

Use of Label-Free Quantitative Proteomics To Distinguish the Secreted Cellulolytic Systems of *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis*^{∇†}

Adriane Lochner,^{1,3,4} Richard J. Giannone,^{2,3} Miguel Rodriguez, Jr.,^{1,3} Manesh B. Shah,² Jonathan R. Mielenz,^{1,3} Martin Keller,^{1,3} Garabed Antranikian,⁴ David E. Graham,^{1,5*} and Robert L. Hettich^{2,3*}

Biosciences Division,¹ Chemical Sciences Division,² and BioEnergy Science Center,³ Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831; Technical Microbiology, Hamburg University of Technology, Kasernenstrasse 12, D-21073 Hamburg, Germany⁴; and Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996⁵

Received 1 December 2010/Accepted 9 April 2011

The extremely thermophilic, Gram-positive bacteria *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis* efficiently degrade both cellulose and hemicellulose, which makes them relevant models for lignocellulosic biomass deconstruction to produce sustainable biofuels. To identify the shared and unique features of secreted cellulolytic apparatuses from *C. bescii* and *C. obsidiansis*, label-free quantitative proteomics was used to analyze protein abundance over the course of fermentative growth on crystalline cellulose. Both organisms' secretomes consisted of more than 400 proteins, of which the most abundant were multidomain glycosidases, extracellular solute-binding proteins, flagellin, putative pectate lyases, and uncharacterized proteins with predicted secretion signals. Among the identified proteins, 53 to 57 significantly changed in abundance during cellulose fermentation in favor of glycosidases and extracellular binding proteins. Mass spectrometric characterizations, together with cellulase activity measurements, revealed a substantial abundance increase of a few bifunctional multidomain glycosidases composed of glycosidase (GH) domain family 5, 9, 10, 44, or 48 and family 3 carbohydrate binding (CBM3) modules. In addition to their orthologous cellulases, the organisms expressed unique glycosidases with different domain organizations: *C. obsidiansis* expressed the COB47_1671 protein with GH10/5 domains, while *C. bescii* expressed the Athe_1857 (GH10/48) and Athe_1859 (GH5/44) proteins. Glycosidases containing CBM3 domains were selectively enriched via binding to amorphous cellulose. Preparations from both bacteria contained highly thermostable enzymes with optimal cellulase activities at 85°C and pH 5. The *C. obsidiansis* preparation, however, had higher cellulase specific activity and greater thermostability. The *C. bescii* culture produced more extracellular protein and additional SDS-PAGE bands that demonstrated glycosidase activity.

The conversion of lignocellulosic feedstock into biofuels has garnered significant interest recently in light of an increasing global demand for transportation fuel alternatives. Clearly, the natural process of plant cell wall degradation by microorganisms serves as an informative guideline for the design and optimization of efficient industrial enzymatic conversion processes. To this end, there is a strong emphasis on a comprehensive understanding of complex microbial biomass degradation systems that consist primarily of glycosidases, including synergistically acting cellulases and hemicellulases (21). The increasing number of whole genome sequences, along with sophisticated experimental and computational technologies, has provided a remarkable glimpse into the molecular processes by which microorganisms degrade cellulosic material.

One of the leading technologies employed for the system level interrogation of various organisms is mass spectrometry, which during the last decade has revolutionized the large-scale, high-throughput proteomic characterization of both microbial isolates and communities (3, 24). In particular, proteomics has accelerated the discovery and quantification of cellulose-degrading proteins from both aerobic and anaerobic microorganisms. For example, mass spectrometry (MS)-based proteomic measurements revealed numerous monofunctional glycosidase proteins found in the secretome of the mesophilic fungus *Trichoderma reesei* (20, 50), currently the source of most commercial cellulose preparations (22, 35). The mesophilic marine gammaproteobacterium *Saccharophagus degradans* also encodes multiple glycosidases, many fused to carbohydrate binding modules or cadherin domains; however, only a handful were identified by proteomic measurement of the culture supernatant (45).

Thermostable glycosidases produced by thermophilic microbes offer many advantages for large-scale biofuel production, including increased protein stability and cellulose degradation rates and a reduced risk of microbial contamination (5, 41). Although several thermophilic microbes are able to degrade cellulosic biomass, they do so by one of several unique strategies. The moderate thermophile (55°C) *Thermobifida*

* Corresponding author. Mailing address for David E. Graham: Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6038. Phone: (865) 574-0559. Fax: (865) 576-8646. E-mail: grahamde@ornl.gov. Mailing address for Robert L. Hettich: Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6131. Phone: (865) 574-4986. Fax: (865) 576-8559. E-mail: hettichrl@ornl.gov.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

∇ Published ahead of print on 15 April 2011.

fusca secretes a diverse set of monofunctional glycosidases, often fused to carbohydrate-binding domains (1, 53). The thermophilic (60°C) bacterium *Clostridium thermocellum* produces glycosidases in cellulosomal complexes (38), while the extremely thermophilic (70°C) bacterium *Caldicellulosiruptor saccharolyticus* secretes several multidomain, multifunctional cellulases (4, 21).

The use of proteomics for system level investigation of cellulose degradation strategies goes beyond general protein identification. Quantitative proteomics enables the derivation of relative or absolute measurements of protein abundance within a given sample. Several unique strategies exist, including isotopic-label-based methods such as ICAT or iTRAQ and "label-free" methods that utilize spectral counts (SpC), intensity, or chromatographic peak area to estimate protein abundance (36, 37). With regard to its application to cellulose-degrading thermophiles, quantitative proteomics has been employed to study *C. thermocellum*'s cellulolytic enzymes based on metabolic labeling of cells grown on cellulose or cellobiose (16), while a label-free approach measured changes in cellulosomal protein composition using cells grown on different substrates (38). In both cases, quantitative analysis permitted estimation of relative protein abundance, providing valuable comparisons of cultures grown on different substrates based on changes in protein expression.

The *Caldicellulosiruptor* genus of anaerobic Gram-positive bacteria includes *C. saccharolyticus*, as well as *Caldicellulosiruptor bescii* (optimal temperature, 75°C; previously *Anaerocellum thermophilum* DSM 6725) (55) and *Caldicellulosiruptor obsidiensis* (78°C) (18). *C. bescii* and *C. obsidiensis* share 97% 16S rRNA gene sequence identity and 88% average nucleotide identity in their genome sequences (18), and both grow on the same set of monomeric and polymeric sugars as carbon sources, including pretreated switchgrass and poplar (18, 49). However, *C. obsidiensis* grows at slightly higher temperatures than *C. bescii* does and only *C. obsidiensis* produces measurable amounts of ethanol during growth on switchgrass or cellulose (18, 42, 54). Putative cellulase genes in both organisms are concentrated in an island associated with prophage genes. This cluster varies in size and gene composition among the *Caldicellulosiruptor* species: it comprises 48 kbp in *C. saccharolyticus* (47), 61 kbp in *C. obsidiensis* (9), and 68 kbp in *C. bescii* (23).

Genomic, proteomic, and physiological studies have shown that catabolic enzymes and pathways evolve rapidly through positive selection, differentiating closely related microbial species (28). Proteomic analysis of the secretomes of two evolved *T. reesei* strains showed the diversification of glycosidase profiles in these fungi, as revealed by changes in expression or secretion efficiency (20). The genetic diversity among *Caldicellulosiruptor* cellulase gene clusters and the rapid evolution of thermostable, multidomain, multifunctional glycosidases in this lineage warranted a comparison of the strains' secreted cellulolytic protein complement.

To compare cellulolytic systems from *C. bescii* and *C. obsidiensis*, we analyzed their secreted protein profiles over the course of crystalline cellulose fermentations. Label-free quantitative proteomic analysis was performed using two-dimensional liquid chromatography (LC)-tandem MS (MS/MS) to identify secreted proteins. Identified proteins were then quan-

tified by normalized spectral abundance factors (NSAF), a method based on spectral counting that corrects for protein size and run-to-run variability (37). The correlation of these NSAF measurements with the carboxymethyl cellulase (CMCase) activity of the culture supernatant was assessed, and the increase in secreted glycosidase proteins over the course of fermentation for both organisms was monitored. Cellulose affinity digestion experiments were employed to enrich cellulose-binding proteins from both organisms to identify the highly thermostable multidomain glycosidases that are shared by and unique to each organism.

MATERIALS AND METHODS

Controlled cultivation and sampling. Fermentations were performed in BIOSTAT Bplus Twin 5-liter jacketed glass fermentors (Sartorius Stedim Biotech) using a 4-liter working volume of basal growth medium (18) that contained 4.5 mM KCl, 4.7 mM NH₄Cl, 2.5 mM MgSO₄, 1.0 mM NaCl, 0.7 mM CaCl₂ · 2H₂O, 0.25 mg/ml resazurin, 5.6 mM cysteine-HCl · H₂O, 6.0 mM NaHCO₃, 1 mM phosphate buffer, 1× ATCC trace minerals, 1.25× MTC medium vitamin solution E (57), 0.02% (wt/vol) yeast extract, and 0.5% (wt/vol) Avicel PH-101 (Fluka). The temperature was controlled at 75°C for *C. bescii* and at 78°C for *C. obsidiensis* using a Polystat Circulator (Cole-Parmer). Fermentors and media were sparged overnight at 200 rpm with a N₂-CO₂ (80:20) gas mixture; the exhaust gas was run through a water trap. The next day, yeast extract, sodium bicarbonate, and vitamins were added and sparged for an additional 3 to 4 h. Inocula were grown in 125-ml serum bottles and added to the fermentors to achieve an initial cell density of 2.6 × 10⁶ cells/ml. The agitation was set at 300 rpm, and the pH was controlled at 6.8 using a 10% sodium bicarbonate solution. Replicate samples were taken at elapsed fermentation time points of 0, 4, 8, 12, 16, 20, 24, 30, 40, and 48 h. These samples were processed as follows. Forty-milliliter samples for proteomic analysis and 10-ml samples for activity assays were centrifuged at 6,500 × g for 15 min, and the supernatant was filtered via a 0.22-μm syringe filter with a polyethersulfone (PES) membrane to obtain a cell-free culture supernatant fraction. Aliquots of 1.8 ml were centrifuged for 5 min at 15,500 × g, and the supernatants and pellets were frozen in liquid nitrogen and stored at -80°C. At 48 h, the fermentations were stopped and the broth was harvested by centrifugation (30 min at 5,500 × g). Two liters of supernatant was filtered through a 0.22-μm PES membrane Steritop device (Millipore), resulting in the cell-free supernatant fraction SN, which was subsequently concentrated 5 times via a Quixstand tangential-flow filtration system equipped with a hollow-fiber filter (GE Healthcare) with a 5,000-molecular-weight cutoff. The supernatant protein concentrate from tangential-flow filtration, designated TFF, was used for the cellulase enrichment procedure.

