ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Glycoside Hydrolase Inventory Drives Plant Polysaccharide Deconstruction by the Extremely Thermophilic Bacterium *Caldicellulosiruptor saccharolyticus*

Amy L. VanFossen, Inci Ozdemir, Samantha L. Zelin, Robert M. Kelly

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695-7905; telephone: 919-515-6396; fax: 919-515-3465; e-mail: rmkelly@eos.ncsu.edu

Received 15 November 2010; revision received 22 January 2011; accepted 1 February 2011
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bit.23093

ABSTRACT: The genome of *Caldicellulosiruptor saccharolyticus* encodes a range of glycoside hydrolases (GHs) that mediate plant biomass deconstruction by this bacterium. Two GH-based genomic loci that appear to be central to the hydrolysis of hemicellulosic and cellulosic substrates were examined. XynB-XynF (Csac_2404-Csac_2411) encodes intracellular and extracellular GHs that are active towards xylan and xylan side-chains, as well as carboxymethyl cellulose (CMC). XynD (Csac_2409) and XynE (Csac_2410) were produced recombinantly and confirmed to be xylanases. XynF (Csac_2411) was produced in two separate polypeptides, each with one GH43 catalytic domain displaying α-L-arabinofuranosidase activity. CelA-ManB (Csac_1076-Csac_1080) encodes four multi-domain, extracellular GHs, including CelB (Csac_1078), a 118 kDa extracellular enzyme not present in the other genome-sequenced member of this genus, *Caldicellulosiruptor bescii* (formerly *Anaerocellum thermophilum*). CelB contains both GH10 and GH5 domains, separated by a family 3 carbohydrate-binding module (CBM3). CelB encoded in Csac_1078 differed from the version originally reported (Saul et al., 1990, Appl Environ Microbiol 56:3117–3124) with respect to linker regions. CelB hydrolyzed xylan and CMC, as well as barley β-glucan, glucomannan, and arabinoxylan. For all substrates tested, intact CelB was significantly more active than either the individual GH5 and GH10 domains or the two discrete domains together, indicating that the multi-domain architecture is essential for complex carbohydrate hydrolysis. Transcriptomes for *C. saccharolyticus* grown at 70°C on glucose, xylose, xyloglucan, switchgrass, and poplar revealed that certain GHs were particularly responsive to growth on switchgrass and poplar and that CelB was in the top decile of all transcripts during growth on the plant biomass.
Biotechnol. Bioeng. 2011;9999: xxx–xxx.

KEYWORDS: *Caldicellulosiruptor saccharolyticus*; extreme thermophile; plant biomass; glycoside hydrolases

Introduction

The use of lignocellulosic biomass, in the form of agricultural residues (e.g., wheat straw, sugarcane bagasse, corn stover), forest products (hardwoods, softwoods), and dedicated crops (switchgrass, salix), as substrate for biofuels (Kumar et al., 2009), is predicated on the chemical/biological conversion of the carbohydrate component to fermentable sugars (Lynd et al., 2008). Of these biomass substrates, focus to date has been on switchgrass and poplar for several reasons. Switchgrass, a perennial herbaceous crop, grows to high yields from seed, and adapts well to poor soils and marginal cropland (Bouton, 2007). Poplar grows rapidly at high density over a wide range of temperatures as a short-rotation woody crop, and has a fully sequenced genome (Simmons et al., 2008). Cellulose and hemicellulose comprise ~55% of switchgrass and ~70% of poplar (Khandeparker and Numan, 2008). For the conversion of biomass to fuel, the cellulose and hemicellulose must be broken down into simple sugars to facilitate fermentation.

Consolidated bioprocessing microorganisms, capable of deconstructing untreated or minimally treated plant biomasses and subsequently producing biofuels, such as liquid alcohols and H_2 (Lynd et al., 2005), are highly desirable (Lynd et al., 2002; Yang et al., 2009a; Zaldivar et al., 2001). *Caldicellulosiruptor saccharolyticus*, a candidate for consolidated bioprocessing, is a Gram-positive, anaerobic

Additional Supporting Information may be found in the online version of this article.
Amy L. VanFossen's present address is Agrivida, Inc., Medford, MA.
Samantha L. Zelin's present address is Department of Microbiology, University of Massachusetts, Amherst, MA.
Correspondence to: R.M. Kelly
Contract grant sponsor: US NSF
Contract grant number: CBT0617272
Contract grant sponsor: Bioenergy Science Center (BESC)
Contract grant sponsor: Office of Biological and Environmental Research
Contract grant sponsor: DOE Office of Science

bacterium that grows optimally at 70°C on a wide variety of biomass-related complex carbohydrates, including crystalline cellulose, hemicellulose, starch, pectin, as well as simple sugars (Blumer-Schuette et al., 2010; Rainey et al., 1994).

Previous studies indicated that both hexoses and pentoses are co-fermented by this microorganism to produce roughly 4 moles H_2 per mole of glucose, an unusual, but attractive, trait for bioenergy production (de Vrije et al., 2007; Kadar et al., 2003; van de Werken et al., 2008; VanFossen et al., 2009). The genome of *C. saccharolyticus* encodes a wide variety of glycoside hydrolases (GHs; Table I) that hydrolyze complex carbohydrate growth substrates to facilitate their conversion to hydrogen (VanFossen et al., 2008). A peculiar aspect of certain β-specific GH encoded in the genomes of *C. saccharolyticus* and other *Caldicellulosiruptor* species is their multi-domain organization (Fig. 1). The structural architecture of these complex proteins involves one or two catalytic GH domains combined with one or more carbohydrate-binding motifs (CBMs). The CBM increases the rate of catalysis by bringing the catalytic domain into close and prolonged association with its substrate (Shoseyov et al., 2006). Having multiple but distinct catalytic capacities creates the possibility of novel enzymatic properties, including mobility of a substrate intermediate between

Table I. Glycoside hydrolases encoded in the *C. saccharolyticus* genome.

