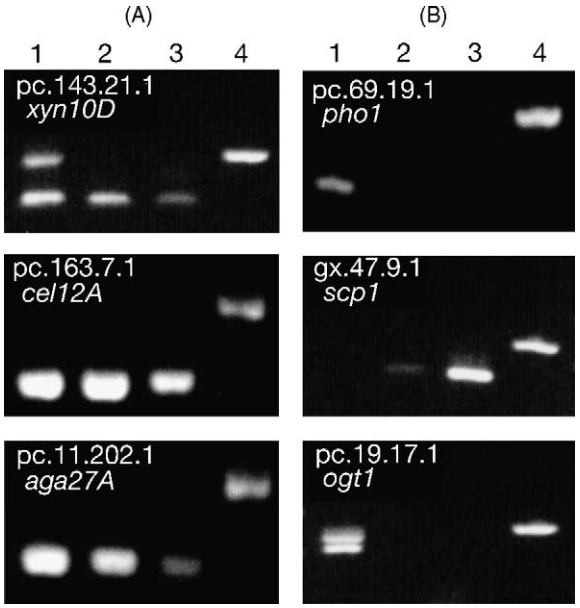

Fig. 2. RT-PCR amplification of genes with low scoring peptides. Panel A and B gene models are listed with peptide scores in Tables 2 and 3, respectively. Gene-specific PCR primers and expected lengths of cDNA and genomic products are listed in Table 1. Lanes 1–3 are amplifications with RNA from 3-day-old cellulose (avicel) cultures, from 6-day-old cellulose (avicel) cultures and from colonized wood chips, respectively, as described in Section 2. Lane 4 contains amplicons from *P. chrysosporium* genomic DNA. A second larger band in lane 1 of ogt1 and xyn10D amplifications indicates genomic DNA contamination of RNA from 3-day-old cultures.

Peptides matching *cel7E* and *cel7F* were identified for the first time. Also consistent with previous investigations, peptides corresponding to CBH2 (*cel6*) were observed.

3.4. Glycosyl hydrolase family GH5 (endoglucanases)

Of 18 putative GH5 gene models within the *P. chrysosporium* genome (Martinez et al., 2004), four of the matching peptides were detected in Avicel cultures (Tables 2 and 3). Comparisons of our corrected models with experimentally determined amino acid sequences (Uzcategui et al., 1991a) show that *cel5A* and *cel5B* correspond to isozymes EG38/36 and EG44, respectively (Table 2). Mass spectroscopy also revealed two new GH5 family members corresponding to gene models gx.102.2.1 and pc.30.50.1 (Table 3). These models are most closely related to β-mannanases within the GH5 family, and have been designated accordingly as *man5C* and *man5D*. All four GH5 model proteins have a highly conserved carbohydrate-binding module (CBM family 1) at their N-termini.

3.5. Glycosyl hydrolase family GH61 (endoglucanases)

A minimum of 17 putative *cel61* sequences were recognized in the *P. chrysosporium* genome (Martinez et al., 2004), and a single cDNA (*cel61A*) was previously characterized (Vanden Wymelenberg et al., 2002). The corresponding protein was not detected. Instead, a protein matching gene model gw.48.27.1 was identified. Designated *cel61B*, the gene has a carboxy terminal CBM1 domain. The predicted sequence is 61 and 36% identical to an *Agaricus bisporus* cellulase (GenBank Q00023) and *Trichoderma reesei* EG4 (GenBank O14405), respectively.

3.6. Glycosyl hydrolase family GH12 (endoglucanases)

In silico comparisons of gene models (Martinez et al., 2004) with experimentally determined amino acid sequence (Henriksson et al., 1999) matched model pc.163.7.1 with endoglucanase EG28 (Table 2). As mentioned above, we amplified and sequenced the full-length cDNA of this *cel12A* gene, and a closely related gene, pc.5.247.1 (*cel12B*), was identified in the database. Both proteins were detected, but a low Spectrum Mill score was obtained for *cel12A* (Table 2). Despite the low score, RT-PCR identified transcripts of gene model pc.163.7.1 under all conditions tested (Fig. 2).