Planktonic cell densities were measured using a Petroff-Hausser microscope counting chamber (Fisher Scientific). Unless stated otherwise, supernatant protein concentrations, as well as concentrations of other protein solutions used in this study, were estimated by Bradford microassay (6) with bovine serum albumin as the standard. Pellet protein was estimated by the Lowry method with Peterson's modification (Sigma) after cell lysis in a 0.2 N NaOH-1% SDS solution as described previously (56).

Cellulase enrichment. The glycosidases in the supernatant protein concentrate (TFF) were enriched via binding to amorphous cellulose and subsequent digestion as described previously (38). Briefly, acid-swollen Avicel PH-105 was added to the TFF fraction of each respective organism in an amount corresponding to 50 mg crystalline cellulose. To maintain a similar protein/substrate ratio, concentrate volumes were adjusted to contain 20 mg total protein. After binding, the amorphous cellulose was separated by centrifugation and resuspended in 10 ml of reaction buffer containing 50 mM sodium acetate, pH 5.5. The remaining protein solution was referred to as affinity digest supernatant (ADSN). Affinity digestion was performed with a dialysis membrane (SpectraPor; 6- to 8-kDa cutoff) against reaction buffer at 75°C for 5 h with frequent changes of the dialysis buffer to prevent possible product inhibition. The reaction was considered complete after all visible traces of the substrate had disappeared. Residual substrate was removed from the affinity digest protein fraction (AD) by centrifugation.

Zymogram. SDS-PAGE (26) was performed with 4 to 20% Precise protein gels with the BupH Tris-HEPES-SDS running buffer (Thermo-Pierce). Protein bands were stained with Coomassie blue dye. CMCase bands were visualized using a modification of the zymogram technique as described previously (40). Gels were incubated in 1% carboxymethyl cellulose (CMC)

solutions (in 50 mM sodium acetate, pH 5.5) instead of incorporating these substrates into the polyacrylamide gel.

Cellulase assays. CMCase activities of the culture supernatant time course samples were assayed in 1-ml reaction mixtures containing 50 mM sodium acetate, pH 5.5, and 1.3% CMC. The protein concentrations in the assays varied from 3 to 10 mg. The reactions were started by combining the substrate solution with the protein-buffer mixture after 10 min of separate preincubation and run for 30 to 90 min, depending on the protein concentration in the assay. For temperature and pH optimum and thermostability determination experiments using the highly concentrated AD fractions, small volumes of enzyme solution were added to the preheated substrate-buffer mixture and the assay was run for 20 min. Temperature assays were conducted within a range of 40 to 100°C. For pH assays at 80°C, 50 mM citrate buffer was used at pH 3.5 to 6.5 and 50 mM potassium phosphate was used at pH 6.5 to 8.0. To determine thermostability, the protein solutions were incubated for 30, 60, or 90 min each at 75, 85, and 95°C prior to reaction start. To stop the reactions, a 250- μ l aliquot of sample was added to 500 μ l of 3,5-dinitrosalicylic acid reagent (2, 31). Reducing sugar concentrations were determined as glucose equivalents after boiling at 98°C for 5 min. The samples were then diluted 1:5 in H₂O, and the absorbance was measured at 540 nm. One unit of activity catalyzed the release of 1 μ mol glucose equivalent per minute. A modified version of the IUPAC standard assay was used to compare the activities of different fractions (SN, TFF, AD, and ADSN) between the two organisms. Enzyme units were determined by assaying different enzyme dilutions for 60 min at 80°C using 1.3% CMC or 2.0% Avicel PH-101 in order to identify the enzyme concentration that yields 4% substrate conversion as previously defined (2, 8).

Analysis of peptides by mini-multidimensional protein identification technology (MudPIT) LC-MS/MS analysis. Cell-free secretome samples were prepared for LC-MS analysis as follows. Proteins were denatured and reduced by addition of SDS lysis buffer (4% SDS in 100 mM Tris-HCl, pH 8.0, with 10 mM dithiothreitol [DTT]) at a 1:1 (vol/vol) ratio, boiled, and sonicated with a Branson sonic disruptor (20% amplitude for 2 min; 10-s pulse, 10-s pause). Trichloroacetic acid was added to a concentration of 20% (wt/vol) to precipitate sample proteins from detergent and solutes. Ice-cold, acetone-washed pelleted proteins were resuspended with 8 M urea in 100 mM Tris-HCl, pH 8.0. The amount of recovered protein was measured using the bicinchoninic acid (BCA) assay (Thermo-Pierce). Proteins were reduced with 5 mM DTT, alkylated with 10 mM iodoacetamide, and digested with two separate and sequential aliquots of sequencing grade trypsin (Promega) at a 1:100 (wt/wt) enzyme-to-protein ratio. As 8 M urea inhibits trypsin, samples were diluted to 4 M for an overnight digestion, followed by dilution to 2 M urea for a 4-h digestion. Samples were then adjusted to 150 mM NaCl and 0.1% formic acid and filtered through a 500- μ l 10-kDa cutoff spin column filter (VWR brand). Peptide concentrations were then measured using the BCA assay.

To compare the extracellular protein complements of the two organisms at each time point, a 25- μ g aliquot of peptides was bomb loaded onto a biphasic MudPIT back column as described previously (30, 52). Loaded peptides were then washed with solvent A (5% acetonitrile, 95% high-performance liquid chromatography [HPLC] grade water, 0.1% formic acid) for 20 min, followed by a 25-min gradient to solvent B (70% acetonitrile, 30% HPLC grade water, 0.1% formic acid) offline. Desalted peptides were then placed in line with an in-house pulled, reverse-phase packed nanospray emitter and subjected to a 4-step analysis as previously described (13), with modifications (salt pulses at 10%, 25%, 50%, and 100% 500 mM ammonium acetate, each followed by a 1-h organic gradient to 50% solvent B), referred to here as mini-MudPIT. LC-separated peptides were analyzed via a hybrid LTQ-Velos/Orbitrap mass spectrometer (Thermo Fisher) operating in a data-dependent fashion. Each full scan (2 microscans) generated by the Orbitrap mass analyzer (30,000 resolution) was followed by 10 parent ion isolations/MS/MS events (2 microscans) by the LTQ-Velos. Two replicate measurements were obtained for each sample.

MS data analysis and evaluation. Acquired MS/MS spectra were assigned to specific peptide sequences using the SEQUEST search algorithm (11) with a FASTA proteome database specific to either *C. obsidiansis* (9) or *C. bescii* (23). Both databases contained common contaminant protein entries, as well as reversed decoy entries to assess protein-level false-discovery rates. SEQUEST-scored peptide sequence data were filtered and assembled into protein loci using DTASelect (43) with the following conservative criteria: XCorr, +1 = 1.8, +2 = 2.5, and +3 = 3.5; DeltCN, 0.08; and 2 unique peptides per protein identification. Prior to semiquantitative analysis, SpC were rebalanced to properly distribute shared, nonunique peptides between their potential parent proteins, based solely on their unique SpC. NSAF values were then calculated for each protein (59).

NSAF values were imported into the JMP Genomics software package (ver.

4.1; SAS Institute) for statistical analyses (17). After transformation with the natural logarithm and standardization to correct for signal intensity, one-way analysis of variance (ANOVA) was used to identify proteins that show significant differences in abundance over time ($P < 0.01$). Cellulase protein abundance data for each organism were compared to growth curves and cellulase assays to ascertain differences in organism-specific cellulose degradation.

The predicted protein sequences of *C. bescii* (23) and *C. obsidiansis* (9) were submitted to the SignalP server (ver. 3.0) (10) to predict the presence of signal peptides. Hidden Markov model analysis, with parameters for Gram-positive bacteria, was used to identify signal peptidase I cleavage sites within the first 70 residues of each protein. Probability scores (SProb) were used to distinguish proteins that could be translocated by the Sec-dependent pathway (12). This analysis did not identify proteins translocated by alternative, less common secretory pathways.

Orthologous genes in *C. bescii*, *C. obsidiansis*, and *C. saccharolyticus* were identified by comparing predicted protein sequences using the BLASTClust program (ver. 2.2.21; Ilya Dondoshansky, National Center for Biotechnology Information) from the BLAST package with default parameters. Custom Perl scripts were used to sort and interpret the output.

RESULTS

Growth characterization. *C. bescii* and *C. obsidiansis* cells were grown in batch fermentations using crystalline cellulose. Although both fermentors were inoculated with similar cell densities, planktonic cell counts showed a lag phase of about 4 h in *C. bescii*, while *C. obsidiansis* displayed immediate entry into the exponential growth phase (Fig. 1A). *C. obsidiansis* reached a maximal cell density of approximately 10^9 /ml after 16 h, and *C. bescii* did so after 20 h, after which the cell densities of both organisms remained stationary. Pellet protein concentrations were below the detection limits of the Lowry assay until 4 h after inoculation (Fig. 1B). In contrast to the planktonic cell densities, pellet protein concentrations remained similar until 12 h, when the *C. bescii* protein levels exceeded those of *C. obsidiansis* by 10 to 30 μ g/ml. Protein concentrations in the cell-free culture supernatant similarly diverged after 20 h. The highest supernatant protein concentrations were measured at the end of the fermentation (48 h), with 47 ± 2 μ g/ml for *C. bescii* and 37 ± 3 μ g/ml for *C. obsidiansis*.