Open reading frame #	Glycoside hydrolase	GH family	MW[a] (kDa)	Signal P[b]	Mode of action	Refs.
β-Glucan acting						
Csac_1076	1,4-β-Glucanase (CelA)	GH48 & 9	193.7	Yes	Endo	Te'o et al. (1995)
Csac_1077[c]	1,4-β-Glucanase (ManA)	GH5 & 44	146.9	Yes	Endo	Gibbs et al. (1992), Luthi et al. (1991)
Csac_1078[d]	1,4-β-Glucanase (CelB)	GH10 & GH5	117.6	Yes	Endo	Saul et al. (1990), This work
Csac_1079	1,4-β-Glucanase (CelC)	GH9	123.8	Yes	Endo	Morris et al. (1995)
Csac_1085	Endoglucanase	GH74	101.7	Yes	Endo	—
Csac_1089	β-Glucosidase (BglA)	GH1	53.2	No	Exo	Hardiman et al. (2010), Hong et al. (2009a), Love et al. (1988)
Csac_2548	Endoglucanase	GH16	125.9	No	Endo	—
β-Xylan acting						
Csac_0204	Endo-1,4-β-xylanase	GH10	46.9	No	Endo	—
Csac_0678	Endo-1,4-β-xylanase	GH5	84.7	Yes	Endo	—
Csac_0696	Endo-1,4-β-xylanase	GH10	78.8	Yes	Endo	—
Csac_2404	β-Xylosidase (XynC)	GH39	56.4	No	Exo	Luthi et al. (1990b)
Csac_2405	Endo-1,4-β-xylanase (ORF4)	GH10	36.5	No	Endo	Luthi et al. (1990b)
Csac_2408	Endo-1,4-β-xylanase (XynA)	GH10	40.5	No	Endo	Luthi et al. (1990a)
Csac_2409	Endo-1,4-β-xylanase (XynD)	GH39	58.5	No	Endo	Bergquist et al. (1999), This work
Csac_2410	Endo-1,4-β-xylanase (XynE)	GH10	80.9	Yes	Endo	Bergquist et al. (1999), This work
Csac_2732	β-Xylosidase	GH39	35.7	No	Exo	—
β-Mannan acting						
Csac_0129	β-Mannosidase	GH2	95.7	No	Exo	—
Csac_0137	β-1,4-Mannanase	GH5	39.0	No	Endo	—
Csac_0663	β-1,4-Mannanase	GH26	52.9	No	Endo	—
Csac_1080	β-1,4-Mannanase (ManB)	GH5	27.5	No	Endo	Morris et al. (1995)
Csac_2528	β-1,4-D-Mannanase	GH5	71.1	No	Endo	—
Side-chain degrading						
Csac_0359	β-Xylosidase/α-*N*-arabinofuranosidase	GH43	60.6	No	Exo	—
Csac_0362	β-Galactosidase	GH2	118.9	No	Exo	—
Csac_1018	β-Galactosidase	GH42	79.2	No	Exo	Park and Oh (2010)
Csac_1118	α-Galactosidase (clan GH-D)	GH36	84.5	No	Exo	—
Csac_1561	α-*N*-arabinofuranosidase	GH51	57.9	No	Exo	Lim et al. (2010)
Csac_2411	α-L-Arabinofuranosidase (XynF)	GH43 & 43	152.0	Yes	Exo	Bergquist et al. (1999), This work
Pectin acting						
Csac_0360	Unsaturated rhamnogalacturonyl hydrolase	GH105 or 88	43.6	No	Exo	—
Csac_0361	Galacturan 1,4-α-galacturonidase	GH28	50.0	No	Exo	—
Csac_0664	Galacturan 1,4-α-galacturonidase	GH28	50.8	No	Exo	—
Csac_1105	α-L-rhamnosidase N-terminal domain	GH78	58.0	No	Exo	—
Csac_1107	α-L-Rhamnosidase C-terminal domain	GH78	47.7	No	Exo	—
Csac_1560	Endo-1,5-α-L-arabinanase	GH43	55.4	No	Endo	Hong et al. (2009b)
Csac_2527	Endo-1,5-α-L-arabinanase	GH43	38.4	Yes	Endo	—
Csac_2689	α-Glucuronidase	GH67	80.6	No	Exo	—
Csac_2730	Unsaturated rhamnogalacturonyl hydrolase	GH105 or 88	44.6	No	Exo	—

[a]MW, molecular weight in kDa.
[b]Signal peptide.
[c]Csac_1077 also acts as mannanase.
[d]Csac_1078 also acts as xylanase.

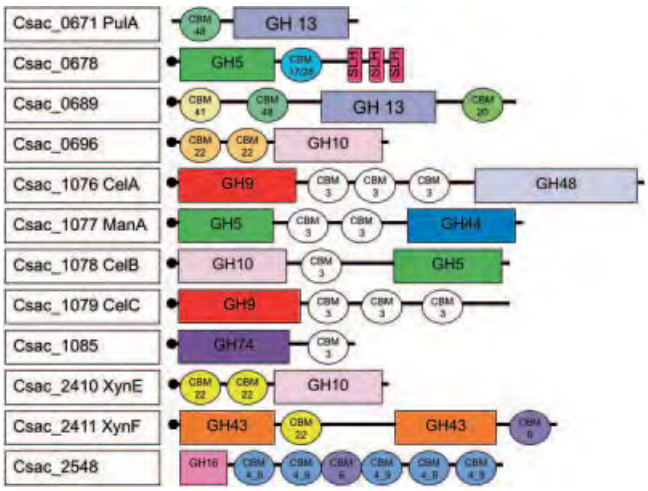

Figure 1. Multi-domain GHs found in *C. saccharolyticus*. These enzymes have domains belonging to different families and are associated with carbohydrate-binding motifs. The circles at the N-terminus indicate signal peptide sequences. GH, glycoside hydrolase family; CBM, carbohydrate-binding motif; SLH, S-layer homology domain.

two active sites and hydrolytic synergism (Khandeparker and Numan, 2008).

Here, two key genomic loci encoding multiple domain GHs were examined with respect to their role during *C. saccharolyticus* growth on plant biomass. Biochemical analysis of recombinant forms of selected GHs from these loci was pursued to provide insights into their contributions to consolidated bioprocessing. Furthermore, the transcriptome of *C. saccharolyticus* growing on poplar and switchgrass was examined to further elucidate the physiological and biochemical strategies used by this bacterium for the deconstruction of lignocellulosic biomass.

Materials and Methods

Cloning, Expression and Purification of Specific *C. saccharolyticus* GHs

Genes encoding selected GHs were amplified from *C. saccharolyticus* genomic DNA by PCR, using primers listed in Table S1. The primers were designed so that any nucleotide sequences encoding putative signal peptides were not incorporated into the PCR products. The amplified fragments were inserted into the pET28b (Novagen, Madison, WI) for expression with N-terminal 6X-His tags, using the noted restriction endonucleases (Table S1). Protein production was done in Rosetta (DE3; Novagen) at 37°C, induced when OD_{600} reached ~0.8, using either 1 mM isopropyl-1-thio-β-D-galactopyranoside (XynE, XynF-N_term, XynF-C_term, CelB, CelB-GH10, and CelB-GH5) or 40 mM α-lactose (XynD); cells were harvested 4 h after induction. The cell pellet was re-suspended in 30 mL of 50 mM sodium phosphate pH 8.0, containing 300 mM NaCl. Cells were lysed by sonication, and heat-treated at 60°C for 20–45 min to precipitate the bulk of *Escherichia coli* proteins. Proteins were further purified using a HiTrap HP column (GE Life Sciences, Piscataway, NJ). For full-length CelB (lacking its signal peptide), a Superdex-75 (GE Life Sciences) gel filtration column was used (XK 16 column; bed volume, 190 mL) followed the HiTrap. Protein concentrations were determined using the Biorad protein assay (Biorad, Hercules, CA) with bovine serum albumin as the standard.

Biochemical Characterization of GHs

All samples and substrates were incubated separately at the temperature of interest for at least 2 min before mixing. GH activity was determined in triplicate by measuring the amounts of reducing sugars released from the substrates, using the dinitrosalicyclic acid (DNS) reagent method (Konig et al., 2002).

Xylanase GH Activity

Activity of enzymes in the xylanase cluster was determined at 70°C. Activity toward substrates birchwood xylan (Sigma–Aldrich, St. Louis, MO), wheat arabinoxylan (Megazyme, Wicklow, Ireland), and rye arabinoxylan (Megazyme) was determined with 0.5% (w/v) substrate. For assays using *ortho*-nitrophenyl β-D-xylopyranoside (oNPX; Sigma–Aldrich), *para*-nitrophenyl D-xylopyranoside (pNPX; Sigma–Aldrich) and *para*-nitrophenyl α-L-arabinofuranoside (pNPara; Sigma–Aldrich), 2 mM of the substrate in 50 mM sodium acetate buffer (pH 5.0) was used. Release of nitrophenyl was measured at 405 nm using a HTS 7000 Plus Bio Assay Reader (Perkin Elmer, Waltham, MA).

pH and Temperature Optima

To determine the optimum pH and temperature of CelB, the purified recombinant enzyme was incubated with 1% (w/v) birchwood xylan (Sigma–Aldrich) or with 0.5% carboxymethyl cellulose (CMC; Sigma–Aldrich) in 50 mM sodium acetate buffer (pH 4.0–6.0) or sodium phosphate buffer (pH 6.0–8.0) in a final reaction volume of 60 μL. After incubation for 30 min (xylans) or 60 min (glucans), at the respective reaction temperature, the activity was measured using the DNS reagent method.