3.7. Glycosyl hydrolase families GH10 and GH11 (xylanases)

Previously sequenced by Simms and co-workers (unpublished, GenBank accessions AF301903, AF301905), cDNAs encoding putative xylanase A and xylanase B correspond to gene models gw.74.21.1 (*xyn10A*) and pc.120.11.1 (*xyn11A*) (Martinez et al., 2004). Mass spectroscopy detected peptides matching both genes. A relatively low Spectrum Mill score was obtained for *xyn10D* (pc.143.21.1)(Table 2),

but subsequent amplification of the *xyn10D* cDNA supported the mass spec identification (Fig. 2). Our analysis also revealed a previously unknown GH10 corresponding to gene model pc.5.251.1, which we designate *xyn10C* (Table 3). Except *xyn10D* (pc.143.21.1), the GH10s are each predicted to encode a single CBM1 at their N-termini.

3.8. Glycosyl hydrolase family GH3

The β-glucosidase gene, *bgl3A* was previously shown to be expressed under cellulose induction (Li and Renganathan, 1998), although at relatively low levels (Vanden Wymelenberg et al., 2002). The *bgl3A* protein was not detected in this study. Genome analysis (Martinez et al., 2004) indicates a minimum of 12 GH3-like sequences and our mass spectral data identified a peptide corresponding to GH3 model gx.66.3.1 (Table 2). Designated *gly3B*, the predicted amino acid sequence lacks a recognizable carbohydrate-binding module and is only 25% identical to *bgl3A*. Recent investigations show that purified BGL3A has substantial glucan 1,3-*β*-glucosidase activity (Igarashi et al., 2003). Most recently, Abbas et al. (2004) identified GLY3B peptides in culture filtrates of *P. chrysosporium* grown in media containing powdered oak wood. The GH family 3 contains β-glucosidases and β-glucanases among other activities, and further enzymatic studies will be required to reveal the true specificity of these two genes.

3.9. Glycosyl hydrolase family 28

We detected two members of glycosyl hydrolase family 28 (Table 3). The putative polygalacturonase gene, gx.1.81.1 (*epg28A*), encodes a protein that is 58% identical to a *Chondrostereum purpureum* polygalacturonase (Williams et al., 2002). Peptides corresponding to model gx.32.31.1 were also detected. This gene, designated *rhg28C*, is closely related to rhamanogalacturonases from ascomycetes, such as *A. aculateus* with amino acid sequence identities >53%. Neither *egp28A* nor *rhg28C*, have obvious binding domains.

3.10. Glycosyl hydrolase family 74

Searches of the *P. chrysosporium* genome had shown only two members of the GH family 74 (Martinez et al., 2004), and peptides matching both genes were present in the cellulose-grown cultures (Table 3). Members of this family are widely distributed in bacterial and fungal systems, and all are believed to be inverting with respect to anomeric configuration of products. The models, pc.129.11.1 and pc.19.88.1, appear to have complete N-termini and secretion signals were detected. In the absence of known substrate specificities, the genes were designated *gly74A* and *gly74B* and the deduced proteins sequences were found to be 63% identical.

3.11. Glycosyl hydrolase family 16

Family 16 is large and functionally heterogeneous and the *P. chrysosporium* genome has at least 20 members. Peptides matched a single gene model, pc.22.125.1 (Table 3). Designated *gly16A*, the gene is most similar (Smith–Waterman score 516) to a mixed-linked glucanase from *Cochiobolus carbonum* (Gorlach et al., 1998). The deduced protein sequence of *gly16A* is 75% identical to *P. chrysosporium* model pc.78.31.1. Abbas et al. (2004) detected a peptide matching model pc.78.31.1, and Samejima and coworkers have deposited the corresponding cDNA sequence (BAC67687).