Extracellular proteome characterization. Average values for SpC, nonredundant peptides (NR pep), and nonredundant proteins (NR pro) from cell-free supernatant time course samples of both cultures followed similar trends, although they were found to be slightly lower in *C. obsidiansis* than in *C. bescii* (Table 1, see Table S6 in the supplemental material for both replicates). At a false-discovery rate of approximately 5%, 494 and 418 nonredundant proteins were identified in *C. bescii* and *C. obsidiansis*, respectively. The numbers of identified proteins in samples taken at 0 and 4 h were too low to be normalized together with the subsequent time points. In both organisms, the NR pro values were marginally reduced by spectral rebalancing to provide corrected nonredundant protein (NR pro corr) values. SignalP analysis predicted secretory signal peptide sequences in only one-third of the identified proteins (Table 1). NSAF values, derived from SpC, were calculated for each identified protein and used as a quantitative measure of protein abundance. Based on these NSAF values, proteins with confidently predicted secretion sequences make up 60 to 70% of the sample composition at each time point (see Tables S1 and S2 in the supplemental material). Although protein diversity is high among putative nonsecreted

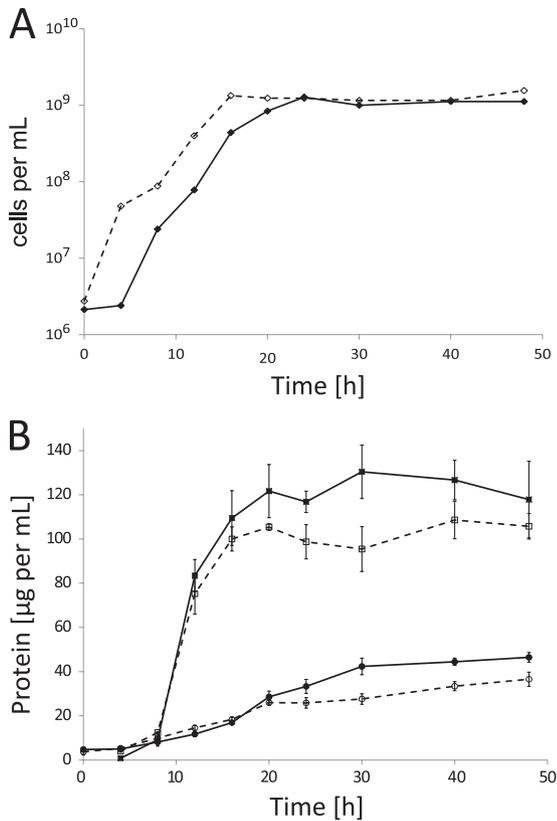


FIG. 1. Characterization of cell growth and protein production. *C. bescii* (solid symbols) and *C. obsidiensis* (open symbols) were grown in a 4-liter batch fermentation on 5 g/liter Avicel crystalline cellulose. Samples removed at each time point were analyzed to assess growth. (A) Planktonic cell counts of both organisms. (B) Pellet (squares) and supernatant (circles) protein analyses showed higher protein levels in the *C. bescii* culture. Error bars indicate the standard deviation among three replicates. Extracellular protein levels continued to increase after the end of the exponential growth phase at 20 h.

proteins, they contributed only marginally to the protein abundance in the samples.

Identification of significantly changing proteins over time.

NSAF values were transformed with the natural logarithm and compared across all time points to identify proteins that exhibit abundance changes over time. Heat maps show the abundance patterns of significantly changing proteins identified by one-way ANOVA, organized into hierarchical clusters based on

abundance trending (Fig. 2). For *C. bescii*, the abundances of 57 (12%) identified proteins changed significantly over time (Fig. 2A), similar to 53 (13%) proteins for *C. obsidiensis* (Fig. 2B). Sample clustering (Fig. 2, bottom) showed a sequential relationship among samples, with two major groups consisting of samples removed 8 to 20 h or 24 to 48 h postinoculation. These time bins represented the exponential and stationary growth phases.

The proteins identified in the cell-free culture supernatant were categorized into different groups. The distributions of protein groups within the clusters were very similar in the two organisms. Most striking was a cluster of extracellular binding proteins (EBPs) and glycosidases (GHs; glycoside hydrolases according to the carbohydrate-active enzyme [CAZy] database) that substantially increased in abundance after 20 h. The GHs in this group are found in close proximity in both organisms' genomes, while the EBPs are distributed across the genome. Another group of EBPs and GHs showed maximum abundance from 12 to 30 h. This cluster also contains a high density of extracellular unknown proteins with uncharacterized function that contain predicted secretory signal peptides. Proteins without signal sequences were categorized as putative cytosolic proteins (PCPs) and are distributed throughout the clusters. Most PCPs clustered with a group of proteins that were highly abundant in the first 16 h but exhibited a subsequent decrease thereafter. This cluster also contained two putative pectate lyase proteins common to both organisms, as well as proteins containing S-layer homology (SLH) domains.

Comparison with a table of orthologous genes shows that 40% of the significantly changing proteins in both organisms are orthologs that display the same abundance pattern over time, while 30% are unique to the respective organism. The other 30% of these proteins have orthologs that did not change significantly in the other organism (see Tables S1 and S2 in the supplemental material). It is important to note that although Fig. 2 displays abundance trends, it does not provide details of absolute protein abundance. For example, the *C. bescii* Athe_0597 and *C. obsidiensis* COB47_0549 EBPs were almost 2 orders of magnitude more abundant than the average of all of the identified proteins, followed by the PCPs Athe_1664 and COB47_0918, with about 40-fold higher abundance, and the most prominent GHs Athe_1867 and COB47_1673, with 20-fold higher abundance. Athe_1664 and COB47_0918 are part of a group of flagellum-associated proteins that were identified in both organisms, including 11 proteins in *C. bescii* and 13

TABLE 1. Extracellular proteome characterization for *C. bescii* and *C. obsidiensis* cell-free culture supernatants by LC-MS/MS

Time (h)	<i>C. bescii</i>					<i>C. obsidiensis</i>				
	SpC	NR pep	NR pro	NR pro corr	SigP ^a	SpC	NR pep	NR pro	NR pro corr	SigP ^a
8	5,529	2,142	216	197	69	7,249	1,961	218	205	67
12	6,358	2,030	221	201	78	7,531	2,393	229	212	76
16	11,207	4,403	350	325	113	8,583	2,597	234	222	81
20	8,739	2,560	189	172	82	9,672	2,365	186	179	73
24	9,366	2,950	226	204	76	6,587	2,143	191	168	61
30	8,532	1,996	172	155	65	7,683	1,739	163	146	60
40	9,432	2,327	219	205	72	6,780	1,732	159	144	62
48	8,480	2,123	227	213	73	6,060	1,569	162	144	63

^a SigP, proteins predicted to have signal sequences for secretion.

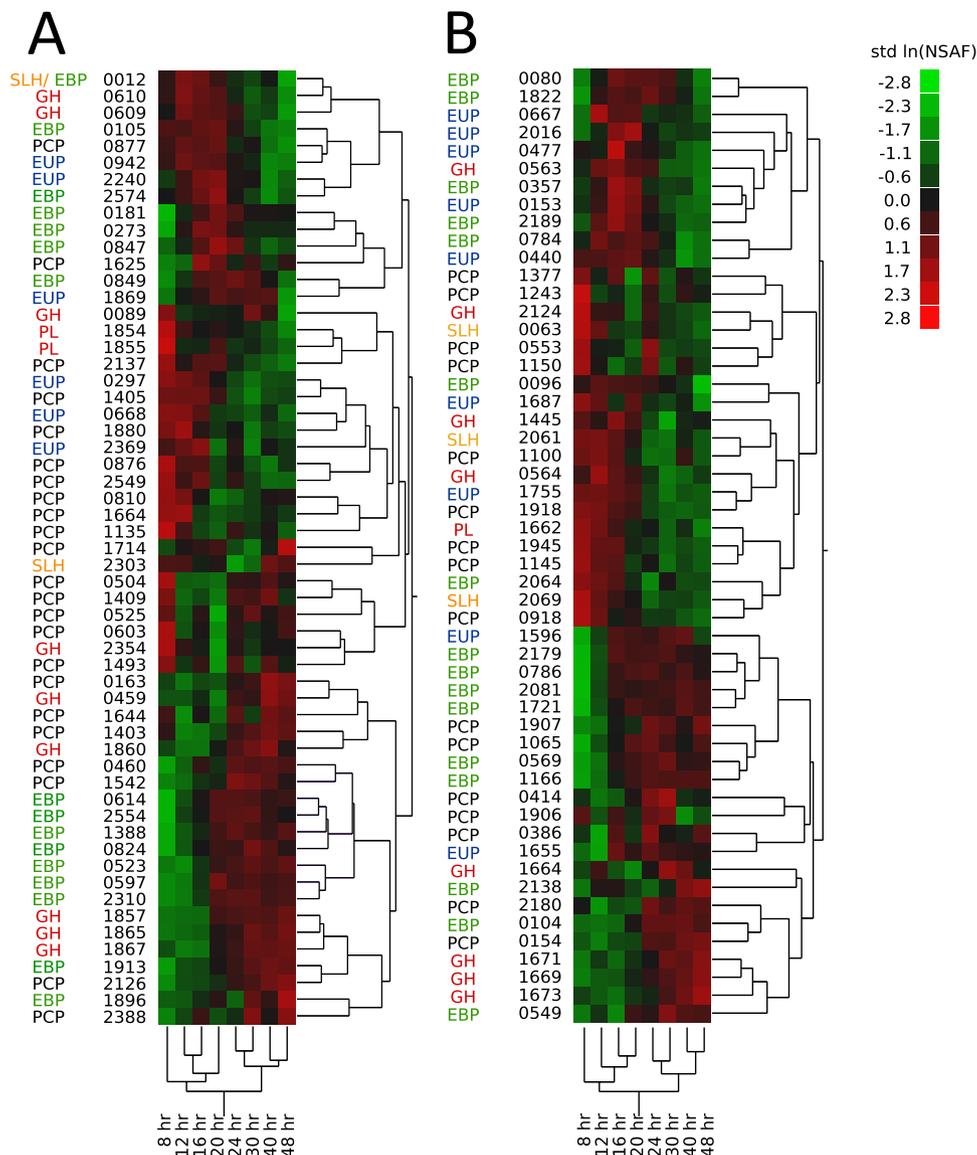


FIG. 2. Heat maps showing proteins whose abundances changed significantly in the cell-free culture supernatant. Sets of LC-MS/MS data were acquired in duplicate for *C. bescii* (A) and *C. obsidiansis* (B) from 8 to 48 h. Peptides were identified using SEQUEST, and NSAF values for each protein were \ln transformed and standardized for one-way ANOVA. Abundances at each time point range from low (green) to high (red). Gene loci are listed without the prefix *Athe_* for *C. bescii* or the prefix *COB47_* for *C. obsidiansis*. Proteins are annotated as EBPs (green), EUP (extracellular unknown proteins; blue), GHs (the CAZy term for glycosidases; red), PCPs (black), PL (pectate lyases; red), or SLH (proteins with SLH domains; orange).