CelB Substrate Specificity

Activity of CelB on barley β-glucan (Megazyme), CMC (Sigma–Aldrich), konjac glucomannan (Megazyme), birchwood xylan, oat spelt xylan (Sigma–Aldrich), wheat arabinoxylan, filter paper (Whatman No. 1, Kent, UK)

and Avicel PH-101 (FMC Corp., Philadelphia, PA) was determined at optimal temperatures and pH. Both glucanase and xylanase activity was measured using 0.5% (w/v) substrate, except for the filter paper assay, for which a disk of paper (8 mm diameter) was added. The reaction mixture of polysaccharide and enzyme was incubated for 30 min.

Enzyme Thermoinactivation

The recombinant enzyme CelB was incubated in 0.05 M sodium phosphate buffer (pH 6.0) for 48 h at selected temperatures. Residual activity on birchwood xylan and CMC was determined for the conditions described above, except the reaction mixture volume was 10 mL.

Effects of Various Metal Ions

To examine the effect of cations on CelB activity, the purified enzyme was treated with EDTA by first being dialyzed overnight at 4°C against 1 L of 50 mM MOPS (pH 7.0) containing 10 mM EDTA. It was then dialyzed twice against 50 mM MOPS (pH 7.0) containing 2 mM EDTA, and finally dialyzed twice again in 50 mM MOPS (pH 7.0). The apo-enzyme was mixed with appropriate metal salts at the final concentration of 1 mM. Xylanase activity was measured at pH 6.0 and 80°C, while CMCase activity was measured at pH 6.0 and a temperature of 95°C.

Growth of *C. saccharolyticus*

C. saccharolyticus DSM 8903 (also, ATCC 43494) was acclimated to the substrate of interest in modified DSMZ 640 medium. This was done by seven passes in serum bottles shaken at 100 rpm under anaerobic conditions (N_2 headspace) on the substrate of interest, before inoculating a 400 mL screw top, batch culture at 70°C, also shaken at 100 rpm. Growth was followed for 18 h on the monosaccharides tested and up to 36 h on poplar and switchgrass. The following growth substrates with the indicated sources were used: D-glucose and D-xylose, both from Sigma–Aldrich; poplar (sieved, −80 mesh fraction; kindly provided by Vincent L. Chiang, Department of Forestry, North Carolina State University), and switchgrass (sieved −20/+80 mesh fraction, dilute acid pretreatment, kindly provided by the National Renewable Energy Laboratory, Golden, CO). Growth medium was supplemented with 0.5 g monosaccharide, 1.0 g poplar, or 4.0 g switchgrass L^{-1}. Cultures were harvested during exponential phase growth by rapidly chilling and then centrifugation at 5,000 × *g*, before storage at −80°C.

Epifluorescence Microscopy

Cells were visualized by first fixing culture samples with glutaraldehyde (Sigma–Aldrich, 0.25% final concentration) for 5 min, and then staining small aliquots of cells (~10 μL) in 5 mL of 0.05% acridine orange (Sigma–Aldrich). The stained cells were filtered onto 0.22 μm polycarbonate filter membranes (GE Water and Process Technologies, Trevose, PA). An epifluorescence microscope (Nikon, Tokyo, Japan) with appropriate filter set (Omega Optical, Brattleboro, VT) that was fitted with a Spot system digital camera (Diagnostic Instruments Inc., Sterling Heights, MI) was used to capture images.

Oligonucleotide Microarray Construction and Transcriptional Response Analysis

A spotted whole-genome *C. saccharolyticus* microarray was constructed based the 2679 open reading frames (ORFs) in the genome (http://genome.ornl.gov/microbial/csac), as described previously (van de Werken et al., 2008), with each probe spotted five times. Total RNA was extracted using a protocol described previously, with an additional sonication step after the cells were re-suspended in TRIzol reagent (Invitrogen, San Diego, CA; van de Werken et al., 2008). Once isolated, the RNA was reverse transcribed (Superscript III, Invitrogen; random primers, Invitrogen; 5-(3-aminoallyl)-2′-deoxyuridine-5′-triphosphate, Ambion, Austin, TX), labeled with either Cy3 or Cy5 dye (GE Life Sciences), and then hybridized to Corning Ultra-Gap II slides (Corning, Acton, MA). Samples were hybridized according to a four-slide loop experimental design (Figure S1). Slides were scanned using a ScanArray Lite microarray scanner (Perkin Elmer); raw signal intensities were quantitated using ScanArray Express (v2.1.8; Perkin-Elmer). Normalization of data and statistical analysis were performed with JMP Genomics 3.0 (SAS, Cary, NC), as described previously (Pysz et al., 2004), using a mixed-effects analysis of variance model (Wolfinger et al., 2001). ORFs that were differentially transcribed twofold or more and met the Bonferroni statistical criterion were considered to be up- or down-regulated (Wolfinger et al., 2001).

Microarray Accession Numbers

The microarray platform used in this study is available in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GPL6681. The raw data along with the final $\log_2$ fold changes have been deposited in the same database under accession number GSE18856.

Results and Discussion

Role of GHs in Carbohydrate Utilization by *C. saccharolyticus*

Genes encoding GHs are widely distributed throughout the *C. saccharolyticus* genome (van de Werken et al., 2008; VanFossen et al., 2009). Table I lists these polysaccharide-

degrading enzymes that include 12 multi-domain GHs (Fig. 1), which attack β-glucan, β-xylan, or β-mannan linkages. The sequential degradation of hemicellulose and glucan substrates by these GHs presumably creates better access of cellulolytic enzymes to cellulose microfibrils that are embedded in plant biomass and facilitates *C. saccharolyticus* growth on plant biomass. Many of the endo-acting GHs in this bacterium are either membrane-associated or exported to the secretome, while some are resident in the cytoplasm to breakdown oligosaccharides transported into the cell. Several endo-acting GHs contain domains belonging to carbohydrate-active enzymes (CAZy; http://www.cazy.org) glycoside hydrolyase family 5 (GH5) (Csac_0137, Csac_0678, ManA, CelB, and ManB) and/or GH10 (Csac_0204, Csac_0696, CelB, ORF4, and XynA). Two loci found within the *C. saccharolyticus* genome appear to be particularly important for plant biomass degradation, based on the number and type of encoded endo-acting, extracellular GHs. These loci were examined in more detail.

xynB-xynF GH Locus

xynB-xynF (Csac_2404-Csac_2411; Fig. 2A), located on the negative strand of the genome, encodes most of the identifiable xylan-degrading enzymes found in *C. saccharolyticus*. Previous reports indicated that XynA (Csac_2408, GH10) exhibits both endoxylanase and β-xylosidase activity, while XynB (Csac_2404, GH39) has β-xylosidase activity only (Luthi et al., 1990a,b). XynC (Csac_2407) is not a GH but rather an acetyl esterase, based on its activity towards α-naphthyl acetate (Luthi et al., 1990b). Biochemical characteristics of the several other enzymes in this locus have yet to be reported: XynD