3.12. Other carbohydrate-active enzyme families

Possibly of importance to hemicellulose degradation, proteins closely related to known acetylxylan esterases (CE family 1, *axe1A*, pc.100.54.1) and α-L-arabinofuranosidase (GH family 51, *arb51A*, pc.50.85.1) were detected. The simultaneous occurrence of these enzymes may be significant because of their potential to act in a synergistic fashion with endoxylanases (Castanares et al., 1995).

Finally, we detected a single α-galactosidase protein (model pc.11.202.1), and this corresponds to the enzyme and gene previously characterized (Brumer et al., 1999; Hart et al., 2000). The peptide Spectum Mill score for this α-galactosidase was relatively low, but expression was supported by the presence of its cDNA (Fig. 2).

3.13. Oxidative enzymes

Peptides corresponding to a putative glucose oxidase gene (model gx.91.6.1) were identified. The

protein prediction of this model is very similar to *A. niger* glucose oxidase (Smith–Waterman score of 480). The enzyme oxidizes β-D-glucose at C-1 producing D-gluconolactone and H_2O_2 with O_2 as the electron acceptor. Therefore, this glucose oxidase is more precisely designated as glucose 1-oxidase, distinct from the glucose 2-oxidase or pyranose 2-oxidase (POX). The POX cDNA was recently sequenced (de Koker et al., 2004). No other extracellular oxidases were detected in cellulose-grown cultures, although cellobiose dehydrogenase was found, as expected.

3.14. Other enzymes

We also observed multiple peptide matches to a sequence with strong similarity to aldose-1-epimerase (gx.10.71.1; Table 4). Also known as mutarotase, this enzyme enhances glucose 1-oxidase activity by catalyzing the inter-conversion of α-D-glucose and β-D-glucose. Mutarotase is generally considered intracellular and so its identification in *P. chrysosporium* culture filtrates was surprising. However, careful inspection of the gene model revealed multiple introns in the 5′-terminus and an archetypal eukaryotic secretion signal was identified with high confidence using both SignalP-NN and SignalP-HMM. Further, we have determined that "hypothetical proteins" in *U. maydis* (UM01772) and *N. crassa* (XP329561) have similar structural features and secretion signals. An aldose-1-epimerase would also have a critical role in cellulose depolymerization involving exo- and endo-glucanases with inverting mechanisms (see Section 4).

Several proteases were identified in cellulose-grown cultures. Two aspartyl proteases (*asp1*, pc.12.138.1; *asp2*, pc.127.30.1) and a low scoring serine protease (*scp1*, gx.47.9.1, Fig. 2) were not previously known (Table 4). In contrast, model pc.111.22.1 (Table 2) encodes a protease isolated from colonized wood shavings and first identified by the match to the *N*-terminal amino acid sequence published by Datta (1992). By sequencing cDNAs and various manual corrections, we have determined that the protease is synthesized with a 184-residue propeptide and 17-residue secretion signal (unpublished results). The deduced protein is most similar to *A. oryzae* aorsin (Lee et al., 2003), a member of the serine–carboxyl family and insensitive to pepstatin.

Based on structural similarities with known proteins, additional expressed sequences include glutaminase, RNase, two thaumitin-like proteins, acid phosphatase, and an *O*-linked UDP-*N*-acetylglucosaminyltransferase (Table 4). The latter two peptides gave relatively low Spectrum Mill scores, and their transcripts were not detected in 6-day-old cellulose-grown cultures (Fig. 2). However, cDNAs were amplified from 3-day-old cellulose-grown cultures, suggesting temporally regulated expression and relatively stable proteins.