proteins in *C. obsidiansis* (*Athe_1653*, *Athe_1654*, *Athe_1675*, *Athe_1674*, *Athe_2162*, *Athe_2165*, *Athe_2167*, *Athe_2173*, *Athe_2174*, *Athe_2337*, *Athe_2338*, *COB47_0906*, *COB47_0909*, *COB47_0910*, *COB47_0930*, *COB47_0931*, *COB47_1934*, *COB47_1943*, *COB47_1946*, *COB47_1947*, *COB47_1948*, *COB47_1956*, *COB47_1957*, and *COB47_2110*).

Glycosidase abundance and domain composition. NSAF values for selected glycosidases were plotted to compare their relative abundances over time (Fig. 3). These GHs were annotated as putative cellulolytic proteins and exhibited an increasing abundance trend over time. Each had a sum of $(100 \times \text{NSAF}) > 0.5$ across all of the time points, indicating substantial representation in the extracellular fractions. *Athe_1867*,

also known as *CelA* (58), and its ortholog *COB47_1673* were among the most abundant supernatant proteins identified. Both proteins have a theoretical molecular mass of 195 kDa, and they share 95% amino acid sequence identity (Fig. 4). The *Athe_1865*, *Athe_1857*, *COB47_1669*, and *COB47_1671* GHs were about 60% less abundant than the *CelA* homologs (Fig. 3). Although *Athe_1865* and *COB47_1669* are orthologs, both *Athe_1857* and *COB47_1671* are unique to their respective strains (Fig. 4). Another GH unique to *C. bescii*, *Athe_1859*, is approximately an order of magnitude less abundant than *Athe_1867*. The GH orthologs *Athe_1860* and *COB47_1664* contribute marginally to the overall cellulolytic component of each organism's secreted proteins (Fig. 3).

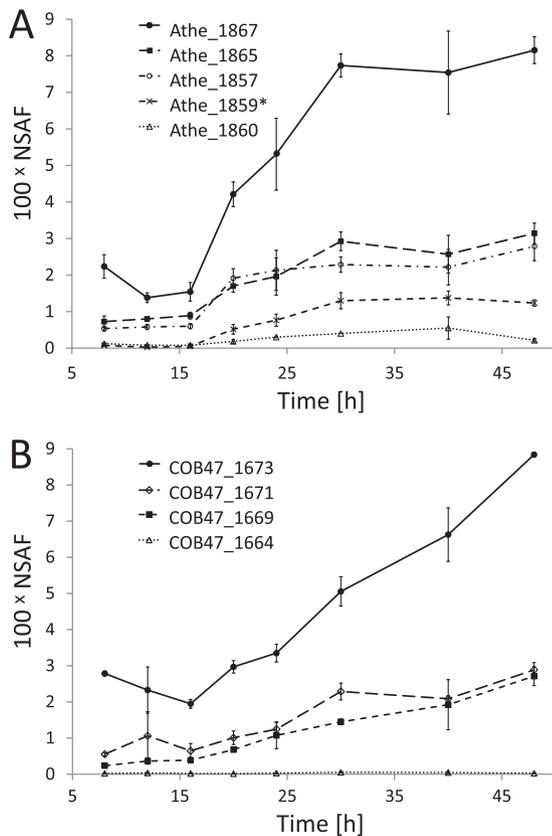


FIG. 3. Glycosidase abundances increased over time. The averages of two proteomic measurements of *C. bescii* (A) and *C. obsidiensis* (B) glycosidase abundances are shown as NSAF values in hundreds. Error bars indicate the standard deviations of duplicate measurements. The signal intensity differs substantially among the enzymes. (A) *C. bescii* supernatant contained five identified glycosidases, including the most abundant protein Athe_1867 (CelA) and the least abundant protein Athe_1860. The Athe_1859 (*) protein abundance did not change significantly during growth but was higher than that of Athe_1860 after 16 h. (B) *C. obsidiensis* cell-free culture supernatant contained four glycosidases, including the most abundant protein COB47_1673 (CelA) and the relatively minor protein COB47_1664.

Cellulase activity correlates with MS data. Time course samples were also analyzed for cellulase activity. Cell-free culture supernatant was incubated with CMC, and the rate of reducing sugar formation was measured. Specific activity in the sample was plotted alongside the summed percent NSAF values of recognized GHs (Fig. 5). Both organisms show a close correlation between activity and GH abundance measurements, with 97% for *C. bescii* and 92% for *C. obsidiensis*, although GH abundance briefly lags activity in *C. obsidiensis*.

Selective cellulase enrichment. After 48 h, when cellulase abundance and activity were found to be highest, cells were harvested by centrifugation and culture supernatant was filtered to obtain a cell-free supernatant fraction (SN). The SN was concentrated 5-fold via tangential-flow filtration, resulting in a retentate that retained roughly 80% of the total protein for *C. bescii* and 54% for *C. obsidiensis* (Table 2). Cellulases were selectively enriched from the TFF fraction via affinity digestion, leaving behind proteins in the supernatant that theoretically do not bind to cellulose. In order to compare the fractions

in the two organisms, the protein concentration that converted 4% of the CMC or Avicel PH-101 substrate to reducing sugars in 60 min was determined (designated CMC U or AV U). The number of CMC U per milligram of protein increased 30-fold for both organisms in TFF. In AD, the activity increased about 500-fold for *C. bescii* and 1,000-fold for *C. obsidiensis*. This indicates that the *C. bescii* AD fraction (15.57 U/mg) is only half as active as the one obtained from *C. obsidiensis* (32.9 U/mg) (Table 2). Insoluble Avicel is a more complex and recalcitrant substrate, and protein concentrations in SN, TFF, and ADSN samples were too low to achieve 4% conversion in 60 min, so that activity could not be determined. The number of AD fraction AV U, however, followed the same trend as observed in the number of CMC U, with AD more than twice as active in *C. obsidiensis* as in *C. bescii*, 0.89 versus 0.39 U/mg, respectively (Table 2).

Zymogram. The TFF, AD, and ADSN fractions were separated on two SDS-PAGE gradient gels under the same conditions. One gel was stained with Coomassie blue dye to visualize all proteins, and the other one was stained for cellulase activity in a zymogram (Fig. 6). The zymograms showed a staggered unstained band pattern of active CMCase with apparent masses of 60 to 250 kDa for both organisms. A separate band at 50 kDa was more distinct in *C. bescii* proteins. The same pattern was retained in the AD fractions. The ADSN fractions showed active bands near 90 kDa at substantially reduced intensities. Additionally, *C. obsidiensis* lacked an active 60-kDa protein in all three fractions that appeared in the *C. bescii* TFF and AD fractions. The apparent molecular masses of the glycosidases could not be matched to the theoretical masses of predicted proteins. Proteins from several gel bands were extracted and analyzed by mass spectrometry; however, the protein diversity was found to be very high and dominated by the most abundant proteins in the samples (data not shown). Larger apparent molecular masses could be due to heterogeneous protein glycosylation, while smaller ones could have resulted from proteolytic degradation of the large, multidomain proteins, as suggested previously (58).

Proteomic analysis of cellulase-enriched fractions. LC-MS/MS measurements of abundant proteins in the TFF, AD, and ADSN fractions are shown in Table 3 for *C. bescii* and in Table 4 for *C. obsidiensis*. Tables S3 and S4 in the supplemental material present the complete set of identified proteins. In both organisms, the number of identified proteins in the AD fraction was about 6 to 7 times lower than in the TFF and ADSN fractions, which have similar numbers of identified proteins. Conversely, abundance values for the GHs increased about 10-fold compared to those in the TFF fraction and were depleted 6-fold in the remaining ADSN fraction. Besides the GHs, AD fractions contained several proteins without carbohydrate binding domains, such as the EBPs Athe_0597 (COB47_0549), Athe_1896 (COB47_1704), and COB47_1166; the flagellin domain protein Athe_1664 (COB47_0918); and the SLH domain protein COB47_2069. However, none of these proteins had been selectively enriched, because the AD/TFF ratio was <1 and the ADSN/TFF ratio was ≥1. The high abundance of these proteins in the TFF fraction suggests carryover through nonspecific binding.