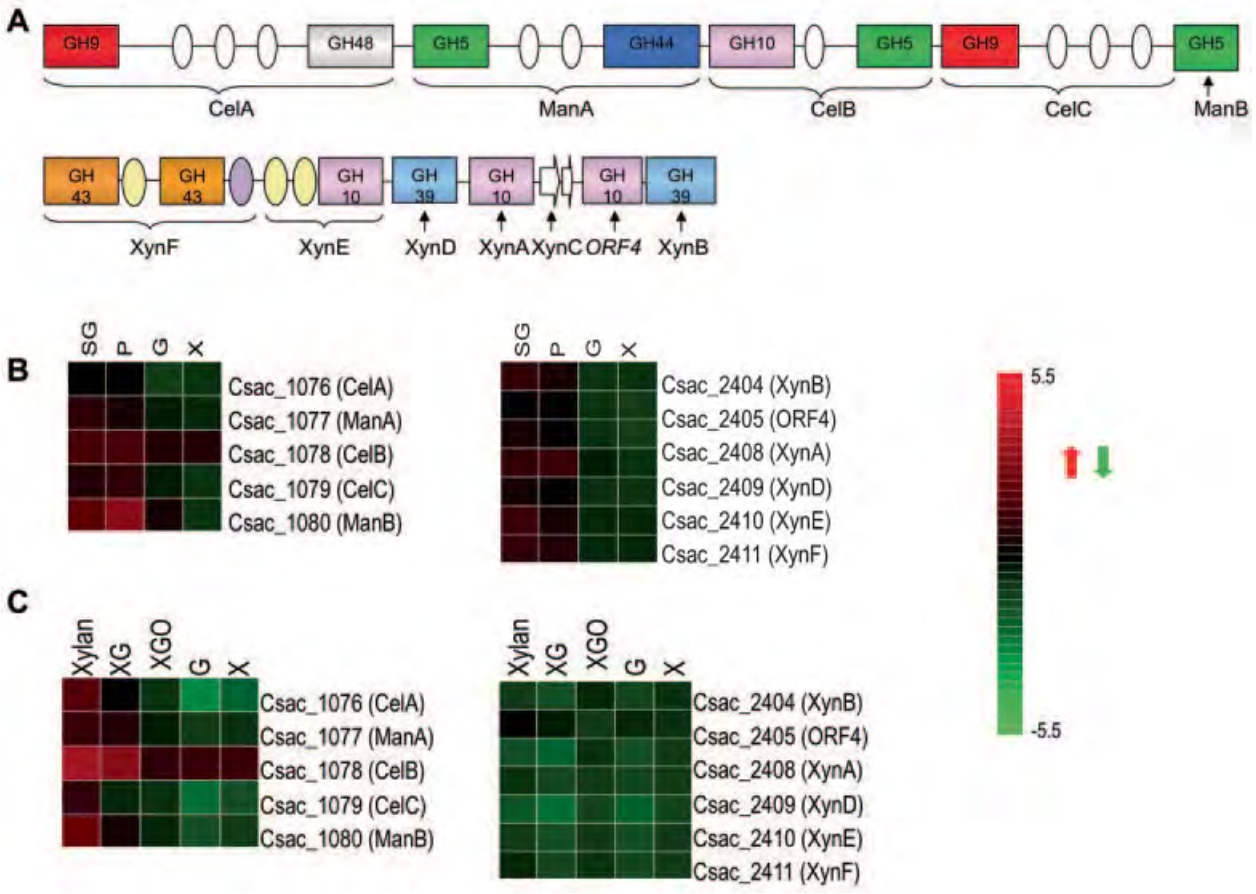

Figure 2. Response of biomass degrading enzyme genes located together in *C. saccharolyticus*. **A**: Cellulase–hemicellulase gene cluster *celA-manB* (Csac_1076–1080) and xylanase gene cluster *xynB-xynF* (Csac_2404–2411). Boxes represent GH domains and the circles are CBM domains: solid white, CBM3; horizontal strips, CBM22; vertical strips, CBM6. **B**: Differential transcription of genes encoding GHs involved in biomass degradation in biomass loop. **C**: Differential transcription of genes encoding GH and an acetyl esterase involved in carbohydrate degradation in the hemicellulose polysaccharide loop. SG, switchgrass; G, glucose; P, poplar; X, xylose; XG, xyloglucan; XGO, xylogluco-oligosaccharide. Least squares mean (lsm) estimates (Materials and Methods Section) of transcript level shown for selected genes (ORF#). White and black denote transcript levels above (black) and below (white) the mean across all genes.

(Csac_2409, GH39), XynE (Csac_2410), and XynF (Csac_2411), the latter two of which are both multi-domain, extracellular enzymes containing GH10 or GH43 family domains, respectively (Bergquist et al., 1999; Fig. 1). Recombinant XynD produced here was active on xylan and oNP-X, while XynE had biochemical features similar to XynA. Bergquist et al. (1999) proposed that XynF (Csac_2411) was the result of gene fusion: two domains comprising an arabinofuranosidase (GH43 and CBM22) and two domains comprising a xylanase (GH43 and CBM6). Here, cloning and expression of intact XynF (Csac_2411) was unsuccessful. As an alternative, the N-terminus (N_term, GH43, and CBM22) and C-terminus (C_term, GH43, and CBM6) of XynF were cloned and expressed separately. Neither segment of XynF was active on xylan, although both of the separate domains hydrolyzed pNP-arabinose and arabinoxylan (wheat and rye), releasing α-L-arabinofurnaoside; in fact, XynF-C_term was fourfold more active on both arabinoxylans than XynF-N_term. This result is not surprising since the two separate GH43 domains have low amino acid sequence homology (identity = 26%) to each other. Although the function of intact XynF is not clear, the results here implicate a role in degrading the arabinoxylan component of hemicellulose.

It is interesting that the xylanolytic enzymes, XynA (Csac_2408) and XynE (Csac_2410) exhibit CMCase activity (Luthi et al., 1990a). Other GH10 family enzymes have been found to hydrolyze *para*-nitrophenyl β-D-cellobioside, but not all have non-specific endoglucanase activity (CMCase; Perez-Avalos et al., 2008). XynD (Csac_2409, GH39) is xylanolytic, which was unexpected, since this GH family contains only characterized α-L-iduronidases and β-xylosidases. XynB (GH39) from *Thermoanaerobacterium saccharolyticum* strain B6A-RI, which has strong similarity to XynD (Csac_2409) (aa identity = 71%, blastp), was active towards xylopentose, xylotriose, xylobiose, and pNPX, but not toward xylan (Lee and Zeikus, 1993). These broadly specific enzymes most probably work in concert with the arabinofuranosidase XynF, to disassociate the cellulose–hemicellulose matrix to facilitate complete degradation of the hemicellulose backbone.

celA-manB GH Locus

The second locus of interest investigated was the cellulase–hemicellulase gene cluster *celA-manB* (Csac_1076-Csac_1080; Fig. 2A). The enzymes encoded in this locus, CelA (Csac_1076), ManA (Csac_1077), CelB (Csac_1078), CelC (Csac_1079), and ManB (Csac_1080), are all extracellular, with the exception of the single domain ManB, which may originally have been part of CelC (Gibbs et al., 1992; Morris et al., 1995; Saul et al., 1990; Te'o et al., 1995). Previous characterization studies with these enzymes focused on the separate domains. CelA's GH9 domain exhibited endo-cellulase activity on CMC, lichenan, and konjac glucomannan. ManA's GH5 domain, like ManB's GH5 domain, hydrolyzes mannan, while its GH44 domain hydrolyzes CMC, oat spelt xylan and lichenin. Characterization of the single catalytic domain enzyme, CelC, found the enzyme had endoglucanase activity on CMC. Elements of the cellulase–hemicellulase locus *celA-manB* is thus far found only in the two genome-sequenced *Caldicellulosiruptor* species: *C. saccharolyticus* and *C. bescii* (formerly *Anaerocellum thermophilum*; Yang et al., 2009b). However, one differentiating feature between *C. bescii* and *C. saccharolyticus* is the multi-domain GH encoded in Csac_1078, referred to as CelB.

