Label-free quantitative proteomics distinguish the secreted cellulolytic systems of
*Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidianis*

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

\(^1\)Biosciences Division; \(^2\)Chemical Sciences Division; \(^3\)BioEnergy Science Center at Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

\(^4\)Technical Microbiology, Hamburg University of Technology, Kasernenstrasse 12, D-21073 Hamburg, Germany

\(^5\)Department of Microbiology, University of Tennessee, Knoxville, TN 37996.

\*To whom correspondence should be addressed. Mailing address: (DEG) Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6038. Phone: 865-574-0559. (RLH) Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6131. Phone: 865-574-4986

This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.
ABSTRACT
The extremely thermophilic Gram-positive bacteria *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis* efficiently degrade both cellulose and hemicellulose, making 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 families 5, 9, 10, 44 or 48 and family 3 carbohydrate binding modules (CBM3). In addition to their orthologous cellulases, the organisms expressed unique glycosidases with different domain organization: *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.
INTRODUCTION

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 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 (20). 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 systems-level interrogation of various organisms is mass spectrometry, which during the last decade has revolutionized large-scale, high-throughput proteomic characterization of both microbial isolates and communities (3, 23). In particular, proteomics has accelerated the discovery and quantification of cellulose-degrading proteins from both aerobic and anaerobic microorganisms. For example, MS-based proteomic measurements revealed numerous monofunctional glycosidase proteins found in the secretome of the mesophilic fungus *Trichoderma reesei* (19, 50), currently the source for most commercial cellulose preparations (21, 35). The mesophilic marine γ-proteobacterium *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 (4, 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 fusca* secretes a diverse set of monofunctional glycosidases, often fused to carbohydrate binding domains (1, 53). The thermophilic bacterium (60 °C) *Clostridium thermocellum* produces glycosidases in cellulosomal complexes (38), while the extremely thermophilic (70 °C) *Caldicellulosiruptor saccharolyticus* secretes several multidomain, multifunctional cellulases (33).

The use of proteomics for systems-level investigation of cellulose degradation strategies goes beyond that of 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, 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 (15), 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 between 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 obsidiansis* (78 °C) (17). *C. bescii* and *C. obsidiansis* share 97 % 16S rRNA sequence identity and 88 % average nucleotide identity in their genome sequences (17), and both grow on the same set of monomeric and polymeric sugars as carbon sources, including pretreated switchgrass and poplar (17, 49). However, *C. obsidiansis* grows at slightly higher temperatures than *C. bescii*, and only *C. obsidiansis* produces measurable amounts of ethanol during growth on switchgrass or cellulose (17, 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. obsidiansis* (8) and 68 kbp in *C. bescii* (22).

Genomic, proteomic and physiological studies have shown that catabolic enzymes and pathways evolve rapidly through positive selection, differentiating closely related microbial species (27). 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 (19). 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. obsidiansis*, we analyzed their secreted protein profiles over the course of crystalline cellulose fermentations. Label-free quantitative proteomic analysis was performed using 2D LC-MS/MS to identify secreted proteins. Identified
proteins were then quantified 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 carboxymethylcellulase (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 and unique to each organism.

MATERIALS AND METHODS

Controlled cultivation and sampling. Fermentations were performed in BIOSTAT Bplus Twin 5-L jacketed glass fermentors (Sartorius Stedim Biotech) using a 4-L working volume of basal growth medium (17) that contained the following: 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, 1x ATCC trace minerals, 1.25X MTC medium vitamin solution E (57), 0.02% (wt/vol) yeast extract, and 0.5% (w/v) Avicel-PH101 (Fluka). The temperature was controlled at 75 °C for C. bescii and at 78 °C for C. obsidiansis using a Polystat Circulator (Cole-Parmer). Fermentors and media were sparged overnight at 200 rpm with a N₂-CO₂ (80:20) gas mix; 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 3-4 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 pH was controlled at 6.8 using a 10% sodium bicarbonate solution. Replicate samples were taken at the following elapsed fermentation time points (h): 0, 4, 8, 12, 16, 20, 24, 30, 40 and 48. These samples were processed...
as follows: 40-ml samples for proteomics analysis and 10-ml samples for activity assays were centrifuged at 6,500 g for 15 min and the supernatant filtered via 0.22 µm syringe filter with a polyethersulfone (PES) membrane to obtain a cell-free culture supernatant fraction. 1.8-ml aliquots were centrifuged for 5 min at 15,500 g and supernatants as well as 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). 2 L of supernatant were 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 Quixstand tangential flow filtration system equipped with a 5000 molecular weight cutoff hollow fiber filter (GE Healthcare). The supernatant protein concentrate from tangential flow filtration, designated as TFF was used for the cellulase enrichment procedure.