Characterization of cellulase-enriched fractions. The cellulase-enriched AD fractions were investigated to determine the

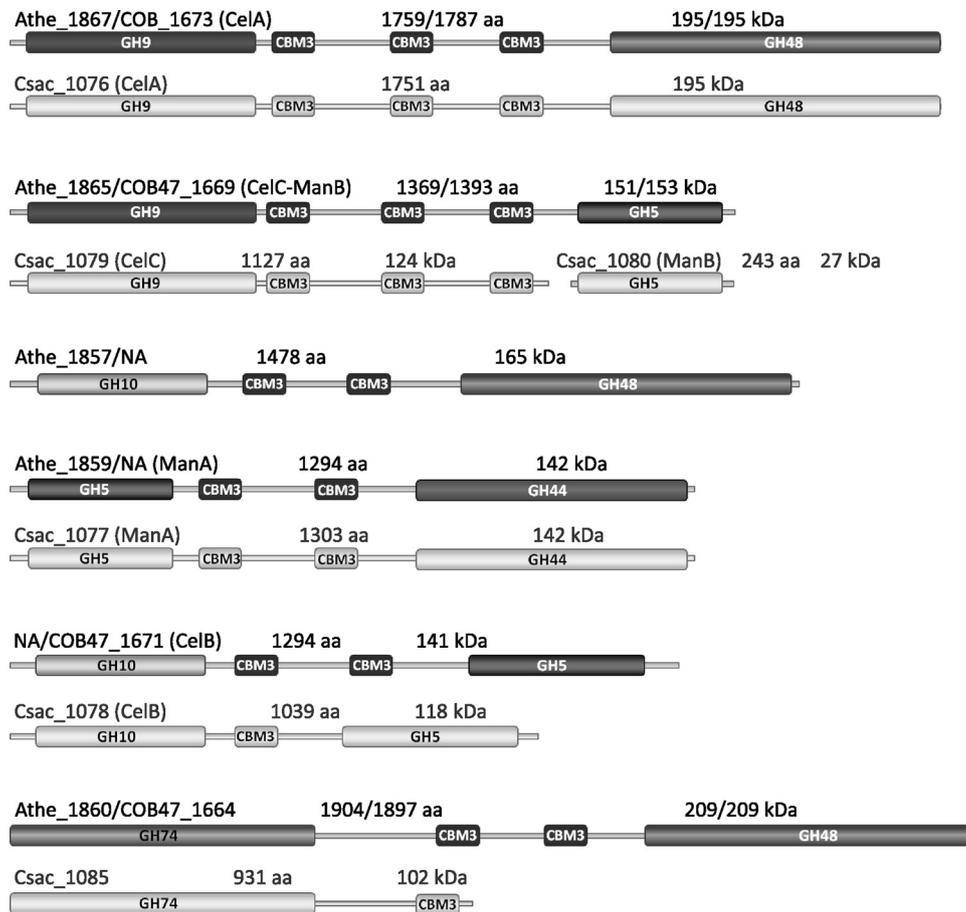


FIG. 4. Glycosidase domain composition. The Pfam, CAZy, and nonredundant protein databases were used to identify functional domains within the multidomain enzymes. These domains are labeled as follows according to CAZy families: GH, glycosidase; CBM3, carbohydrate-binding module family 3. The Athe_1857 and Athe_1859 proteins have no orthologs in *C. obsidiansis*, and COB47_1671 has no *C. bescii* ortholog. aa, amino acids.

optimum reaction temperature and pH. Cellulase mixtures obtained from both *C. bescii* and *C. obsidiansis* hydrolyzed CMC optimally at 85°C at pH 5 (Fig. 7). Thermostability was determined by preincubating the protein fractions at 75, 85,

and 95°C and then measuring CMCase activity at 80°C. Figure 7C and D show that the activity of both cellulase mixtures increased slightly following preincubation at 75°C. At 85°C, the activity decreased to 80% after 30 min of preincubation but

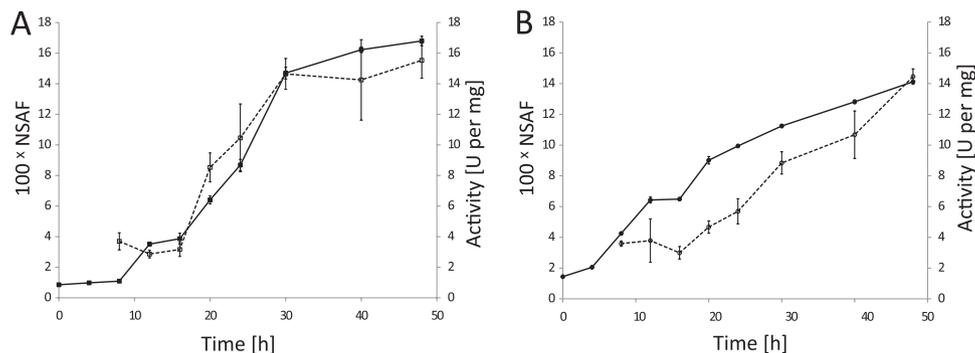


FIG. 5. Total glycosidase abundance correlates with cellulase activity during the course of fermentation. CMCCase activity of the cell-free supernatant was measured for *C. bescii* (A) and *C. obsidiansis* (B) from 0 to 48 h after inoculation. Enzyme activity in units per milligram of total supernatant protein (solid symbols, solid lines) was plotted together with the sum of NSAF values (in hundreds) for the glycosidases identified in Fig. 3 (open symbols, dashed lines). The error bars indicate standard deviations for each value. Abundance and activity increases correlate for both cultures; however, the correlation is closer for *C. bescii* data than for *C. obsidiansis* data, suggesting that an active component may not have been included in the summed NSAF values.

TABLE 2. *C. bescii* and *C. obsidiensis* glycosidase enrichment at 48 h

Organism and fraction	Vol (ml)	Protein concn (µg/ml)	Total protein amt (mg)	CMC U/ml	CMC U/mg	AV ^a U/ml	AV U/mg
<i>C. bescii</i>							
SN ^b	550	38.4	21.1	0.58	0.02	<0.046 ^c	ND ^d
TFF	110	152.6	16.8	4.92	0.03	<0.046	ND
ADSN	110	82.8	9.1	0.99	0.01	<0.046	ND
AD	9	900.7	8.1	14.02	15.57	0.35	0.39
<i>C. obsidiensis</i>							
SN	1,180	28.5	33.6	0.55	0.02	<0.046	ND
TFF	200	90.6	18.1	3.08	0.03	<0.046	ND
ADSN	200	73.5	14.7	0.7	0.01	<0.046	ND
AD	10	827.3	8.3	27.2	32.9	0.74	0.89

^a AV, Avicel PH-101.
^b SN, culture supernatant after 48 h.
^c Below detection limit.
^d ND, not determined.

recovered after 60 min to 90% for *C. bescii* and to 100% for *C. obsidiensis*. Both enzyme preparations were highly thermostable: the residual activity after preincubation for 90 min at 95°C was ~20% for *C. bescii* and ~40% for *C. obsidiensis*.

DISCUSSION

LC-MS/MS characterizations of *Caldicellulosiruptor* sp. culture supernatants showed that the extracellular protein composition was diverse, with about 200 different proteins identified for each organism at any given time point and over 400 proteins across all time points. Among these proteins, one-third contained SignalP-predicted signal sequences for Sec-dependent protein translocation, including the most abundant proteins. Additional extracellular proteins, such as flagellar

subunits, are translocated by mechanisms not considered by the SignalP predictions (4, 10).

For evaluation and comparison of the sets of data, NSAF were calculated. A portion of the peptides detected could not be uniquely assigned to a multidomain cellulase protein, due to a high degree of sequence similarity between common domains such as CBM3 modules. Rebalancing the SpC based on unique peptides corrected for the overestimation of shared peptides in MS data analysis. However, this correction could cause an underrepresentation of small proteins that share numerous peptide sequences relative to larger homologs that contain more unique peptides. This problem does not appear to be significant for either of the *Caldicellulosiruptor* proteomes described here.

Based on these corrected NSAF values, significant changes in extracellular protein abundance over time were observed for more than 50 proteins from each bacterium. With a few exceptions, most of the orthologous proteins displayed similar abundance trends over time in both organisms. These changes could be due to differential gene expression, protein turnover, proteolytic degradation, or protein sorption to the substrate. Other proteins appeared unique to one species, reflecting characteristic differences between the strains. These quantitative proteomic measurements refine our model of the cells' genetic potential for carbohydrate degradation to focus on enzymes specifically produced during growth on crystalline cellulose. Comparisons with proteomic surveys of similarly grown microbes provide insight into the disparate biological strategies that cells have evolved to degrade plant biomass.

Surprisingly few glycosidase enzymes were abundantly expressed to degrade crystalline cellulose: only five were revealed in *C. bescii*, and three were revealed in *C. obsidiensis*. Previous analyses identified 88 carbohydrate-active enzymes encoded in the *C. bescii* genome (4, 7, 19) and 87 in that of *C. obsidiensis*, including 53 orthologs found in both bacteria (see Table S5 in the supplemental material). The findings presented here elaborate on two recent studies of *Caldicellulosiruptor* gene expression and protein secretion. Similar to other members of the genus, such as *C. saccharolyticus* and *Caldicellulosiruptor* sp. strain ToK7B.1, the cellulolytic components of the *C. bescii* and *C. obsidiensis* secretomes consist primarily of multifunc-

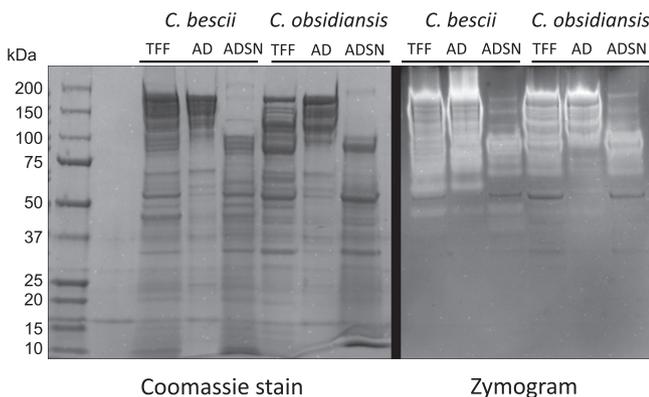


FIG. 6. SDS-PAGE analysis and zymogram of cellulase enrichment fractions. Protein samples from the *C. bescii* and *C. obsidiensis* cellulase enrichments analyzed in Table 2 were separated on two 4 to 20% gradient gels together with a gradient protein ladder of 10 to 250 kDa. One gel was stained with Coomassie brilliant blue dye to visualize all of the proteins (left). The other gel was incubated in a CMC solution and subsequently stained with Congo Red dye (right). Unstained areas indicate CMCase activity. In both organisms, high-molecular-weight proteins from TFF were concentrated in AD and depleted in ADSN, except for an ~90-kDa component. The total protein and activity patterns of the two organisms are similar, with the exception of an ~60-kDa activity band that is unique to *C. bescii*.