4. Discussion

Diverse extracellular proteins were observed in basal salts with cellulose as the sole carbon source. The current database of 11,722 sequences contains many incomplete and inaccurate gene models, but it allowed the identification of 178 unique peptides representing 48 genes. We believe all these identifications to be reliable because conservative scoring thresholds were imposed and expression was supported by cDNA amplification of the lowest scoring peptides. Only two additional genes, pc.69.19.1 (*pho1*) and pc.168.16.1 (hypotheticasl protein) were identified using a larger set of ORFs. Nevertheless, the ORF database is recommended for complete analysis and in two cases, pc.69.19.1 and pc.50.85.1 (*arb51A*), ORF 'hits' enabled extension of current models. In terms of optimizing the number of identifications at lowest cost, trypsin digests of concentrated filtrate and AspN digestion of CBIND eluates were together able to provide peptide sequence for all but six genes.

In contrast to the gene model and ORF databases, the computationally derived 'secretome' of 268 models is less useful. Of the 50 models assigned to specific peptides, only 25 were predicted. Given the inaccuracies inherent in fungal gene models, particularly incomplete N-termini, it seems unlikely that computational approaches will be useful for generating reliable and complete databases suitable for peptide searches. However, the list of gene models predicted to encode secreted, soluble proteins includes many interesting sequences which merit future study. In addition to the well-known lignin peroxidases and manganese peroxidases, a minimum of 30 glycosyl hydrolases and 21 proteases were identified. Among the latter group were 10 of the12 glutamic proteases, recently analyzed by Sims et al. (Sims et al., 2004). Sequences potentially involved in

degradation of lignin and related aromatic compounds include those similar to various oxidoreductases, oxalate decarboxylases, and hydoxylases. Perhaps more significant, at least 37 models showed some similarity (Smith–Waterman scores >50 using BLOSUM 62 matrix) only to hypothetical proteins in public databases, and 23 of these were very similar (score >150) to conserved proteins of unknown function.

While an impressive range of soluble proteins was identified in culture filtrates, the results pale in comparison to the number of gene models, which on the basis of structure, might be predicted to be important in degrading cellulose. For example, the genome harbors a minimum of 166 glycoside hydrolases in comparison to the 32 detected in out studies.

Families of structurally related genes are well known in *P. chrysosporium*, and the 32 identified glycoside hydrolases were unevenly distributed among 19 GH families. No peptides were identified for GH families 31, 47, or 92, which are predicted to contain 6, 5, and 6 members, respectively. Families 3, 16 and 61 are predicted to have a minimum of 12, 20 and 17 members, respectively, but only one representative protein was found for each of these families. In contrast, both predicted GH12s and GH74s, most GH7s (four out of six) and most GH10 (four out of six) were identified in the culture filtrate. The biological role of gene multiplicity remains uncertain. Substrate-based differential regulation among closely related genes is well established among the cellobiohydrolases and peroxidases, and it seems likely that many of the glycosyl hydrolases are expressed under cultural conditions other than that evaluated here.

Expression of genes involved in extracellular oxidations appeared to be limited to CDH and GOX in the cellulose-containing medium. Glucose oxidase activity was previously reported in *P. chrysosporium* cultures and the protein purified (Kelley and Reddy, 1986,1988). However, efforts to repeat those results suggest that the oxidase was most likely a POX and not the glucose 1-oxidase (Volc et al., 1996). Therefore, the indication of a putative GOX in cellulose-grown culture is particularly intriguing considering the past confusion regarding these oxidases. The cDNA encoding *P. chrysosporium* POX was recently sequenced and Northern blots showed high transcript levels in C-starved cultures relative to cellulose-grown cultures (de Koker et al., 2004). These results, together with our inability to detect POX peptides herein, suggest that POX is not required for cellulose degradation.