Planktonic cell densities were measured using a Petroff-Hauser 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 (5) 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 and 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-PH105 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). The affinity digestion was performed in a dialysis membrane (SpectraPor, 6-8 kDa cut-off) 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 (25) was performed in 4-20 % Precise protein gels with the BupH Tris-HEPES-SDS running buffer (Thermo-Pierce). Protein bands were stained with Coomassie Blue dye. Carboxymethylcellulase (CMCase) bands were visualized using a modification of the zymogram technique as described previously (40). Gels were incubated in 1 % carboxymethylcellulose (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. 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 pre-incubation, 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 from 40 to 100 °C. For pH assays at 80 °C, 50 mM citrate buffer was used from
pH 3.5 to 6.5 and 50 mM potassium phosphate was used from 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-µL aliquot of sample was added to 500 µL of 3,5-dinitrosalicylic acid (DNS) reagent (2, 30). 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 measured at 540 nm. One unit (U) of activity catalyzed the release of 1 µmol 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 ºC 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, 7).

**Analysis of peptides by mini-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 DTT) at a 1:1 (v/v) ratio, boiled and sonicated with a Branson sonic disruptor (20% amplitude, 2 min; 10 sec pulse; 10 sec pause). Trichloroacetic acid was added to a concentration of 20 % w/v 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 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 enzyme to protein ratio (w/w). As 8M urea inhibits trypsin, samples were diluted to 4 M for an overnight digest followed by dilution to 2 M urea for a 4 h digest. Samples were then was adjusted to 150 mM NaCl, 0.1 % formic acid, and
filtered through a 500-µl 10 kDa cut-off spin column filter (VWR brand). Peptide concentrations were then measured using the BCA assay.

To compare the extracellular protein complement of both organisms at each time point, a 25-µg aliquot of peptides was bomb-loaded onto a biphasic MudPIT back column as described previously (29, 52). Loaded peptides were then washed with solvent A (5 % acetonitrile, 95 % 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 analyzed by a 4-step analysis as previously described (12), with modifications (salt pulses at 10 %, 25 %, 50 %, 100 % of 500 mM ammonium acetate each followed by a 1 h organic gradient to 50 % solvent B), in this study referred to 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 K resolution) was followed by 10 parent ion isolations per 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 (10) with FASTA proteome databases specific to either *C. obsidiansis* (8) or *C. bescii* (5). 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, +3 = 3.5, DelCN 0.08, and 2 unique peptides per protein identification. Prior to semi-quantitative analysis, spectral counts were rebalanced to properly distribute non-unique/shared peptides between their...
potential parent proteins, based solely on their unique spectral counts. NSAF values were then calculated for each protein (59).

NSAF values were imported into the JMP Genomics software package (SAS Institute; ver. 4.1) for statistical analyses (16). 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* (22) and *C. obsidiansis* (8) were submitted to the SignalP server (ver. 3.0) (9) 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 (11). 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 (Ilya Dondoshansky; National Center for Biotechnology Information; ver. 2.2.21) 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 exponential growth phase (Figure 1A). *C. obsidiansis* reached a maximal cell density of approximately $10^9$ cells/ml after 16 h and *C. bescii* after 20 h, after which cell densities remained stationary in both organisms. Pellet protein concentrations were below the detection limits of the Lowry assay until 4 h after inoculation (Figure 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 µg/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 µg/ml for *C. bescii* and 37 ± 3 µg/ml for *C. obsidiansis*.

Extracellular proteome characterization. Average values for spectral counts (SpC), non-redundant peptides (NR pep), and non-redundant proteins (NR pro) from cell-free supernatant time course samples of both cultures follow similar trends, although they were found to be slightly lower in *C. obsidiansis* relative to *C. bescii* (Table 1, See Supplemental Table S6 for both replicates). At a false-discovery rate of approximately 5 %, a total of 494 and 418 non-redundant 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 non-redundant proteins (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 for protein abundance. Based on these NSAF values, proteins with confidently predicted secretion sequences make up 60-70% of the sample composition at each time point (Supplemental Tables S1 and S2). Although protein diversity is high among putative non-secreted 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 in hierarchical clusters based on abundance trending (Figure 2). For *C. bescii*, abundances of 57 (12%) identified proteins changed significantly over time (Figure 2A), similar to 53 (13%) proteins for *C. obsidiansis* (Figure 2B). Sample clustering (Figure 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 post-inoculation. These time bins represented exponential and stationary growth phases.