TABLE 3. Proteomic analysis of GH enrichment in 48-h supernatant from *C. bescii*

Locus, description, or parameter	AD ^a	TFF ^a	ADSN ^a	AD/TFF ratio	ADSN/TFF ratio	SignalP ^b
Athe_1867, GH9/3(CBM3)/GH48	34.7	3.8	1.2	9.1	0.3	Y
Athe_1865, GH9/3(CBM3)/GH5	15.1	1.1	0.4	13.5	0.4	Y
Athe_1857, GH10/2(CBM3)/GH48	13.0	1.6	0.5	8.3	0.3	Y
Athe_1859, GH5/2(CBM3)/GH44	8.6	0.8	0.0	11.3	0.0	Y
Athe_0597, EBP	5.6	3.4	4.4	1.7	1.3	Y
Athe_1860, GH74/2(CBM3)/GH48	3.6	0.6	0.1	6.0	0.1	Y
Athe_1664, flagellin domain protein	3.0	5.4	8.2	0.6	1.5	N
Athe_1896, EBP	1.4	1.8	2.4	0.7	1.3	Y
Sum for glycosidases	74.9	7.9	2.2	9.5	0.3	
No. of identified proteins	59	363	373			

^a NSAF values in hundreds are shown, and only proteins with values of >1 are shown.

^b Y, yes; N, no.

tional glycosidase enzymes composed of a small number of GH5, GH9, GH10, GH43, GH44, GH48, and GH74 modules interspersed with highly conserved CBM3 domains. Different *Caldicellulosiruptor* species express different permutations of these GH and CBM domains on single polypeptide chains. This construction implies evolution via domain shuffling (14). Enzymes are thought to work together in an orchestrated fashion to efficiently degrade cellulosic substrate (34, 53). It is possible that the multidomain architecture in *Caldicellulosiruptor* is an adaptation to high-temperature environments that exhibit increased enzyme/substrate diffusion rates, as it provides close spatial proximity for synergistic effects, as well as stronger binding to the substrate, due to the multiple CBM modules. In contrast to these multidomain proteins, the main cellulose-degrading systems employed by the bacterium *T. fusca* and the fungus *T. reesei* include six or seven cell-free enzymes consisting of a single catalytic domain together with one carbohydrate-binding module.

Transcriptional and descriptive proteomic analyses of *C. bescii* cells grown on glucose, cellulosic filter paper, and xylan identified enhanced expression of the multidomain glycosidases described below during growth on cellulose (7). Another set of putative xylanase genes was transcribed at significantly higher levels in that study; however, those proteins were not abundant in the present set of quantitative proteomic data. *C. saccharolyticus* cells grown on glucose

secreted a similar set of glycosidases (described below) dominated by the CelA protein (4).

The *C. bescii* and *C. obsidiansis* proteomic and activity measurements reported here detected an increase in glycosidases after the cell cultures entered stationary phase at 20 h, which could be due either to continued expression or to release from the hydrolyzed substrate. These three sets of data suggest that *Caldicellulosiruptor* species express a small group of large, bifunctional, multidomain enzymes with broad substrate specificities to degrade a variety of plant cell wall polymers. One might infer that *Caldicellulosiruptor* cells constitutively express multidomain glycosidases but mono- or oligosaccharides induce higher expression levels. This small set of highly active, generic glycosidases offers a strong contrast to the multienzyme clostridial system, where the expression levels of single-domain GHs were shown to be specifically regulated, depending on the carbon source (16).

The most abundant enzyme in *Caldicellulosiruptor* supernatants was the bifunctional CelA protein Athe_1867 (COB47_1673), which consists of a GH9 domain, three CBM3 domains, and a GH48 domain. The GH9 domain has endo- β -1,4-D-glucanase activity, each family 3 carbohydrate-binding module binds cellulose, and the GH48 domain has processive exoglucanase (cellobiohydrolase) activity. The native CelA protein from *C. bescii* catalyzes the hydrolysis of crystalline cellulose, producing glucose and cellobiose, but has the highest

TABLE 4. Proteomic analysis of GH enrichment in 48-h supernatant from *C. obsidiansis*

Locus, description, or parameter	AD ^a	TFF ^a	ADSN ^a	AD/TFF	ADSN/TFF	SignalP ^b
COB47_1673, GH9/3(CBM3)/GH48	39.7	4.5	2.9	8.8	0.6	Y
COB47_1671, GH10/3(CBM3)/GH5	18.0	1.9	0.8	9.5	0.4	Y
COB47_1669, GH9/2(CBM3)/GH5	13.5	1.0	0.5	13.7	0.5	Y
COB47_0549, EBP	9.2	5.1	10.5	1.8	2.0	Y
COB47_0918, flagellin domain protein	2.1	3.9	7.3	0.5	1.8	N
COB47_2069, SLH-containing protein	1.6	2.1	2.1	0.8	1.0	Y
COB47_1166, EBP	1.3	1.4	1.7	0.9	1.2	Y
COB47_1704, EBP	1.3	1.8	2.8	0.7	1.5	Y
Sum for glycosidases	71.2	7.4	4.2	9.7	0.6	
No. of identified proteins	49	335	272			

^a NSAF values in hundreds are shown, and only proteins with values of >1 are shown.

^b Y, yes; N, no.

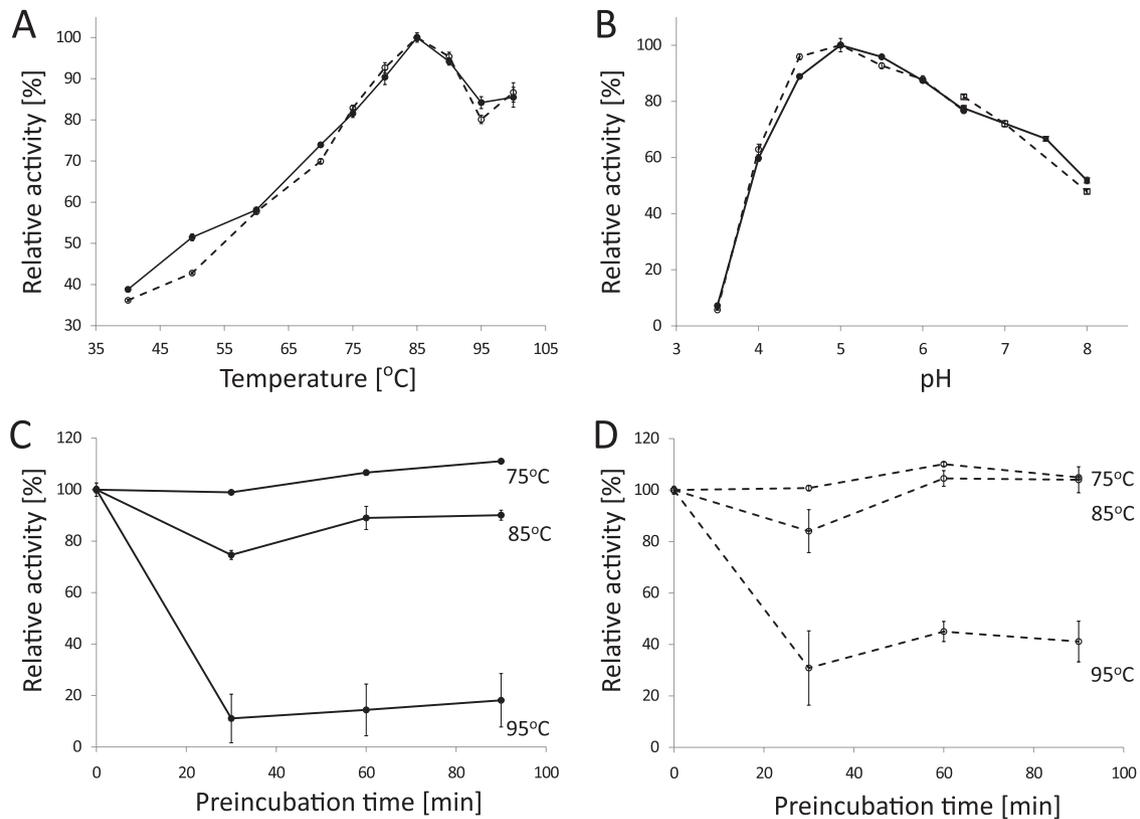


FIG. 7. Biochemical characterization of glycosidase preparations. The protein fractions isolated by affinity digestion of the cell-free culture supernatants of *C. bescii* (solid symbols, solid lines) and *C. obsidiansis* (open symbols, dashed lines) were examined for optimal reaction parameters and thermostability in triplicate CMCase assays at 80°C (B to D) or at various temperatures of 40 to 100°C (A). To determine the pH optimum, values were measured at pHs of 3.5 to 6.5 in 50 mM citrate buffer (circles) and at pHs of 6.5 to 8 in 50 mM potassium phosphate buffer (squares). The thermostability of the *C. bescii* (C) and *C. obsidiansis* (D) enzyme mixtures was measured by incubating the protein samples for 30, 60, or 90 min at 75, 85, or 95°C before starting the CMCase assay by substrate addition. Both preparations showed slightly increased activity upon preincubation for 30 and 60 min. The *C. bescii* enzyme mixture retained 10 to 20% of its activity upon preincubation for 90 min at 95°C, while the *C. obsidiansis* mixture retained up to 40%.

activity on CMC (Gluβ-1→4 Glu), β-glucan (Gluβ-1→3,4 Glu), and xylan (Xylβ-1→4 Xyl) (58). A recombinant CelA protein from *C. saccharolyticus* did not exhibit xylanase activity (46).