The importance of an aldose 1-epimerase in metabolism is related to the anomeric specificity of the enzymes that further transform the substrate. For example, only the β-anomer of cellobiose is a substrate for CDH of *P. chrysosporium* (Higham et al., 1994), and therefore a cellobiose 1-epimerase would greatly accelerate cellobiose oxidation by CDH. Similarly, glucose 1-oxidase is reported to only oxidize the β-anomer of glucose, while pyranose oxidase will oxidize both α- and β-anomers (Taguchi et al., 1985). Therefore, a 1-epimerase is critical for the rapid and complete oxidation of glucose by glucose 1-oxidase. The anomeric configuration of an aldose can also be important with hydrolases, even when the anomeric carbon is not the site of hydrolysis; the β-glucosidase from *Aspergillus niger* preferentially hydrolyzes the β-anomer of cellobiose (Seidle et al., 2004). Consequently, a role for an aldose 1-epimerase in cellulose metabolism is probable especially considering the presence of inverting-type cellulases reported here (endoglucanases pc.129.11.1 and pc.19.88.1 of GH74 family and cellobiohydrolase pc.3.82.1 of glycoside hydrolase family 6). Enzymes that would conceivably benefit from an epimerase are GOX (gx.91.6.1), CDH (pc.5.157.1) and β-glycosidase (gx.66.3.1). Whether the epimerase ALE1 (gx.10.71.1) has activity with glucose and cellobiose is not known but importance in extracellular metabolism is supported by its presence in culture filtrates and by the prediction of secretion signals in *P. chrysosporium* and in other cellulolytic fungi.

In summary, mass spec analysis holds considerable promise for solving fundamental questions related to lignocellulose degradation. Hundreds of genes are predicted to encode extracellular enzymes and LC–MS/MS provides a direct approach for assessing expression. However, mass spectral analysis can be costly and detection limited by technical issues (glycosylation, resistance to proteolytic cleavage, etc.). Database interpretation should recognize that current gene models are computer generated and continually subject to revisions (e.g. additions, deletions, corrections). Twenty-five proteins detected in our study correspond to gene models without secretion signals as predicted by SignalP. We presume that many, if not all, of these models are incorrect at their N-termini, but manual inspection revealed only seven additional

secretion signals. New and improved modeling approaches are being implemented (Igor Grigoriev, Joint Genome Institute, personal communication), and computational analysis of future databases is likely to better predict secreted proteins.

Acknowledgements

We are much indebted to Bernard Henrissat (CNRS Marseille, France) for clarifying gene nomenclature and for thoughtful comments regarding binding modules. This research was supported by U.S. Department of Energy grant DE-FG02-87ER13712, by Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture, under Agreement No. 2001-35103-11246, and by The Swedish Foundation for International Cooperation in Research and Higher Education (STINT), Sweden. Diego Martinez was supported by U.S. Department of Energy Contract W-7405-ENG-36a.

References

Abbas, A., Koc, H., Liu, F., Tien, M., 2004. Fungal degradation of wood: initial proteomic analysis of extracellular proteins of *Phanerochaete chrysosporium* grown on oak substrate. Curr. Genet. 47, 49–56.

Akhtar, M, 1997. Method for enhancing biopulping efficacy. U.S. Patent. Wisconsin Alumni Research Foundation.

Birney, E., Durbin, R., 2000. Using GeneWise in the *Drosophila* annotation experiment. Genome Res. 10, 547–548.

Brumer 3rd, H., Sims, P.F., Sinnott, M.L., 1999. Lignocellulose degradation by *Phanerochaete chrysosporium*: purification and characterization of the main alpha-galactosidase. Biochem. J. 339 (Part 1), 43–53.

Cameron, M.D., Aust, S.D., 2001. Cellobiose dehydrogenase—an extracellular fungal flavocytochrome. Enzyme Microb. Technol. 28, 129–138.

Castanares, A., Hay, A.J., Gordon, A.H., McCrae, S.I., Wood, T.M., 1995. D-Xylan-degrading enzyme system from the fungus *Phanerochaete chrysosporium*: isolation and partial characterisation of an alpha-(4-O-methyl)-D-glucuronidase. J. Biotechnol. 43, 183–194.

Copa-Patino, J., Kim, Y.G., Broda, P., 1993. Production and initial characterisation of the xylan-degrading enzyme system of *Phanerochaete chrysosporium*. Appl. Microbiol. Biotechnol. 40, 69–76.