Differences were found between the genome amino acid sequence of the *C. saccharolyticus* CelB (Csac_1078) (van de Werken et al., 2008) and the other versions of CelB (Gibbs et al., 2000; Saul et al., 1990). For example, CelB, encoded in Csac_1078, has an additional 18 amino acids in the first linker region and 53 additional amino acids in the second linker region, compared to the CelB first reported (Saul et al., 1990; Fig. 3). The nucleotides encoding the repeated linkers may act as introns, enabling discrete domains to be excised and fused to other genes, thus generating novel hybrid enzymes (Bergquist et al., 1999). In fact, the sequence of the first linker addition (TVTATPTPTPTPVSTPAT) is also found in the *manA* and *celC* genes. Since no homologs to the second linker addition could be identified, it may be the product of multiple homologous recombinations. From a biochemical function perspective, the PT-rich linkers likely optimize the geometry between the binding and catalytic domains. Thus, the catalytic domains and linker sequences are positioned to create a conformational compromise between the requirement for tight binding to the substrate and the need to attack other hydrolysis sites (Receveur et al., 2002; von Ossowski et al., 2005). A third version of CelB, identified in *Caldicellulosiruptor* sp. Tok7B.1 (Gibbs et al., 2000), has three carbohydrate-binding domains, instead of the single domain present in the other two (Figure S2). The catalytic domains of the *Caldicellulosiruptor* sp. Tok7B.1 CelB have >85% amino acid identity to CelB. The GH5 domain of CelB from *Caldicellulosiruptor* sp. Tok7B.1 was active on CMC (Gibbs et al., 2000).

Biochemical Characteristics of CelB (Csac_1078)

Based on the sequence differences from previously characterized CelB, biochemical characterization of cellulase/hemicellulase CelB on model lignocelluloses substrate was undertaken. The influence of pH and temperature on CelB activity was investigated for the substrates CMC and birchwood xylan. Recombinant CelB was optimally active on birchwood xylan and CMC at pH 6.0. However, temperature optima were different for xylanase (80°C) and CMCase (95°C) activities. A similar observation was made for CelA from *C. bescii* (Athe_1867), which had different optimal temperatures for Avicelase activity (85°C) and CMC hydrolysis (95–100°C; Zverlov et al., 1998). A

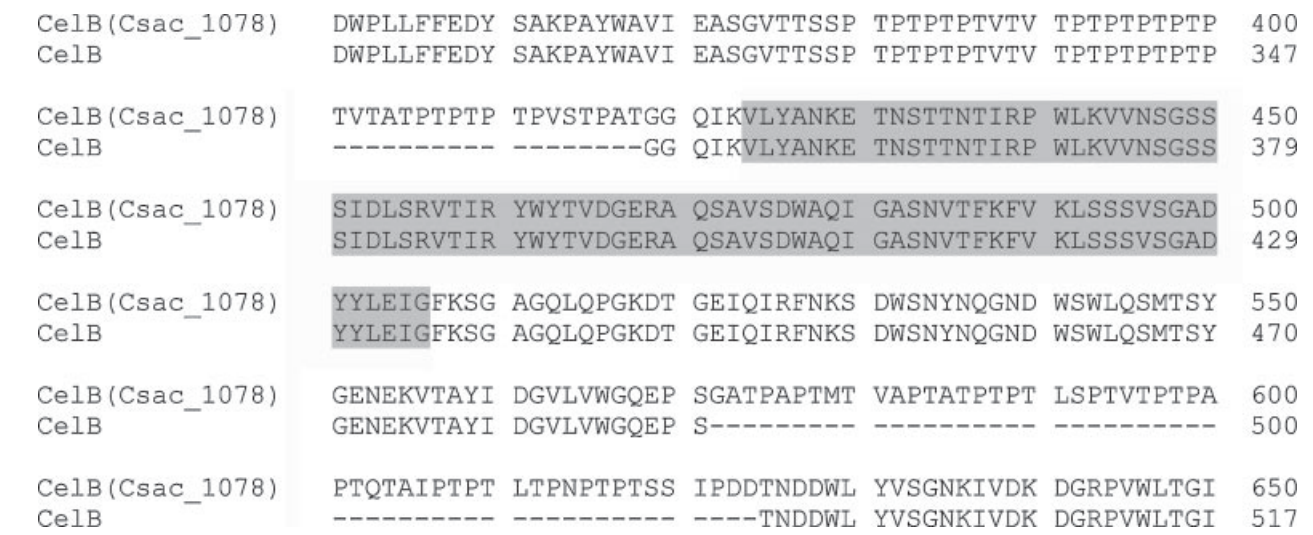

Figure 3. Selected sections of a full multiple sequence alignment of CelB (Csac_1078) from the genome sequence of *C. saccharolyticus* (van de Werken et al., 2008) and the original CelB from Saul et al. (1990) performed by CLUSTALW. CelB has extended PT boxes, linking the GH5 and GH 10 domains to the CBM (shaded region).

half-life of 29 h ($k_d = 0.02\,h^{-1}$) at 70°C for CelB against CMC was previously reported by Saul et al. (1990). Here, CMCase inactivation for CelB was bi-phasic, with the initial phase of rapid inactivation, k_{d1}, of $2.0\,h^{-1}$, followed by a second phase with a much lower k_{d2} (i.e., $0.03\,h^{-1}$). CelB xylanase inactivation was also similarly bi-phasic (k_{d1} of $1.2\,h^{-1}$ and k_{d2} of $0.30\,h^{-1}$). The reason for the variation in thermostability between the two GH domains is unclear, but could be related to lateral gene transfer events from different sources with dissimilar thermophilicities. This could also be attributed to the respective roles of the two GH domains in *C. saccharolyticus* growth physiology on polysaccharides.

Also examined was the effect of metal ions on hydrolysis activity of CelB. Enzyme assays were carried out after treatment with ethylenediaminetetraacetic acid (EDTA) by the addition of metal ions such as Fe^{3+}, Cu^{2+}, Mn^{2+}, Ca^{2+}, Zn^{2+}, Ba^{2+}, Mg^{2+}, and K^{+}, each at 1 mM. Only Cu^{2+}, which reduced CMCase activity to 40% and the xylanase activity to 76%, impacted enzyme function to any significant extent. This is not surprising, since this cation is known to adversely impact endoglucanases and xylanases (Bhiri et al., 2008; Gao et al., 2008; Picart et al., 2008; Zhang et al., 2007).

The specific activity of recombinant CelB was determined at 80°C for a number of polysaccharide substrates that differ in backbone composition, types of glycosidic linkages, and side-chain residues (Table II). CelB was active towards crystalline and amorphous cellulose (CMC, filter paper, Avicel), barley β-glucan, various xylans (birchwood and oat

Table II. Substrate specificity of CelB (Csac_1078) at 80°C.