Most cellulolytic organisms express extracellular or cell surface-associated GH9 or GH48 protein for crystalline cellulose hydrolysis. Proteomic analysis of the *C. thermocellum* cellulosome identified the GH48 (CelS) exoglucanase as the most abundant enzyme among 14 other less abundant GH9 domain-containing proteins when cultures were grown on crystalline cellulose (25, 38, 51). Cellulose-grown *T. fusca* also expresses a GH48 (Cel48A) exoglucanase, with concurrent upregulation of several endoglucanases, including a GH9 (Cel9B) protein (1). These findings demonstrate the synergistic importance of the GH9 and GH48 protein domains in the degradation of crystalline cellulose. Perhaps an exception to the rule, *S. degradans* proteins lack any detectible GH48 domain. This organism employs mostly endoglucanases, including a GH9 (Cel9B) protein, for the degradation of crystalline cellulose (45).

The CelC-ManB bifunctional glycosidase Athe_1865 (COB47_1669) was the second most abundant protein identified by LC-MS/MS. These orthologs contain a GH9 domain, three CBM3 domains, and a GH5 domain. In *C. saccharolyti-*

cus, the orthologous domains can be found in different open reading frames. Csac_1079 (A4XIF8, CelC) encodes a GH9 domain with three CBM3 domains and has been shown to have endo-1,4-β-D-glucanase activity, while Csac_1080 (A4XIF9, ManB) is a β-mannanase homolog of the GH5 domain (15, 33). In addition to its role in cellulose hydrolysis, the bifunctional protein may facilitate hemicellulose degradation by catalyzing glucomannan hydrolysis. GH5 enzymes seem to be essential for cellulose degradation in *T. reesei* (Cel5A) (20), *T. fusca* (Cel5A) (1), and *S. degradans* (Cel5I and Cel5J) (45), where all three are annotated as endoglucanases.

C. bescii and *C. obsidiansis* both express GH10 domains, but in nonorthologous multidomain proteins. Athe_1857 contains GH10 and GH48 domains, while in COB47_1671, the GH10 domain is associated with another GH5 module. GH10 domains often confer endo-1,4-β-D-xylanase activity, although a homologous domain from *C. saccharolyticus* was reported to have exo-β-1,4-D-glucanase (cellobiohydrolase) activity (39). The COB47_1671 protein is homologous to CelB from *C. saccharolyticus* (Csac_1078, A4XIF7). This enzyme was recently heterologously expressed, purified, and shown to have the highest hydrolytic activity on xylan and β-glucan, followed by glucomannan and CMC. When expressed separately, the

GH5 and GH10 domains both independently exhibited the same broad substrate specificity, although at decreased rates of hydrolysis. Combining the single enzymes did not completely restore the activity of the full-length version, demonstrating the synergistic effects of multidomain proteins (48). Nevertheless, several cellulolytic bacteria express single GH10 domain proteins. *C. thermocellum* has four xylanolytic GH10 domains dispersed over several proteins, although their expression seems to be downregulated on crystalline cellulose (16, 38). In contrast, *T. fusca* expresses significant levels of GH10 xylanase together with other hemicellulases when grown on cellulose (1), while *T. reesii* and *S. degradans* do not encode any GH10 domain proteins at all (20, 45).

The *C. bescii* protein Athe_1859 (ManA) was less abundant than CelC-ManB and has no ortholog in *C. obsidiansis*. It contains an N-terminal GH5 β -mannanase, two CBM3 domains, and a GH44 endo- β -1,4-glucanase domain. The orthologous *C. saccharolyticus* protein (Csac_1077, A4XIF6) exhibited endo- β -1,4 xylanase activity (15, 27).

The orthologs Athe_1860 and COB47_1664 are present at very low levels in the supernatant, as well as in the cellulase-enriched fractions. These proteins contain an amino-terminal Asp box repeat commonly found in sialidases, two CBM3 domains, and a GH48 processive cellobiohydrolase domain. *C. saccharolyticus* does not have a full-length ortholog but does encode a protein that contains a similar N-terminal domain (Csac_1085; A4XIG4). The Asp box repeat forms a β -propeller sequence motif that has also been found in GH74 proteins like the xyloglucanase Xgh74A from *C. thermocellum* (29). The natural function of these proteins may be to facilitate glucoxy-lan hydrolysis in hemicellulose.

Besides the highly abundant glycosidases, other proteins that were found in the secretome include putative pectate lyases: Athe_1854 (COB47_1662) and Athe_1855. These genes are found to be in close genomic proximity to the multifunctional glycosidase genes; however, their abundance patterns do not cluster with the identified GHs. Rather, they exhibited a reverse trend, with higher abundances in the early growth stages that decreased over time, consistent with previous transcriptional analysis of cells grown on cellulose versus glucose (7). Other GHs showed a similar proteomic trend, although their transcript abundance was higher during growth on cellulose: those proteins declining in abundance included a putative pul-lulanase, Athe_0609 (OB47_0563); a putative α -amylase, Athe_0610 (COB47_0564); a putative β -xylosidase with a GH3 domain, Athe_2354 (COB47_2124); the endo- β -1,4-xylanase Athe_0089, and the GH73-containing putative peptidoglycan hydrolase Athe_1080 (COB47_1445) (see Tables S1 and S2 in the supplemental material). Since the abundance of these proteins decreased over time, it is possible that their expression was triggered either by impurities in the cellulose substrate or by yeast extract components in the medium.

Cellulase protein stability is valuable in enzyme mixtures for large-scale cellulose saccharification; therefore, it is noteworthy that the cellulase-enriched fractions from both organisms had maximum CMCase activity at pH 5 and 85°C, exceeding the thermoactivity of commercial cellulase preparations (35). In a previous study, a similar pH optimum for CMCase activity of purified *C. bescii* CelA was measured, while the hydrolytic activity continued increasing at elevated temperatures, even at

95 to 100°C (58). Enzyme preparations from both *Caldicellulosiruptor* spp. appeared equally thermostable following preincubation at various temperatures, although *C. obsidiansis* proteins retained slightly more stability after preincubation at 95°C. A similar study with *Caldicellulosiruptor lactoaceticus* culture supernatant reported optimal CMCase activity at 80°C and pH 6. However, the proteins' thermostability was lower: 60% activity remained after 90 min preincubation at 70°C, 50% at 80°C, and about 30% at 90°C (32). Significant differences in the assay methods used to measure cellulase activity prevent comparison of *Caldicellulosiruptor* AD activity data measured at 80°C with results from previous studies. However, commercial cellulase mixtures that were also assayed according to the IUPAC standard show a slightly lower range (5 to 25 CMC U/mg at 50°C [35]) of specific activities.

Most of the abundant glycosidase proteins are expressed from loci situated in islands on the chromosomes of *C. bescii*, *C. obsidiansis*, and *C. saccharolyticus*. These 61- to 80-kbp islands are hot spots of genetic recombination, insertion, and deletion events that produced the diversity of multidomain glycosidases and accessory enzymes observed in these bacteria. These islands include genes COB47_1657 to COB47_1693, Athe_1845 to Athe_1886, or Csac_1060 to Csac_1108 encoding glycosidase integrases, transposases, and pilin or prophage elements, as well as numerous hypothetical proteins. However, there are no transporter components in close genomic proximity.

ABC transporters are the primary mode of monosaccharide and oligosaccharide uptake by *Caldicellulosiruptor* cells. The most abundant extracellular proteins identified in this study, EBPs, bind their cognate substrates and deliver them to the membrane-bound components of ABC transport systems. The *C. saccharolyticus* genome contains at least 177 ABC transporter genes, including components of 25 putative sugar transporters (49). *C. bescii* and *C. obsidiansis* share orthologs for 14 of these systems, which include the most abundant EBP subunits detected in these proteomics analyses. In both organisms, the abundance levels of one group of EBPs were low at the beginning of the fermentations, peaked in the exponential growth stage, and decreased in the stationary phase. Among these EBPs are COB47_0357 (orthologous to Athe_0399 and Csac_0440) and COB47_0096 (orthologous to Athe_0105 and Csac_2506). Expression of the orthologous *C. saccharolyticus* genes was upregulated during growth on the monosaccharides xylose, glucose, fructose, and galactose (49). Like the abundance of the identified hemicellulases, that of these proteins decreased over time and therefore they could also be induced by impurities in the cellulose substrate or yeast extract components and/or be derived from a basal level of constitutive expression.

Another group of EBPs increased over time in both organisms, with the highest levels observed in the late stationary phase; these proteins clustered and trended with the major GHs. These EBPs included COB47_0549 (orthologous to Athe_0597 and Csac_0681), which was predicted to transport xyloglucans (49). Due to its high abundance in cellulose-grown cultures, it most likely has a substrate specificity for cellooligosaccharides. COB47_0569 (orthologous to Athe_0614 and Csac_0692), on the other hand, may bind a broad range of monosaccharides (49). These EBPs could work synergistically

with the major GHs of each species, binding carbohydrates released by the multifunctional cellulases.

The flagellin proteins in both species, Athe_1664 and COB47_0918, decreased in abundance during growth but remained the second most abundant extracellular proteins. Furthermore, these proteins were abundant in AD fractions. Less abundant components of the flagellar apparatus were also identified in the supernatant fractions. Electron micrographs of *C. bescii* and *C. obsidiansis* did not indicate flagella, and motility has not been observed in these organisms. However, their genomes contain conserved clusters of flagellar motor and filament assembly genes. In *Clostridium difficile*, the FlhC and FlhD proteins mediate cellular attachment to mucus (44), suggesting a potential role for flagella in cellulose attachment, as well as possible motility.

Conclusions. *Caldicellulosiruptor* spp. have proven to be highly versatile cellulose- and hemicellulose-degrading thermophiles. This quantitative proteomic analysis identified the strategy *C. bescii* and *C. obsidiansis* use to degrade crystalline cellulose, i.e., by secreting a small number of cellulose-binding, multifunctional glycosidases dominated by the CelA protein. A few EBPs are expressed at high levels to sequester released glucose and small glucans for cellular metabolism. Predictions of the hemicellulose-degradative potential of the same glycosidases expressed during growth on purified cellulose suggests that future studies of differential protein expression in cells grown with diverse polymeric substrates will help differentiate the cells' general and specific responses to lignocellulosic substrates. Closely related species of the same genus differ in their secreted enzyme yields, enzymes, and catalytic efficiencies. Therefore, the assumption that genomic similarity equates to physiological similarity can lead to a missed opportunity to garner biotechnological potential.