Covert, S., Vanden Wymelenberg, A., Cullen, D., 1992b. Structure, organization and transcription of a cellobiohydrolase gene cluster from *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. 58, 2168–2175.

Cullen, D., Kersten, P.J., 2004. Enzymology and Molecular Biology of Lignin Degradation. In: Brambl, R., Marzulf, G.A. (Eds.), The Mycota. III. Biochemistry and Molecular Biology. Springer-Verlag, Berlin, pp. 249–273.

Datta, A., 1992. Purification and characterization of a novel protease from solid substrate cultures of *Phanerochaete chrysosporium*. J. Biol. Chem. 267, 728–732.

de Boer, H.A., Zhang, Y.Z., Collins, C., Reddy, C.A., 1987. Analysis of nucleotide sequences of two ligninase cDNAs from a white-rot filamentous fungus, *Phanerochaete chrysosporium*. Gene 60, 93–102.

Decelle, B., Tsang, A., Storm, R., 2004. Cloning, functional expression and characterization of three *Phanerochaete chrysosporium* endo-1,4-β-xylanases. Curr. Genet. 46, 166–175.

de Koker, T.H., Mozuch, M.D., Cullen, D., Gaskell, J., Kersten, P.J., 2004. Pyranose 2-oxidase from *Phanerochaete chrysosporium*: isolation from solid substrate, protein purification, and characterization of gene structure and regulation. Appl. Environ. Microbiol. 70, 5794–5800.

Deshpande, V., Eriksson, K-E., Pettersson, B., 1978. Production, purification and partial characterization of 1,4-β-glucosidase enzymes from *Sporotrichum pulverulentum*. Eur. J. Biochem. 90, 191–198.

Eriksson, K.-E., Pettersson, B., Westermark, U., 1974. Oxidation: an important enzyme reaction in fungal degradation of cellulose. FEBS Lett. 49, 282–285.

Eriksson, K.-E., Pettersson, B., 1975a. Extracellular enzyme system utilized by the fungus *Sporotrichum pulverulentum* (*Chrysosporium lignorum*) for the breakdown of cellulose. 1. Separation, purification, and physio-chemical characterization of five endo-1,4-β-glucanases. Eur. J. Biochem. 51, 193–206.

Eriksson, K.-E., Pettersson, B., 1975b. Extracellular enzyme system used by the fungus *Sporotrichum pulverulentum* (*Chrysosporium lignorum*) for the breakdown of cellulose. 3. Purification, physiochemical characterization of an exo-1,4-β-glucanases. Eur. J. Biochem. 51, 213–218.

Eriksson, K.-E., Hamp, S.G., 1978. Regulation of endo-1,4-β-glucanase production in *Sporotrichum pulverulentum*. Eur. J. Biochem. 90, 183–190.

Eriksson, K.-E.L., Blanchette, R.A., Ander, P., 1990. Microbial and Enzymatic Degradation of Wood and Wood Components. Springer-Verlag, Berlin.

Gorlach, J.M., Van Der Knaap, E., Walton, J.D., 1998. Cloning and targeted disruption of *MLG1*, a gene encoding two of three extracellular mixed-linked glucanases of *Cochliobolus carbonum*. Appl. Environ. Microbiol. 64, 385–391.

Hart, D.O., He, S., Chany 2nd, C.J., Withers, S.G., Sims, P.F., Sinnott, M.L., Brumer 3rd, H., 2000. Identification of Asp-130 as the catalytic nucleophile in the main alpha-galactosidase from *Phanerochaete chrysosporium*, a family 27 glycosyl hydrolase. Biochemistry 39, 9826–9836.

Henriksson, G., Nutt, A., Henriksson, H., Pettersson, B., Stahlberg, J., Johansson, G., Pettersson, G., 1999. Endoglucanase 28 (Cel12A), a new *Phanerochaete chrysosporium* cellulase. Eur. J. Biochem. 259, 88–95.