		Activity (U mol^{-6})[a]			
Polysaccharides	Backbone	CelB	CelB-GH10	CelB-GH5	CelB-GH10 + CelB-GH5
Xylan birchwood	Xyl β1 → 4 Xyl	262 ± 11	84.7 ± 5.6	20.6 ± 1.6	88.5 ± 1.4
Xylan oat spelt	Xyl β1 → 4 Xyl	61.1 ± 8.1	15.0 ± 0.6	5.0 ± 0.4	30.6 ± 0.6
Arabinoxylan	Xyl β1 → 4 Xyl	134.1 ± 9.9	84.7 ± 2.1	25.1 ± 1.2	68.6 ± 0.9
CMC	Glu β1 → 4 Glu	13.5 ± 0.6	4.0 ± 0.2	5.1 ± 0.7	5.0 ± 0.5
Barley β-glucan	Glu β1 → 3,4 Glu	224.8 ± 3.6	77.6 ± 8.3	64.4 ± 8.4	69.9 ± 3.5
Glucomannan	Glu β1 → 4 Man	158.9 ± 1.4	6.7 ± 0.9	34.8 ± 3.2	49.2 ± 1.8
Filter paper	Glu β1 → 4 Glu	1.8 ± 0.2	ND	0.9 ± 0.1	0.85 ± 0.09
Avicel	Glu β1 → 4 Glu	0.40 ± 0.05	ND	0.17 ± 0.03	ND
Xyloglucan	Glu β1 → 4 Glu Xyl β1 → 4 Glu	ND[b]			
Mannan	Man β1 → 4 Man	ND			
Galactomannan	Man β1 → 4 Man Gal α1 → 6 Man	ND			

[a]U, reducing sugar equivalents released per min.
[b]ND, activity not detectable.

spelt) and wheat arabinoxylan. However, no activity was detected on xyloglucan, galactomannan, or mannan.

The previous study of CelB from another *Caldicellulosiruptor* sp. utilized a series of sub-clones and deletion mutants to show that the carbohydrate-binding module (CBM) affected neither CMC nor 4-methyl umbelliferyl-p-D-cellobioside (MUC) hydrolysis (Saul et al., 1990). Here, recombinant versions of the individual GH5 and GH10 domains, lacking the CBM, each exhibited broad xylanase and endoglucanase function, with highest activity on birchwood xylan and barley β-glucan (Table II). This contrasts with previous work (Saul et al., 1990), where neither CMCase activity nor MUCase activity were both present in the separate catalytic domains. The intact CelB had higher activity on all substrates tested than either the individual GH5 (CelB-GH5) and GH10 (CelB-GH10) domains or the two separate domains added together. This indicates that the CBM and the spatial relationship among the domains play an important role in polysaccharide hydrolysis (Table II).

Biomass Transcriptomes

To provide insights into the strategies used by *C. saccharolyticus* to utilize complex carbohydrates, the bacterium was grown separately on switchgrass and poplar and the resulting transcriptomes were compared to growth on glucose and xylose (Tables S2–S5). *C. saccharolyticus* exhibited significant growth on all substrates tested, in each case reaching cell densities of $\sim 10^8$ cells L^{-3}. Growth rates varied from $t_d = 90$ min on glucose and xylose, to $t_d = 120$ min on switchgrass, to $t_d = 230$ min on poplar.

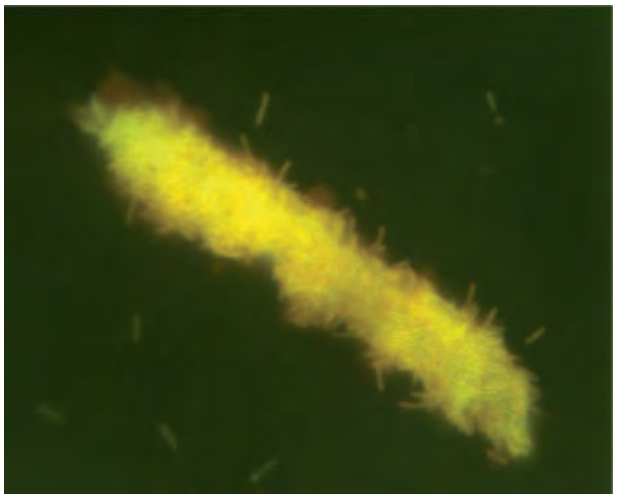

Figure 4. Epifluorescence micrograph of *C. saccholyticus* grown on switchgrass. *C. saccharolyticus* was grown in batch culture on 4% (wt/vol) switchgrass at 70°C. Note the bacterial rods attached to the surface of the plant biomass.

Table III. Number of differentially transcribed ORFs for *C. saccharolyticus* growth on monosaccharides and biomass.

	Switchgrass	Poplar	Glucose	Xylose
Switchgrass		15	98	83
Poplar	28		41	72
Glucose	32	18		86
Xylose	66	70	52	

The numbers in the boxes represent differentially transcribed ORFs (≥2-fold) for comparisons between column and corresponding row. For example, 15 ORFs were up-regulated on poplar compared to switchgrass.

C. saccharolyticus attachment to switchgrass was particularly evident, while attachment to poplar was minimal (Fig. 4).

Table III summarizes the number of ORFs differentially transcribed twofold or more for these experiments. A number of ORFs (~14%) responding to growth on plant biomasses have current annotations as "hypothetical proteins." Comparing the plant biomass to the monosaccharide response, a majority of the genes that responded were, as might be expected, components of carbohydrate transporters (VanFossen et al., 2009) and GHs. These results indicate regulation of carbohydrate metabolism at the transcription level.

Many carbohydrate transporters responded to growth on biomass. Carbohydrate uptake family 1 (CUT1) transporters Csac_0679-Csac_0682, Csac_1557-Csac_1559, Csac_2412-Csac_2414, and Csac_2417-Csac_2419 all responded to biomass, consistent with their annotation as xyloglucan or xylooligosaccharide transporters (VanFossen et al., 2009). CUT2 transporters components (Csac_0238/Csac_0240-Csac_0242 and Csac_2504-Csac_2506) responded to one of the monosaccharide and at least one of the biomass substrates. For example, the substrate-binding protein encoded in Csac_0242 responded to xylose compared to glucose, switchgrass, and poplar, and on switchgrass compared to glucose. The differential transcription demonstrates *C. saccharolyticus* adaptability to lignocelluloses substrates for maximal capture of available carbohydrates.

Heat plots depicting the response of the ORFs encoding the GHs in Table I to the plant biomass substrates (switchgrass, poplar) compared to glucose and xylose are shown in Figure 5. As is evident from these plots, the ORFs encoding GHs vary widely in terms of transcriptional response and level. Results showed that ~20% of all the GHs responded twofold or more for the plant biomass versus monosaccharide contrasts, including several ORFs in *celA-manB* (Csac_1076-Csac_1080) and *xynB-xynF* (Csac_2404-Csac_2411) loci (Fig. 2B). The results were compared to those found previously with growth on the hemicelluloses xyloglucan and xylogluco-oligosaccharides (provided by H. Brumer of the Royal Institute of Technology, Stockholm, Sweden) and oat spelt xylan (Figs. 2C and 5; GSE17436). Growth rates and yields on the hemicelluloses ($t_d = 90$ min) were comparable to those found for the monosaccharides glucose and xylose.