The proteins identified in the cell-free culture supernatant were categorized into different groups. The distribution of protein groups within the clusters was very similar in both organisms. Most striking was a cluster of extracellular binding proteins (EBP) and glycosidases (GH, glycoside hydrolases according to the 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 (EUP) with uncharacterized function that contain predicted secretory signal peptides.

Proteins without signal sequences were categorized as putative cytosolic proteins (PCP) 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 domains (SLH).

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 (Supplemental Tables S1 and S2). It is important to note that although Figure 2 displays abundance trends, it does not provide details about absolute protein abundance. For example, the C. bescii Athe_0597 and C. obsidiansis COB47_0549 EBPs were almost two orders of magnitude more abundant than the average of all the identified proteins, followed by the PCPs Athe_1664 & COB47_0918 with about forty fold higher abundance, and the most prominent GHs Athe_1867 and COB47_1673 with twenty fold. Athe_1664 and COB47_0918 are part of a group of flagella-associated proteins that were identified in both organisms, including 11 proteins in C. bescii and 13 proteins in C. obsidiansis (Athe_1653, 1654, 1675, 1674, 2162, 2165, 2167 2173, 2174, 2337, 2338 and COB47_0906, 0909, 0910, 0930, 0931, 1934, 1943, 1946, 1947, 1948, 1956, 1957, 2110).

Glycosidase abundance and domain composition. NSAF values for selected glycosidases were plotted to compare their relative abundances over time (Figure 3). These GHs were annotated as putative cellulolytic proteins and exhibited an increasing abundance trend over time. Each had a
sum of (100 × NSAF) > 0.5 across all 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 theoretical molecular masses of 195 kDa, and they share 95% amino acid sequence identity (Figure 4). The Athe_1865, Athe_1857, COB47_1669 and COB47_1671 GHs were about 60% less abundant than the CelA homologs (Figure 3). Although Athe_1865 and COB47_1669 are orthologs, both Athe_1857 and COB47_1671 are unique to their respective strains (Figure 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 (Figure 3).

**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 percentage NSAF values of recognized GHs (Figure 5). Both organisms show a close correlation between activity and GH abundance measurements, with 97% for *C. bescii* and 92% for *C. obsidiansis*, although GH abundance briefly lags activity in *C. obsidiansis*.

**Selective cellulase enrichment.** After 48 h, when cellulase abundance and activity were found to be highest, cells were harvested by centrifugation and culture supernatant filtered to obtain a cell-free supernatant fraction (SN). SN was concentrated 5-fold via tangential flow filtration, resulting in the retentate (TFF) that retained roughly 80% of the total protein for *C. bescii* and 54% for *C. obsidiansis* (Table 2). Cellulases were selectively enriched from the TFF fraction via affinity digestion (AD), leaving behind proteins in the supernatant that theoretically do not bind
to cellulose (ADSN). In order to compare the fractions in the two organisms, the protein concentration that converted 4% of the CMC or Avicel-PH101 substrate to reducing sugars in 60 min was determined (designated as a CMC-U or AV-U). CMC-U's per mg protein increased 30-fold for both organisms in TFF. In AD, activity increased about 500 fold for *C. bescii* and 1000 fold for *C. obsidiansis*. This indicates that the *C. bescii* AD fraction (15.57 U/mg) is only half as active as the one obtained from *C. obsidiansis* (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 AD fraction AV-U, however, followed the same trend as observed in CMC-U with AD more than twice as active in *C. obsidiansis* compared to *C. bescii*, 0.89 vs. 0.39 U/mg respectively (Table 2).

**Zymogram.** 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 (Figure 6). The zymograms showed a staggered unstained band pattern of active CMCases 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. obsidiansis* lacked an active 60-kDa protein in all three fractions that appeared in *C. bescii* TFF and AD. 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, multi-domain proteins as suggested previously (58).

