Journal of proteome research

ARTICLE

pubs.acs.org/jpr

Label-free Quantitative Proteomics for the Extremely Thermophilic Bacterium *Caldicellulosiruptor obsidiansis* Reveal Distinct Abundance Patterns upon Growth on Cellobiose, Crystalline Cellulose, and Switchgrass

Adriane Lochner,[†,‡,§,∥] Richard J. Giannone,[‡,∥,⊥] Martin Keller,[†,‡] Garabed Antranikian,[§] David E. Graham,[*,†,#] and Robert L. Hettich[*,‡,⊥]

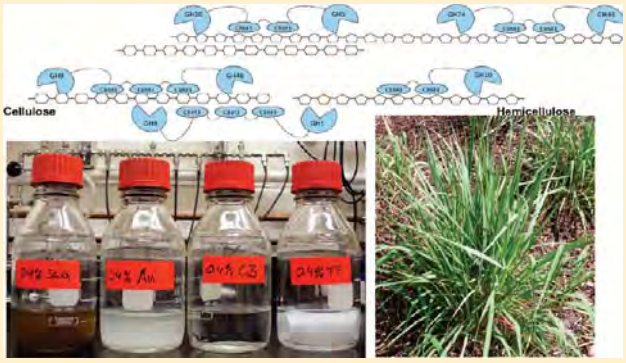
[†]Biosciences Division; [⊥]Chemical Sciences Division; [‡]BioEnergy Science Center at Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

[§]Technical Microbiology, Hamburg University of Technology, Kasernenstrasse 12, D-21073 Hamburg, Germany

[#]Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996, United States

Supporting Information

ABSTRACT: Mass spectrometric analysis of *Caldicellulosiruptor obsidiansis* cultures grown on four different carbon sources identified 65% of the cells' predicted proteins in cell lysates and supernatants. Biological and technical replication together with sophisticated statistical analysis were used to reliably quantify protein abundances and their changes as a function of carbon source. Extracellular, multifunctional glycosidases were significantly more abundant on cellobiose than on the crystalline cellulose substrates Avicel and filter paper, indicating either disaccharide induction or constitutive protein expression. Highly abundant flagellar, chemotaxis, and pilus proteins were detected during growth on insoluble substrates, suggesting motility or specific substrate attachment. The highly abundant extracellular binding protein COB47_0549 together with the COB47_1616 ATPase might comprise the primary ABC-transport system for cellooligosaccharides, while COB47_0096 and COB47_0097 could facilitate monosaccharide uptake. Oligosaccharide degradation can occur either via extracellular hydrolysis by a GH1 β-glycosidase or by intracellular phosphorolysis using two GH94 enzymes. When *C. obsidiansis* was grown on switchgrass, the abundance of hemicellulases (including GH3, GH5, GH51, and GH67 enzymes) and certain sugar transporters increased significantly. Cultivation on biomass also caused a concerted increase in cytosolic enzymes for xylose and arabinose fermentation.



KEYWORDS: microbial proteomics, microbial cellulose degradation, quantitative proteomics, thermophilic bacteria, bioenergy research

INTRODUCTION

The bioconversion of lignocellulosic biomass is beginning to address the increased global demand for renewable transportation fuels. Among the suggested strategies for a commercially viable process is consolidated bioprocessing (CBP), an industrial workflow that allows simultaneous enzyme production, biomass hydrolysis and fermentation by the deployment of a cellulolytic, hemicellulolytic, and ethanologenic microbe or a microbial consortium.[1,2] To select a suitable CBP organism, it is important to fully understand the diverse protein machinery used by microorganisms to degrade biomass. This molecular-level insight has been enabled by recent advances in high-throughput technologies such as whole genome sequencing, transcriptomics, and proteomics. Specifically, the application of quantitative proteomics enables comparisons of protein abundances across a myriad of conditions and time points. Due to post-transcriptional regulation, measurements of protein abundance via mass spectrometry can be a more direct proxy for metabolic processes compared to the transcriptome.[3]

Understanding the deconstruction of model substrates is an important first step to degrading real biomass, such as switchgrass, whose structural and compositional complexity complicates both data acquisition and interpretation. In plant cell walls, crystalline cellulose fibers are embedded in a matrix of other polymers like hemicellulose and lignin. Of the many potential bioenergy crops in America, switchgrass (*Panicum virgatum*) is one of the most promising due to its high productivity,

Received: June 5, 2011
Published: October 11, 2011

robustness and low ecological impact. This perennial grass does not require extensive management or fertilization, and it can be grown on marginal lands that do not support food crops.[4]

The most abundant hemicellulose polymers in grasses are glucuronoarabinoxylans, which are composed of a D-xylose backbone substituted with single units of L-arabinose and D-glucuronic acid.[5] Their cell walls also contain lower levels of D-galactose, D-mannose and D-rhamnose sugars.[6] To make cellulose fibers in switchgrass accessible to enzymatic treatment, the surrounding hemicellulose and lignin matrix must be removed by pretreatment, such as dilute acid hydrolysis under elevated temperatures. Even after this harsh pretreatment, the remaining solids contain significant amounts of carbohydrates other than D-glucose.[6] Potential CBP organisms must be able to hydrolyze mixed carbohydrate polymers and ferment these diverse sugars to produce high yields of biofuels.