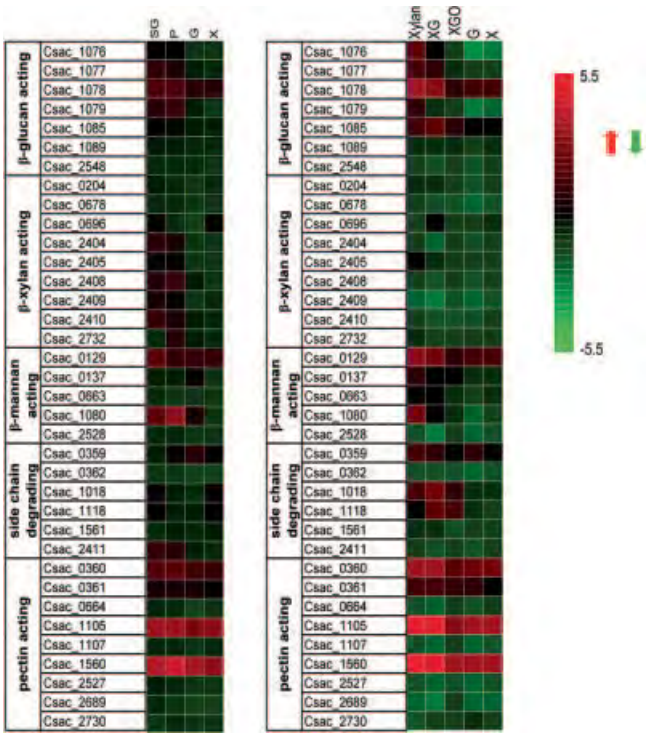

Figure 5. Gene expression levels (least square mean estimates) of ORFs encoding GHs in the *C. saccharolyticus* genome. Separate experimental loops were used for poplar and switchgrass/poplar (**left**) and hemicelluloses (**right**). SG, switchgrass; P, poplar; G, glucose; X, xylose; XG, xyloglucan; XGO, xylogluco-oligosaccharide. Least squares mean (lsm) estimates (Materials and Methods Section) of transcript level shown for selected genes (ORF#). Shading denotes transcript levels above (black) and below (white) the mean (gray) across all genes.

Examination of the transcriptome found that the side-chain acting GHs (Table I) did not respond to poplar, switchgrass, or the monosaccharides. A similar response was also seen with the pectin-acting enzymes encoded in Csac_1105 and Csac_1560, although each was consistently transcribed at very high values (top 6% of lsm values) under all conditions. The lack of regulation of these GHs at the transcription level may be the result of constitutive expression to take advantage of substrates in *C. saccharolyticus*' natural environment.

The transcription level (lsm) of genes in the polysaccharide-degrading locus *celA-manB* on poplar and switchgrass were responsive relative to glucose and xylose (Fig. 2B). The normalized transcription levels for the endoglucanases (containing GH5, GH9, GH16, GH44, GH48, and GH74 domains) on the plant biomass were such that *celB* > *manA*/*celC* > Csac_1085 > *celA* > Csac_2548, although only *manA* and *celC* were differentially transcribed (twofold or more) compared to glucose and xylose. While several β-xylan-degrading enzymes (containing GH10 and GH39 domains), including *xynB*, *ORF4*, *xynA*, *xynD*, and *xynE* were all differentially transcribed twofold or more for one of the plant biomass-monosaccharide contrasts (Fig. 2B), transcription levels for *CelB* were higher than any gene with a GH10 or GH39 domain in the *C. saccharolyticus* genome suggesting the importance of CelB in the hydrolysis of lignocellulose.

manB was differentially transcribed on both plant biomasses relative to both monosaccharides, while *manA* was only up-regulated on switchgrass. This is an unexpected pattern due to the higher reported levels of mannan in raw poplar compared to switchgrass (Esteghlalian et al., 1997). Also, it has been reported that the pH profile, temperature profile, and specific activity of the two enzymes are identical (Morris et al., 1995). It is possible the acid pretreatment of the switchgrass may have released manno-oligosaccharides that trigger higher transcription of *manA*.

The *celA-manB* locus was up-regulated on xylan and xyloglucan, compared to growth on one or both monosaccharides (Fig. 2C). *celA-manB* was also up-regulated on the hemicellulose polysaccharides compared to xyloglucan-oligosaccharides. *celB*, which was highly transcribed on switchgrass and poplar, showed large fold changes on the model hemicellulose substrates compared to the monosaccharides. *celB* was highly transcribed (lsm $\geq$ 3.0) on both xylan and xyloglucan. In contrast, the xylanases in *xynB*, *ORF4*, and *xynA-xynF* did not respond to xylan or xyloglucans and were transcribed at low levels, in general. Growth on xylogluco-oligosaccharides resulted in transcription of *xynB* and *xynD* compared to xyloglucan, suggesting that the transcription of these genes is be triggered by oligosaccharides.

Although differential transcription was not observed on the complex carbohydrates compared to the monosaccharides, the presence of CelA, ManA, CelB, and CelC in the secretome of *C. saccharolyticus* grown on glucose has been reported (Andrews et al., 2010). This base level of expression indicates that mechanisms for carbohydrate uptake and processing are always functional in *C. saccharolyticus*.

Conclusions

C. saccharolyticus can simultaneously ferment a range of simple sugars to H_2 at high rates and with high yields (Kadar et al., 2003; van de Werken et al., 2008; VanFossen et al., 2009). The uniquely broad substrate range for certain GHs in *C. saccharolyticus* facilitates its capacity to use lignocellulose as a source of these fermentable sugars. The catabolic machinery regulation at the level of transcription and constitutive expression of GHs, suggest that *C. saccharolyticus* is adaptable, dynamic, and designed for complex carbohydrate utilization. These factors make this bacterium a good model for examining fundamental microbiological mechanisms related to consolidated bioprocessing. More information from other members of this genus has been gathered but more further is needed to help clarify the intrinsic metabolic and enzymological bases required for plant biomass deconstruction and efforts along these lines are now underway (Blumer-Schuette et al., 2010).

R.M.K. acknowledges support from the US National Science Foundation (CBT0617272) and the Bioenergy Science Center (BESC), a U.S. DOE Bioenergy Research Center supported by the Office of Biological and Environmental Research. A.L.V. acknowledges support from a US Department of Education GAANN Fellowship. We thank H. Brumer of the Royal Institute of Technology, Stockholm, Sweden for providing hemicellulose samples.

REFERENCES

Andrews G, Lewis D, Notey J, Kelly R, Muddiman D. 2010. Part I: Characterization of the extracellular proteome of the extreme thermophile *Caldicellulosiruptor saccharolyticus* by GeLC-MS2. Anal Bioanal Chem 398(1):377–389.

Bergquist PL, Gibbs MD, Morris DD, Te'o VS, Saul DJ, Morgan HW. 1999. Molecular diversity of thermophilic cellulolytic and hemicelluloytic bacteria. FEMS Microbiol Ecol 28:99–110.

Bhiri F, Chaabouni SE, Limam F, Ghrir R, Marzouki N. 2008. Purification and biochemical characterization of extracellular beta-glucosidases from the hypercellulolytic Pol6 mutant of *Penicillium occitanis*. Appl Biochem Biotechnol 149(2):169–182.

Blumer-Schuette SE, Lewis DL, Kelly RM. 2010. Phylogenetic, microbiological and glycoside hydrolase diversity within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. Appl Environ Microbiol 76:8084–8092.

Bouton JH. 2007. Molecular breeding of switchgrass for use as a biofuel crop. Curr Opin Genet Dev 17(6):553–558.

de Vrije T, Mars AE, Budde MA, Lai MH, Dijkema C, de Waard P, Claassen PA. 2007. Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus*. Appl Microbiol Biotechnol 74(6):1358–1367.

Esteghlalian A, Hashimoto AG, Fenske JJ, Penner MH. 1997. Modeling and optimization of the dilute-sulfuric-acid pretreatment of corn stover, poplar and switchgrass. Bioresour Technol 59(2–3):129–136.

Gao J, Weng H, Xi Y, Zhu D, Han S. 2008. Purification and characterization of a novel endo-beta-1,4-glucanase from the thermoacidophilic *Aspergillus terreus*. Biotechnol Lett 30(2):323–327.

Gibbs MD, Saul DJ, Luthi E, Bergquist PL. 1992. The beta-mannanase from *Caldocellum saccharolyticum* is part of a multidomain enzyme. Appl Environ Microbiol 58(12):3864–3867.

Gibbs MD, Reeves RA, Farrington GK, Anderson P, Williams DP, Bergquist PL. 2000. Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile *Caldicellulosiruptor* isolate Tok7B.1. Curr Microbiol 40(5):333–340.