**Proteomic analysis of the 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. obsidiansis*. Supplemental Tables S3 and S4 present the complete set of identified proteins. In both organisms, the number of identified proteins in the AD fraction was about 6-7 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 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 proteins Athe_1664 (COB47_0918); and the SLH domain protein COB47_2069. However, none of these proteins had been selectively enriched, because AD/TFF < 1 and ADSN/TFF ≥ 1. The high abundance of these proteins in the TFF fraction suggests carry over through non-specific binding.

**Characterization of cellulase-enriched fractions.** The cellulase-enriched AD fractions were investigated to determine optimum reaction temperature and pH. Both cellulase mixtures obtained from *C. bescii* and *C. obsidiansis* hydrolyzed CMC optimally at 85 °C at pH 5 (Figure 7). Thermostability was determined by pre-incubating the protein fractions at 75, 85, and 95 °C, and then measuring CMCase activity at 80 °C. Figures 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 recovered after 60 min to 90 % for *C. bescii*. 
and to 100% for *C. obsidiansis*. 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. obsidiansis*.

**DISCUSSION**

LC-MS/MS characterizations of *Caldicellulosiruptor* spp. 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 (33).

For evaluation and comparison of the datasets, normalized spectral abundance factors (NSAF) were calculated. A portion of the detected peptides 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 spectral counts 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 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 in *C. obsidiansis*. Previous analyses identified 88 carbohydrate active enzymes (CAZymes) encoded in the *C. bescii* genome (6, 33) (18) and 87 in *C. obsidiansis*, including 53 orthologs found in both bacteria (Supplemental Table S5). 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. ToK7B.1, the cellulolytic components of the *C. bescii* and *C. obsidiansis* secretomes consist primarily of multifunctional, glycosidase enzymes composed of a small number of GH 5, 9, 10, 43, 44, 48 and 74 modules interspersed with highly conserved CBM family 3 domains. Different *Caldicellulosiruptor* species express different permutations of these GH and CBM domains on single polypeptide chains. This construction implies evolution via domain shuffling (13). Enzymes are thought to work together in an orchestrated fashion to efficiently degrade cellulosic substrate (34, 53). It is possible that the multi-domain 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, cellulose filter paper and xylan identified enhanced expression of the multidomain glycosidases described below during growth on cellulose (6). Another set of putative xylanase genes were transcribed at significantly higher levels in that study; however, those proteins were not abundant in the present quantitative proteomic data set. *C. saccharolyticus* cells grown on glucose secreted a similar set of glycosidases described below, dominated by the CelA protein (33).

*C. bescii* and *C. obsidiansis* proteomics and activity measurements reported here detected an increase in glycosidases after the cell cultures entered stationary phase at 20 h, which could be due to either continued expression or to release from hydrolyzed substrate. These three data sets 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 (15).

The most abundant enzyme in *Caldicellulosiruptor* supernatants was the bifunctional CelA protein, Athe_1867 (COB47_1673) that consists of GH9, three CBM3 and GH48 domains. The 20
GH9 domain has endo-\(\beta\)-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 from *C. bescii* catalyzes the hydrolysis of crystalline cellulose, producing glucose and cellobiose, but was has highest activity on CMC (Glu \(\beta\)-1→4 Glu), \(\beta\)-glucan (Glu \(\beta\)-1→3,4 Glu) and xylan (Xyl \(\beta\)-1→4 Xyl) (58). A recombinant CelA from *C. saccharolyticus* did not exhibit xylanase activity (46).

Most cellulolytic organisms express extracellular or cell surface-associated GH9 or GH48 proteins for crystalline cellulose hydrolysis. Proteomic analysis of the *C. thermocellum* cellulosome identified the GH48 (CelS) exoglucanase to be the most abundant enzyme among 14 other less abundant GH9 domain-containing proteins, when cultures were grown on crystalline cellulose (24, 38, 51). Cellulose-grown *T. fusca* also expresses a GH48 (Cel48A) exoglucanase, with concurrent upregulation of several endoglucanases, including a GH9 (Cel9B) (1). These findings demonstrate the synergistic importance of 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 GH9, three CBM3 and GH5 domains. In *C. saccharolyticus*, the orthologous domains can be found in different open reading frames. Csac_1079 (A4XIF8, CelC) encodes a GH9 with three CBM3 domains and has been shown to have endo-1,4-\(\beta\)-D-glucanase activity, while Csac_1080 (A4XIF9, ManB) is a \(\beta\)-mannanase homolog of the GH5 domain (14, 32). 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) (19), *T. fusca* (Cel5A) (1), and *S. degradans* (Cel5 I &J) (45), where all three are annotated as endoglucanases.