Henriksson, G., Johansson, G., Pettersson, G., 2000a. A critical review of cellobiose dehydrogenases. J. Biotechnol. 78, 93–113.

Henriksson, G., Zhang, L., Li, J., Ljungquist, P., Reitberger, T., Pettersson, G., Johansson, G., 2000b. Is cellobiose dehydrogenase from *Phanerochaete chrysosporium* a lignin degrading enzyme? Biochim. Biophys. Acta 1480, 83–91.

Henrissat, B., 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 280 (Part 2), 309–316.

Higham, C.W., Gordon-Smith, D., Dempsey, C.E., Wood, P.M., 1994. Direct 1H NMR evidence for conversion of beta-D-cellobiose to cellobionolactone by cellobiose dehydrogenase from *Phanerochaete chrysosporium*. FEBS Lett. 351, 128–132.

Igarashi, K., Tani, T., Kawai, R., Samejima, M., 2003. Family 3 β-glucosidase from cellulose-degrading cultures of the white-rot fungus *Phanerochaete chrysosporium* is a glucan 1,3-β-glucosidase. J. Biosci. Bioeng. 95, 572–576.

Janse, B.J.H., Gaskell, J., Akhtar, M., Cullen, D., 1998. Expression of *Phanerochaete chrysosporium* genes encoding lignin peroxidases, manganese peroxidases, and glyoxal oxidase in wood. Appl. Environ. Microbiol. 64, 3536–3538.

Kelley, R.L., Reddy, C.A., 1986. Purification and characterization of glucose oxidase from ligninolytic cultures of *Phanerochaete chrysosporium*. J. Bacteriol. 166, 269–274.

Kelley, R.L., Reddy, C.A., 1988. Glucose oxidase of *Phanerochaete chrysosporium*. Methods Enzymol. 161, 307–316.

Kersten, P., Cullen, D., 1993. Cloning and characterization of a cDNA encoding glyoxal oxidase, a peroxide-producing enzyme from the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. Proc. Nat. Acad. Sci. U.S.A. 90, 7411–7413.

Lee, B.R., Furukawa, M., Yamashita, K., Kanasugi, Y., Kawabata, C., Hirano, K., Ando, K., Ichishima, E., 2003. Aorsin, a novel serine proteinase with trypsin-like specificity at acidic pH. Biochem. J. 371, 541–548.

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

Mansfield, S.D., deJong, E., Saddler, J.N., 1997. Cellobiose dehydrogenase, an active agent in cellulose depolymerization. Appl. Environ. Microbiol. 63, 3804–3809.

Martinez, D., Larrondo, L.F., Putnam, N., Sollewijn Gelpke, M.D., Huang, K., Chapman, J., Helfenbein, K.G., Ramaiya, P., Detter, J.C., Larimer, F., Coutinho, P.M., Henrissat, B., Berka, R., Cullen, D., Rokhsar, D., 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. Nat. Biotechnol. 22, 695–700.

Munoz, I.G., Ubhayasekera, W., Henriksson, H., Szabo, I., Pettersson, G., Johansson, G., Mowbray, S.L., Stahlberg, J., 2001. Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. J. Mol. Biol. 314, 1097–1111.

Raices, M., Paifer, E., Cremata, J., Montesino, R., Stahlberg, J., Divne, C., Szabo, I.J., Henriksson, G., Johansson, G., Pettersson, G., 1995. Cloning and characterization of a cDNA encoding a cellobiose dehydrogenase from the white rot fungus *Phanerochaete chrysosporium*. FEBS Lett. 369, 233–238.

Seidle, H.F., Marten, I., Shoseyov, O., Huber, R.E., 2004. Physical and kinetic properties of the family 3 beta-glucosidase from *Aspergillus niger* which is important for cellulose breakdown. Protein J. 23, 11–23.