ACKNOWLEDGMENTS

We thank Barbara Klippel, Scott Hamilton-Brehm, Babu Raman, and James Elkins for helpful discussions.

This study was funded by the BioEnergy Science Center, a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. Oak Ridge National Laboratory is managed by and this report has been authored by University of Tennessee-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy.

REFERENCES

- Adav, S. S., C. S. Ng, M. Arulmani, and S. K. Sze. 2010. Quantitative iTRAQ secretome analysis of cellulolytic *Thermobifida fusca*. *J. Proteome Res.* **9**:3016–3024.
- Adney, W. S., and J. O. Baker. 1996. Measurement of cellulase activities. National Renewable Energy Laboratory, Washington, DC.
- Aebersold, R., and M. Mann. 2003. Mass spectrometry-based proteomics. *Nature* **422**:198–207.
- Andrews, G., D. Lewis, J. Notey, R. Kelly, and D. Muddiman. 2010. Part I: characterization of the extracellular proteome of the extreme thermophile *Caldicellulosiruptor saccharolyticus* by GeLC-MS2. *Anal. Bioanal. Chem.* **398**:377–389.
- Blumer-Schuette, S. E., I. Kataeva, J. Westpheling, M. W. Adams, and R. M. Kelly. 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr. Opin. Biotechnol.* **19**:210–217.
- Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248–254.
- Dam, P., et al. 2011. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. *Nucleic Acids Res.* **39**:3240–3254.
- Decker, S. R., W. S. Adney, E. Jennings, T. B. Vinzant, and M. E. Himmel. 2003. Automated filter paper assay for determination of cellulase activity. *Appl. Biochem. Biotechnol.* **105–108**:689–703.
- Elkins, J. G., et al. 2010. Complete genome sequence of the cellulolytic thermophile *Caldicellulosiruptor obsidiansis* OB47^T. *J. Bacteriol.* **192**:6099–6100.
- Emanuelsson, O., S. Brunak, G. von Heijne, and H. Nielsen. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**:953–971.
- Eng, J. K., A. L. McCormack, and J. R. Yates. 1994. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**:976–989.
- Erickson, B. K., et al. 2010. Computational prediction and experimental validation of signal peptide cleavages in the extracellular proteome of a natural microbial community. *J. Proteome Res.* **9**:2148–2159.
- Giannone, R. J., et al. 2007. Dual-tagging system for the affinity purification of mammalian protein complexes. *Biotechniques* **43**:296–302.
- Gibbs, M. D., et al. 2000. Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile *Caldicellulosiruptor* isolate Tok7B. *1. Curr. Microbiol.* **40**:333–340.
- Gibbs, M. D., D. J. Saul, E. Lüthi, and P. L. Bergquist. 1992. The β -mannanase from *Caldocellum saccharolyticum* is part of a multidomain enzyme. *Appl. Environ. Microbiol.* **58**:3864–3867.
- Gold, N. D., and V. J. J. Martin. 2007. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J. Bacteriol.* **189**:6787–6795.
- Griffin, N. M., et al. 2010. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat. Biotechnol.* **28**:83–89.
- Hamilton-Brehm, S. D., et al. 2010. *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. *Appl. Environ. Microbiol.* **76**:1014–1020.
- Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**(Pt. 2):309–316.
- Herpöel-Gimbert, I., et al. 2008. Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnol. Biofuels* **1**:18.
- Himmel, M. E., et al. 2010. Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels* **1**:325–343.
- Kabel, M. A., M. J. E. C. van der Maarel, G. Klip, A. G. J. Voragen, and H. A. Schols. 2006. Standard assays do not predict efficiency of commercial cellulase preparations towards plant materials. *Biotechnol. Bioeng.* **93**:56–63.
- Kataeva, I. A., et al. 2009. Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium “*Anaerocellum thermophilum*” DSM 6725. *J. Bacteriol.* **191**:3760–3761.
- Keller, M., and R. Hettich. 2009. Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol. Mol. Biol. Rev.* **73**:62–70.
- Kruus, K., W. K. Wang, J. T. Ching, and J. H. D. Wu. 1995. Exoglucanase activities of the recombinant *Clostridium thermocellum* CelS, a major cellulosome component. *J. Bacteriol.* **177**:1641–1644.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
- Lüthi, E., N. B. Jasmal, R. A. Grayling, D. R. Love, and P. L. Bergquist. 1991. Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding for a β -mannanase from the extremely thermophilic bacterium *Caldocellum saccharolyticum*. *Appl. Environ. Microbiol.* **57**:694–700.
- Lynd, L. R., P. J. Weimer, W. H. van Zyl, and I. S. Pretorius. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**:506–577.
- Martinez-Fleites, C., et al. 2006. Crystal structures of *Clostridium thermocellum* xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation. *J. Biol. Chem.* **281**:24922–24933.
- McDonald, W. H., R. Ohi, D. T. Miyamoto, T. J. Mitchison, and J. R. Yates. 2002. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int. J. Mass Spectrom.* **219**:245–251.
- Miller, G. L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**:426–428.
- Mladenovska, Z., I. M. Mathrani, and B. K. Ahring. 1995. Isolation and characterization of *Caldicellulosiruptor lactoaceticus* sp. nov., an extremely thermophilic, cellulolytic, anaerobic bacterium. *Arch. Microbiol.* **163**:223–230.
- Morris, D. D., R. A. Reeves, M. D. Gibbs, D. J. Saul, and P. L. Bergquist. 1995. Correction of the β -mannanase domain of the CelC pseudogene from *Caldocellulosiruptor saccharolyticus* and activity of the gene-product on kraft pulp. *Appl. Environ. Microbiol.* **61**:2262–2269.
- Nidetzky, B., M. Hayn, R. Macarron, and W. Steiner. 1993. Synergism of *Trichoderma reesei* cellulases while degrading different celluloses. *Biotechnol. Lett.* **15**:71–76.
- Nieves, R. A., C. I. Ehrman, W. S. Adney, R. T. Elander, and M. E. Himmel. 1998. Survey and analysis of commercial cellulase preparations suitable for biomass conversion to ethanol. *World J. Microbiol. Biotechnol.* **14**:301–304.

36. Ong, S. E., and M. Mann. 2005. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**:252–262.
37. Paoletti, A. C., et al. 2006. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci. U. S. A.* **103**:18928–18933.
38. Raman, B., et al. 2009. Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS One* **4**:e5271.
39. Saul, D. J., L. C. Williams, D. R. Love, L. W. Chamley, and P. L. Bergquist. 1989. Nucleotide sequence of a gene from *Caldocellum saccharolyticum* encoding for exocellulase and endocellulase activity. *Nucleic Acids Res.* **17**:439.
40. Schwarz, W. H., K. Bronnenmeier, F. Grabnitz, and W. L. Staudenbauer. 1987. Activity staining of cellulases in polyacrylamide gels containing mixed linkage β -glucans. *Anal. Biochem.* **164**:72–77.
41. Skinner, K. A., and T. D. Leathers. 2004. Bacterial contaminants of fuel ethanol production. *J. Ind. Microbiol.* **31**:401–408.
42. Svetlichnyi, V. A., T. P. Svetlichnaya, N. A. Chernykh, and G. A. Zavarzin. 1990. *Anaerocellum thermophilum* gen. nov., sp. nov.—an extremely thermophilic cellulolytic eubacterium isolated from hot-springs in the Valley of Geysers. *Microbiology* **59**:598–604.
43. Tabb, D. L., W. H. McDonald, and J. R. Yates. 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**:21–26.
44. Tasteyre, A., M. C. Barc, A. Collignon, H. Boureau, and T. Karjalainen. 2001. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect. Immun.* **69**:7937–7940.
45. Taylor, L. E., et al. 2006. Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2–40^T. *J. Bacteriol.* **188**:3849–3861.
46. Te'o, V. S. J., D. J. Saul, and P. L. Bergquist. 1995. *celA*, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. *Appl. Microbiol. Biotechnol.* **43**:291–296.
47. van de Werken, H. J., et al. 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl. Environ. Microbiol.* **74**:6720–6729.
48. Vanfossen, A. L., I. Ozdemir, S. L. Zelin, and R. M. Kelly. 17 February 2011, posting date. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Biotechnol. Bioeng.* [Epub ahead of print.] doi:10.1002/bit.23093.
49. Vanfossen, A. L., M. R. A. Verhaart, S. M. W. Kengen, and R. M. Kelly. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. *Appl. Environ. Microbiol.* **75**:7718–7724.
50. Vinzant, T. B., et al. 2001. Fingerprinting *Trichoderma reesei* hydrolases in a commercial cellulase preparation. *Appl. Biochem. Biotechnol.* **91–93**:99–107.
51. Wang, W. K., K. Kruus, and J. H. D. Wu. 1993. Cloning and DNA sequence of the gene coding for *Clostridium thermocellum* cellulase-S₈ (CelS), a major cellulosome component. *J. Bacteriol.* **175**:1293–1302.
52. Washburn, M. P., D. Wolters, and J. R. Yates. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**:242–247.
53. Wilson, D. B. 2004. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *Chem. Rec.* **4**:72–82.
54. Yang, S. J., et al. 2009. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe “*Anaerocellum thermophilum*” DSM 6725. *Appl. Environ. Microbiol.* **75**:4762–4769.
55. Yang, S. J., et al. 2010. Reclassification of ‘*Anaerocellum thermophilum*’ as *Caldicellulosiruptor bescii* strain DSM 6725T sp. nov. *Int. J. Syst. Evol. Microbiol.* **60**(Pt. 9):2011–2015.
56. Zhang, Y., and L. R. Lynd. 2003. Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: development of an enzyme-linked immunosorbent assay-based method with application to *Clostridium thermocellum* batch cultures. *Anal. Chem.* **75**:219–227.
57. Zhang, Y. H., and L. R. Lynd. 2005. Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. *J. Bacteriol.* **187**:99–106.
58. Zverlov, V., S. Mahr, K. Riedel, and K. Bronnenmeier. 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**:457–465.
59. Zybailov, B., et al. 2006. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**:2339–2347.