*C. bescii* and *C. obsidiansis* both express GH10 domains, but in non-orthologous multidomain proteins. Athe_1857 contains GH10 and GH48 domains, while in COB47_1671 the GH10 domain is associated with another GH5 module. GH10s 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 (cellbiohydrolase) 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, both GH5 and GH10 domains 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 (15, 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 GH10s at all (19, 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 CBM3s, and GH44
endo-\(\beta\)-1,4-glucanase domains. The orthologous *C. saccharolyticus* protein (Csac_1077, A4XIF6) exhibited endo-\(\beta\)-1,4 xylanase activity (14, 26).

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 celllobiohydrolase 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 \(\beta\)-propeller sequence motif that has also been found in GH74 proteins like the xyloglucanase Xgh74A from *C. thermocellum* (28). The natural function of these proteins may be to facilitate glucoxylan hydrolysis in hemicellulose.

Beside 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 (6). Other GHs showed a similar proteomic trend, although their transcript abundance was higher during growth on cellulose: these proteins declining in abundance included a putative pullulanase Athe_0609 (OB47_0563), a putative \(\alpha\)-amylase Athe_0610 (COB47_0564), a putative \(\beta\)-xylosidases with a GH3 domain, Athe_2354 (COB47_2124), the endo-\(\beta\)-1,4-xylanase Athe_0089, and the GH73 containing, putative peptidoglycan hydrolase Athe_1080 (COB47_1445) (Supplemental Tables S1 and S2). 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 95-100 °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 (31). Significant differences in the assay methods used to measure cellulase activity prevent comparing 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 (5-25 CMC-U/mg at 50 °C (35)) of specific activity.

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 61-80 kbp islands are hotspots of genetic recombination, insertion and deletion events that produced the diversity of multi-domain glycosidases and accessory enzymes observed in these bacteria. These islands
include genes COB47_1657-1693, Athe_1845-1886 or Csac_1060-1108 encoding glycosidases, 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 \textit{Caldicellulosiruptor} cells. The most abundant extracellular proteins identified in this study, extracellular binding proteins (EBPs), bind their cognate substrates and deliver them to the membrane-bound components of ABC transport systems. The \textit{C. saccharolyticus} genome encodes at least 177 ABC transporter genes, including components of 25 putative sugar transporters (49). \textit{C. bescii} and \textit{C. obsidiansis} share orthologs for 14 of these systems, which include the most abundant EBP subunits detected in these proteomics analyses. In both organisms, one group of EBPs displayed low abundance levels in the beginning of the fermentations, peaked in the exponential growth stage, and decreased in 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 \textit{C. saccharolyticus} genes were upregulated during growth on the monosaccharides xylose, glucose, fructose, and galactose (49). Like the identified hemicellulases, the abundance of these proteins decreased over time and therefore could also be induced by impurities in the cellulose substrate, 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 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 cellobiose oligosaccharides. 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 multi-functional 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 FliC and FliD 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 proteomics analysis identified the strategy of *C. bescii* and *C. obsidiansis* to degrade crystalline cellulose by secreting a small number of cellulose-binding, multi-functional 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 hemicellulose degradative potential by 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. To that end, the assumption that genomic similarity equates to physiological similarity can lead to a missed opportunity to garner biotechnological potential.

ACKNOWLEDGEMENTS

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 University of Tennessee-Battelle LLC for the Department of Energy under contract DOE-AC05-00OR22725.
REFERENCES


281:24922-24933.


**FIGURE LEGENDS**

**Figure 1.** Characterization of cell growth and protein production. *C. bescii* (solid symbols) and *C. obsidiansis* (open symbols) were grown in a 4 L batch fermentation on 5 g/L Avicel crystalline cellulose. Samples removed at each time point were analyzed to assess growth. A: Planktonic cell counts for 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 exponential growth phase at 20 h.

**Figure 2.** Heat maps show proteins whose abundances changed significantly in the cell-free culture supernatant. LC-MS/MS data sets 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 analysis. Abundances at each time point range from low (green) to high (red). Gene loci encoding are listed without the prefixes Athe_ for *C. bescii* or COB47_ for *C. obsidiansis*. Proteins are annotated as EBP, extracellular solute binding proteins (green); EUP, extracellular unknown proteins (blue); GH glycoside hydrolases, the CAZy term for glycosidases (red); PCP, putative cytosolic proteins (black); PL, pectate lyases (red) or SLH, proteins with S-layer homology domains (orange).