Hardiman E, Gibbs M, Reeves R, Bergquist P. 2010. Directed evolution of a thermophilic beta-glucosidase for cellulosic bioethanol production. Appl Biochem Biotechnol 161(1–8):301–312.

Hong MR, Kim YS, Park CS, Lee JK, Oh DK. 2009a. Characterization of a recombinant beta-glucosidase from the thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. J Biosci Bioeng 108(1):36–40.

Hong MR, Park CS, Oh DK. 2009b. Characterization of a thermostable endo-1,5-alpha-L-arabinanase from *Caldicellulorsiruptor saccharolyticus*. Biotechnol Lett 31(9):1439–1443.

Kadar Z, De Vrije T, Budde MA, Szengyel Z, Reczey K, Claassen PA. 2003. Hydrogen production from paper sludge hydrolysate. Appl Biochem Biotechnol 105–108:557–566.

Khandeparker R, Numan MT. 2008. Bifunctional xylanases and their potential use in biotechnology. J Ind Microbiol Biotechnol 35(7): 635–644.

Konig J, Grasser R, Pikor H, Vogel K. 2002. Determination of xylanase, beta-glucanase, and cellulase activity. Anal Bioanal Chem 374(1):80–87.

Kumar P, Barrett DM, Delwiche MJ, Stroeve P. 2009. Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. Ind Eng Chem Res 48(8):3713–3729.

Lee YE, Zeikus JG. 1993. Genetic organization, sequence and biochemical characterization of recombinant beta-xylosidase from *Thermoanaerobacterium saccharolyticum* strain B6A-RI. J Gen Microbiol 139(Pt 6): 1235–1243.

Lim YR, Yoon RY, Seo ES, Kim YS, Park CS, Oh DK. 2010. Hydrolytic properties of a thermostable alpha-L-arabinofuranosidase from *Caldicellulosiruptor saccharolyticus*. J Appl Microbiol 109(4):1188–1197.

Love DR, Fisher R, Bergquist PL. 1988. Sequence structure and expression of a cloned beta-glucosidase gene from an extreme thermophile. Mol Gen Genet 213(1):84–92.

Luthi E, Jasmat NB, Bergquist PL. 1990a. Xylanase from the extremely thermophilic bacterium *Caldocellum saccharolyticum*: Overexpression of the gene in *Escherichia coli* and characterization of the gene product. Appl Environ Microbiol 56(9):2677–2683.

Luthi E, Love DR, McAnulty J, Wallace C, Caughey PA, Saul D, Bergquist PL. 1990b. Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophile *Caldocellum saccharolyticum*. Appl Environ Microbiol 56(4):1017–1024.

Luthi E, Jasmat NB, Grayling RA, Love DR, Bergquist PL. 1991. Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding for a beta-mannanase from the extremely thermophilic bacterium *Caldocellum saccharolyticum*. Appl Environ Microbiol 57(3):694–700.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol Mol Biol Rev 66(3):506–577. Table of contents.

Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: An update. Curr Opin Biotechnol 16(5):577–583.

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE. 2008. How biotech can transform biofuels. Nat Biotechnol 26(2):169–172.

Morris DD, Reeves RA, Gibbs MD, Saul DJ, Bergquist PL. 1995. Correction of the beta-mannanase domain of the celC pseudogene from *Caldocellulosiruptor saccharolyticus* and activity of the gene product on kraft pulp. Appl Environ Microbiol 61(6):2262–2269.

Park AR, Oh DK. 2010. Effects of galactose and glucose on the hydrolysis reaction of a thermostable beta-galactosidase from *Caldicellulosiruptor saccharolyticus*. Appl Microbiol Biotechnol 85(5):1427–1435.

Perez-Avalos O, Sanchez-Herrera LM, Salgado LM, Ponce-Noyola T. 2008. A bifunctional endoglucanase/endoxylanase from *Cellulomonas flavigena* with potential use in industrial processes at different pH. Curr Microbiol 57(1):39–44.

Picart P, Diaz P, Pastor FI. 2008. Stachybotrys atra BP-A produces alkali-resistant and thermostable cellulases. Antonie Van Leeuwenhoek 94(2):307–316.

Pysz MA, Conners SB, Montero CI, Shockley KR, Johnson MR, Ward DE, Kelly RM. 2004. Transcriptional analysis of biofilm formation processes in the anaerobic, hyperthermophilic bacterium *Thermotoga maritima*. Appl Environ Microbiol 70(10):6098–6112.

Rainey FA, Donnison AM, Janssen PH, Saul D, Rodrigo A, Bergquist PL, Daniel RM, Stackebrandt E, Morgan HW. 1994. Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. FEMS Microbiol Lett 120(3):263–266.

Receveur V, Czjzek M, Schulein M, Panine P, Henrissat B. 2002. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. J Biol Chem 277(43):40887–40892.

Saul DJ, Williams LC, Grayling RA, Chamley LW, Love DR, Bergquist PL. 1990. celB, a gene coding for a bifunctional cellulase from the extreme thermophile *Caldocellum saccharolyticum*. Appl Environ Microbiol 56(10):3117–3124.

Shoseyov O, Shani Z, Levy I. 2006. Carbohydrate binding modules: Biochemical properties and novel applications. Microbiol Mol Biol Rev 70(2):283–295.

Simmons BA, Loque D, Blanch HW. 2008. Next-generation biomass feedstocks for biofuel production. Genome Biol 9(12):242.

Te'o VS, Saul DJ, Bergquist PL. 1995. celA, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. Appl Microbiol Biotechnol 43(2):291–296.

van de Werken HJ, Verhaart MR, VanFossen AL, Willquist K, Lewis DL, Nichols JD, Goorissen HP, Mongodin EF, Nelson KE. van Niel EW, et al. 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Appl Environ Microbiol 74(21): 6720–6729.

VanFossen AL, Lewis DL, Nichols JD, Kelly RM. 2008. Polysaccharide degradation and synthesis by extremely thermophilic anaerobes. Ann NY Acad Sci 1125:322–337.

VanFossen AL, Verhaart MR, Kengen SM, Kelly RM. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl Environ Microbiol 75(24):7718–7724.

von Ossowski I, Eaton JT, Czjzek M, Perkins SJ, Frandsen TP, Schulein M, Panine P, Henrissat B, Receveur-Brechot V. 2005. Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulase. Biophys J 88(4):2823–2832.

Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 8(6):625–637.

Yang SJ, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doeppke C, Davis M, Westpheling J, Adams MW. 2009a. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe *Anaerocellum thermophilum* DSM 6725. Appl Environ Microbiol 75(14):4762–4769.

Yang SJ, Kataeva I, Wiegel J, Yin Y, Dam P, Xu Y, Westpheling J, Adams MW. 2009b. Classification of *Anaerocellum thermophilum* strain DSM 6723 as *Caldicellulosiruptor bescii* sp. nov. Int J Syst Evol Microbiol 60(Pt 9):2011–2015.

Zaldivar J, Nielsen J, Olsson L. 2001. Fuel ethanol production from lignocellulose: A challenge for metabolic engineering and process integration. Appl Microbiol Biotechnol 56(1–2):17–34.

Zhang GM, Huang J, Huang GR, Ma LX, Zhang XE. 2007. Molecular cloning and heterologous expression of a new xylanase gene from *Plectosphaerella cucumerina*. Appl Microbiol Biotechnol 74(2):339–346.

Zverlov V, Mahr S, Riedel K, Bronnenmeier K. 1998. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile *Anaerocellum thermophilum* with separate glycosyl hydrolase family 9 and 48 catalytic domains. Microbiology 144(Pt 2):457–465.