Sims, A.H., Dunn-Coleman, N.S., Robson, G.D., Oliver, S.G., 2004. Glutamic protease distribution is limited to filamentous fungi. FEMS Microbiol. Lett. 239, 95–101.

Sims, P., James, C., Broda, P., 1988. The identification, molecular cloning and characterisation of a gene from *Phanerochaete chrysosporium* that shows strong homology to the exo-cellobiohydrolase I gene from *Trichoderma reesei*. Gene 74, 411–422.

Smith, M.H., Gold, M.H., 1979. *Phanerochaete chrysosporium* β-glucosidase: induction, cellular localization, and physical characterization. Appl. Environ. Microbiol. 37, 938–942.

Smith, T.F., Waterman, M.S., 1981. Identification of common molecular subsequences. J. Mol. Biol. 147, 195–197.

Taguchi, T.K., Ohwaki, K., Okuda, J., 1985. Glucose 2-oxidase (*Coriolus versicolor*) and its application to D-glucose colorimetry. J. Appl. Biochem. 7, 289–295.

Tempelaars, C., Birch, P., Sims, P., Broda, P., 1994. Isolation, characterization, and analysis of the *cbhII* gene of *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. 60, 4387–4393.

Tuisel, H., Sinclair, R., Bumpus, J.A., Ashbaugh, W., Brock, B.J., Aust, S.D., 1990. Lignin peroxidase H2 from *Phanerochaete chrysosporium*: purification, characterization and stability to temperature and pH. Arch. Biochem. Biophys. 279, 158–166.

Uzcategui, E., Johansson, G., Ek, B., Pettersson, G., 1991a. The 1,4-D-glucan glucanohydrolases from *Phanerochaete chrysosporium*. Re-assessment of their significance in cellulose degradation mechanisms. J. Biotechnol. 21, 143–160.

Uzcategui, E., Raices, M., Montesino, R., Johansson, G., Pettersson, G., Eriksson, K.-E., 1991b. Pilot-scale production and purification of the cellulolytic enzyme system from the white-rot fungus *Phanerochaete chrysosporium*. Biotechnol. Appl. Biochem. 13, 323–334.

Uzcategui, E., Ruiz, A., Montesino, R., Johansson, G., Pettersson, G., 1991c. The 1,4-β-D-glucan cellobiohydrolase from *Phanerochaete chrysosporium*. I. A system of synergistically acting enzymes homologous to *Trichoderma reesei*. J. Biotechnol. 19, 271–286.

Vallim, M.A., Janse, B.J., Gaskell, J., Pizzirani-Kleiner, A.A., Cullen, D., 1998. *Phanerochaete chrysosporium* cellobiohydrolase and cellobiose dehydrogenase transcripts in wood. Appl. Environ. Microbiol. 64, 1924–1928.

Vanden Wymelenberg, A., Covert, S., Cullen, D., 1993. Identification of the gene encoding the major cellobiohydrolase of the white-rot fungus *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. 59, 3492–3494.

Vanden Wymelenberg, A., Denman, S., Dietrich, D., Bassett, J., Yu, X., Atalla, R., Predki, P., Rudsander, U., Teeri, T.T., Cullen, D., 2002. Transcript analysis of genes encoding a family 61 endoglucanase and a putative membrane-anchored family 9 glycosyl

hydrolase from *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. 68, 5765–5768.

Volc, J., Kubatova, E., Daniel, G., Prikrylova, V., 1996. Only C-2 specific glucose oxidase activity is expressed in ligninolytic cultures of the white rot fungus *Phanerochaete chrysosporium*. Arch. Microbiol. 165, 421–424.

Williams, H.L., Tang, Y., Hintz, W.E., 2002. Endopolygalacturonase is encoded by a multigene family in the basidiomycete *Chondrostereum purpureum*. Fungal Genet. Biol. 36, 71–83.

Xu, Y., Uberbacher, E.C., 1997. Automated gene identification in large-scale genomic sequences. J. Comput. Biol. 4, 325–338.