**Figure 3.** Glycosidase abundances increased over time. The averages of two proteomic measurements of *C. bescii* (A) and *C. obsidiansis* (B) glycosidase abundances are shown as NSAF values multiplied by 100. Error bars indicate the standard deviation between duplicate measurements. The signal intensity differs substantially among the enzymes. A. *C. bescii*
supernatant contained five identified glycosidases, including the most abundant Athe_1867 (CelA) and the least abundant Athe_1860 proteins. *The Athe_1859 protein abundance did not change significantly during growth, but was higher than Athe_1860 after 16 h. C. obsidiansis cell-free culture supernatant contained four glycosidases, including the most abundant COB47_1673 (CelA) and the relatively minor COB47_1664 proteins.

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

Figure 5. Total glycosidase abundance correlates with cellulase activity during the course of fermentation. CMCase 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 mg total supernatant protein (solid symbols, solid lines) was plotted together with the sum of NSAF × 100 values for the glycosidases identified in Figure 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 than C. obsidiansis data, suggesting an active component may not have been included the summed NSAF values.

Figure 6. SDS-PAGE analysis and zymogram of cellulase enrichment fractions. Protein samples from the C. bescii and C. obsidiansis cellulase enrichments analyzed in Table 2 were separated on two 4-20% gradient gels together with a gradient protein ladder from 10 to 250 kDa. One gel
was stained with Coomassie brilliant blue dye to visualize all 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 (tangential flow filtration concentrate) were concentrated in AD (protein fraction isolated in affinity digest) and depleted in ADSN (supernatant from affinity digest), except for a ~90-kDa component. Total protein and activity patterns are similar across the two organisms with the exception of a ~60-kDa activity band that is unique to *C. bescii*.

Figure 7. Biochemical characterization of glycosidase preparations. The protein fractions isolated by affinity digest from 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 a triplicate CMCase assay at 80 °C (B-D) or at varying temperatures from 40 to 100 °C (A). To determine the pH optimum, pH values from 3.5 to 6.5 were measured in 50 mM citrate buffer (circles), pH-values from 6.5 to 8 in 50 mM potassium phosphate buffer (squares). The thermostability of the *C. bescii* (C) and *C. obsidiansis* (D) enzyme mixtures were 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 – 20 % of its activity upon preincubation for 90 min at 95 °C, while the *C. obsidiansis* mixture retained up to 40 %.
TABLE 1. Extracellular proteome characterization for *C. bescii* and *C. obsidiansis* cell-free culture supernatants by LC-MS/MS

<table>
<thead>
<tr>
<th>Time [h]</th>
<th>SpC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NR pep&lt;sup&gt;b&lt;/sup&gt;</th>
<th>NR pro&lt;sup&gt;c&lt;/sup&gt;</th>
<th>NR pro corr&lt;sup&gt;d&lt;/sup&gt;</th>
<th>SigP&lt;sup&gt;e&lt;/sup&gt;</th>
<th>SpC</th>
<th>NR pep</th>
<th>NR pro</th>
<th>NR pro corr</th>
<th>SigP</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>5529</td>
<td>2142</td>
<td>216</td>
<td>197</td>
<td>69</td>
<td>7249</td>
<td>1961</td>
<td>218</td>
<td>205</td>
<td>67</td>
</tr>
<tr>
<td>12</td>
<td>6358</td>
<td>2030</td>
<td>221</td>
<td>201</td>
<td>78</td>
<td>7531</td>
<td>2393</td>
<td>229</td>
<td>212</td>
<td>76</td>
</tr>
<tr>
<td>16</td>
<td>11207</td>
<td>4403</td>
<td>350</td>
<td>325</td>
<td>113</td>
<td>8583</td>
<td>2597</td>
<td>234</td>
<td>222</td>
<td>81</td>
</tr>
<tr>
<td>20</td>
<td>8739</td>
<td>2560</td>
<td>189</td>
<td>172</td>
<td>82</td>
<td>9672</td>
<td>2365</td>
<td>186</td>
<td>179</td>
<td>73</td>
</tr>
<tr>
<td>24</td>
<td>9366</td>
<td>2950</td>
<td>226</td>
<td>204</td>
<td>76</td>
<td>6587</td>
<td>2143</td>
<td>191</td>
<td>168</td>
<td>61</td>
</tr>
<tr>
<td>30</td>
<td>8532</td>
<td>1996</td>
<td>172</td>
<td>155</td>
<td>65</td>
<td>7683</td>
<td>1739</td>
<td>163</td>
<td>146</td>
<td>60</td>
</tr>
<tr>
<td>40</td>
<td>9432</td>
<td>2327</td>
<td>219</td>
<td>205</td>
<td>72</td>
<td>6780</td>
<td>1732</td>
<td>159</td>
<td>144</td>
<td>62</td>
</tr>
<tr>
<td>48</td>
<td>8480</td>
<td>2123</td>
<td>227</td>
<td>213</td>
<td>73</td>
<td>6060</td>
<td>1569</td>
<td>162</td>
<td>144</td>
<td>63</td>
</tr>
</tbody>
</table>

<sup>a</sup> SpC, Spectra counts  
<sup>b</sup> NR pep, non-redundant peptides  
<sup>c</sup> NR pro, non-redundant proteins  
<sup>d</sup> NR pro corr, non-redundant proteins  
<sup>e</sup> SigP, proteins predicted to have signal sequences for secretion
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C. bescii</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN&lt;sup&gt;c&lt;/sup&gt;</td>
<td>550</td>
<td>38.4</td>
<td>21.1</td>
<td>0.58</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFF&lt;sup&gt;d&lt;/sup&gt;</td>
<td>110</td>
<td>152.6</td>
<td>16.8</td>
<td>4.92</td>
<td>0.03</td>
<td>&lt;0.046&lt;sup&gt;g&lt;/sup&gt;</td>
<td>ND&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>ADSN&lt;sup&gt;e&lt;/sup&gt;</td>
<td>110</td>
<td>82.8</td>
<td>9.1</td>
<td>0.99</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD&lt;sup&gt;f&lt;/sup&gt;</td>
<td>9</td>
<td>900.7</td>
<td>8.1</td>
<td>14.02</td>
<td>15.57</td>
<td>0.35</td>
<td>0.39</td>
</tr>
<tr>
<td>C. obsidiansis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>1180</td>
<td>28.5</td>
<td>33.6</td>
<td>0.55</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFF</td>
<td>200</td>
<td>90.6</td>
<td>18.1</td>
<td>3.08</td>
<td>0.03</td>
<td>&lt;0.046</td>
<td>ND</td>
</tr>
<tr>
<td>ADSN</td>
<td>200</td>
<td>73.5</td>
<td>14.7</td>
<td>0.7</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>10</td>
<td>827.3</td>
<td>8.3</td>
<td>27.2</td>
<td>32.9</td>
<td>0.74</td>
<td>0.89</td>
</tr>
</tbody>
</table>

<sup>a</sup> CMC, carboxymethyl-cellulose  
<sup>b</sup> AV, Avicel PH101  
<sup>c</sup> SN, culture supernatant after 48 h  
<sup>d</sup> TFF, tangential flow filtration concentrate  
<sup>e</sup> ADSN, supernatant from affinity digest  
<sup>f</sup> AD, protein fraction isolated in affinity digest  
<sup>g</sup> below detection limit  
<sup>h</sup> ND, not determined
### TABLE 3. Proteomic analysis of GH enrichment in *C. bescii* from time point 48 supernatant

<table>
<thead>
<tr>
<th>Locus and Description</th>
<th>Affinity digest isolate (AD)</th>
<th>Tangential flow filtration concentrate (TFF)</th>
<th>Affinity digest supernatant (ADSN)</th>
<th>AD/TFF</th>
<th>ADSN/TFF</th>
<th>SignalP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athel_1867 GH^{9}/3 (CBM3)/ GH48</td>
<td>34.7</td>
<td>3.8</td>
<td>1.2</td>
<td>9.1</td>
<td>0.3</td>
<td>Y</td>
</tr>
<tr>
<td>Athel_1865 GH9/3 (CBM3)/ GH5</td>
<td>15.1</td>
<td>1.1</td>
<td>0.4</td>
<td>13.5</td>
<td>0.4</td>
<td>Y</td>
</tr>
<tr>
<td>Athel_1857 GH10/2 (CBM3)/ GH48</td>
<td>13.0</td>
<td>1.6</td>
<td>0.5</td>
<td>8.3</td>
<td>0.3</td>
<td>Y</td>
</tr>
<tr>
<td>Athel_1859 GH5/2 (CBM3)/ GH44</td>
<td>8.6</td>
<td>0.8</td>
<td>0.0</td>
<td>11.3</td>
<td>0.0</td>
<td>Y</td>
</tr>
<tr>
<td>Athel_0597 EBP(b)</td>
<td>5.6</td>
<td>3.4</td>
<td>4.4</td>
<td>1.7</td>
<td>1.3</td>
<td>Y</td>
</tr>
<tr>
<td>Athel_1860 GH74/2 (CBM3)/ GH48</td>
<td>3.6</td>
<td>0.6</td>
<td>0.1</td>
<td>6.0</td>
<td>0.1</td>
<td>Y</td>
</tr>
<tr>
<td>Athel_1664 flagellin domain protein</td>
<td>3.0</td>
<td>5.4</td>
<td>8.2</td>
<td>0.6</td>
<td>1.5</td>
<td>N</td>
</tr>
<tr>
<td>Athel_1896 EBP</td>
<td>1.4</td>
<td>1.8</td>
<td>2.4</td>
<td>0.7</td>
<td>1.3</td>
<td>Y</td>
</tr>
</tbody>
</table>

| sum for glycosidases | 74.9 | 7.9 | 2.2 | 9.5 | 0.3 |

| number of identified proteins | 59 | 363 | 373 |

\(a\) 100 × NSAF, only proteins with values >1 are shown  
\(b\) GH: glycosidase, EBP: extracellular binding protein

### TABLE 4. Proteomic analysis of GH enrichment in *C. obsidiansis* from time point 48 supernatant

<table>
<thead>
<tr>
<th>Locus and Description</th>
<th>Affinity digest isolate (AD)</th>
<th>Tangential flow filtration concentrate (TFF)</th>
<th>Affinity digest supernatant (ADSN)</th>
<th>AD/TFF</th>
<th>ADSN/TFF</th>
<th>SignalP</th>
</tr>
</thead>
<tbody>
<tr>
<td>COB47_1673 GH^{9}/3 (CBM3)/ GH48</td>
<td>39.7</td>
<td>4.5</td>
<td>2.9</td>
<td>8.8</td>
<td>0.6</td>
<td>Y</td>
</tr>
<tr>
<td>COB47_1671 GH10/3 (CBM3)/ GH5</td>
<td>18.0</td>
<td>1.9</td>
<td>0.8</td>
<td>9.5</td>
<td>0.4</td>
<td>Y</td>
</tr>
<tr>
<td>COB47_1669 GH9/2 (CBM3)/ GH5</td>
<td>13.5</td>
<td>1.0</td>
<td>0.5</td>
<td>13.7</td>
<td>0.5</td>
<td>Y</td>
</tr>
<tr>
<td>COB47_0549 EBP(b)</td>
<td>9.2</td>
<td>5.1</td>
<td>10.5</td>
<td>1.8</td>
<td>2.0</td>
<td>Y</td>
</tr>
<tr>
<td>COB47_0918 flagellin domain protein</td>
<td>2.1</td>
<td>3.9</td>
<td>7.3</td>
<td>0.5</td>
<td>1.8</td>
<td>N</td>
</tr>
<tr>
<td>COB47_2069 SLH(b)-containing protein</td>
<td>1.6</td>
<td>2.1</td>
<td>2.1</td>
<td>0.8</td>
<td>1.0</td>
<td>Y</td>
</tr>
<tr>
<td>COB47_1166 EBP</td>
<td>1.3</td>
<td>1.4</td>
<td>1.7</td>
<td>0.9</td>
<td>1.2</td>
<td>Y</td>
</tr>
<tr>
<td>COB47_1704 EBP</td>
<td>1.3</td>
<td>1.8</td>
<td>2.8</td>
<td>0.7</td>
<td>1.5</td>
<td>Y</td>
</tr>
</tbody>
</table>

| sum for glycosidases | 71.2 | 7.4 | 4.2 | 9.7 | 0.6 |

| number of identified proteins | 49 | 335 | 272 |

\(a\) 100 × NSAF, only proteins with values >1 are shown  
\(b\) GH: glycosidase, EBP: extracellular binding protein, SLH: S-layer homology domain