Most studies of microbial biomass degradation begin using soluble model substrates for growth, like D-glucose and D-cellobiose (a β-1,4 linked glucose-disaccharide), or purified, processed crystalline cellulose, like filter paper and Avicel. Soluble substrates minimize protein losses due to protein or cell sorption to substrates. More complicated insoluble, cellulosic substrates better mimic natural biomass, but residual substrate depletes the pool of soluble, carbohydrate-binding proteins in cultures.[7] Homogeneous Avicel PH-101 is a colloidal dispersion of microcrystalline cellulose fibers obtained by mineral acid hydrolysis of wood pulp. Another widely used, relatively homogeneous cellulosic substrate is Whatman number 1 filter paper, which is produced from cotton fibers. While both share a similar crystallinity index value,[8] the enzymatic accessibility of filter paper seems to be higher than that of Avicel. Filter paper also contains a higher maximum protein absorption capacity and a higher proportion of β-glycosidic bonds available to endoglucanases.[9]

Gram-positive, thermophilic bacteria belonging the genus *Caldicellulosiruptor* hydrolyze both cellulose and hemicellulose, and they ferment both C_5 and C_6 sugars.[10,11] Growing cells on different defined substrates provides the opportunity to interrogate specific physiological responses by measuring changes in mRNA or protein abundance. In *Caldicellulosiruptor bescii*, transcript abundances of 1203 genes changed significantly in a comparison of cells growing on glucose versus crystalline cellulose substrates.[12] The expression of gene clusters encoding secreted, multifunctional glycosidases increased on cellulose. A semiquantitative proteomic study compared the secretomes (the suite of extracellular proteins secreted by an organism) of *Caldicellulosiruptor saccharolyticus* and *Thermotoga maritima*, when the cells were grown independently and in coculture.[13] The investigators used normalized spectra abundance factors (NSAF), which are derived from counting the number of spectra obtained for any given protein identified in the sample. NSAF values have been shown to deliver a reliable measure of protein abundance.[14] We recently reported the time series analysis of secreted protein abundance for two highly thermophilic bacteria, *Caldicellulosiruptor obsidiansis* and *Caldicellulosiruptor bescii,* during growth on Avicel.[7] That project identified more than 400 extracellular proteins, providing valuable information about growth-phase dependent enrichment of glycosidase enzymes. These proteomics experiments are overcoming the obstacles of sample preparation, dynamic range, and variable peptide sensitivity to provide the first reliable measurements of protein abundance changes.

Several different quantitative proteomics strategies endeavor to estimate protein abundance in a relative or absolute fashion. These strategies have recently been applied to compare cellular strategies to degrade different carbohydrates and biomass. For example, an isobaric tagging technique coupled with liquid chromatography−tandem mass spectrometry (LC−MS/MS) revealed that the actinobacterium *Thermobifida fusca* only expresses 6 of 36 glycosidase genes in significant amounts for the degradation of cellulose.[15] Another study, employing stable isotope labeling for reductive dimethylation, identified stoichiometric changes in the production of over 100 carbohydrate active enzymes in *Clostridium phytofermentans*.[16] Likewise, a metabolic isotope labeling strategy applied to *Clostridium thermocellum* uncovered the expression pattern of 41 cellulosomal components when grown on cellulose or soluble cellobiose.[17] In addition to these isotopic label-based methods, the application of semiquantitative, label-free quantitation, based on spectral counts, intensity, and/or chromatographic peak area, has been successfully used to estimate protein abundance, providing similar results compared to those using label-based strategies.[18,19]

To understand how *C. obsidiansis* degrades and ferments cellulose-based substrates of increasing complexity, cells were grown on four different substrates: cellobiose, Avicel, filter paper and dilute acid-pretreated switchgrass. Label-free, semiquantitative proteomics was applied to measure protein levels in both intra- and extracellular protein fractions, from multiple biological and technical replicates of each growth condition. The results provide a robust estimate of changes in protein abundance that help to explain how these cellulolytic microorganisms attack compositionally different substrates.

MATERIALS AND METHODS

Cultivation and Sampling

C. obsidiansis OB47 cultures were grown anaerobically in 150 mL of basal growth medium prepared as described previously[20] using 250-mL bottles (42 mm neck) to allow for ample head space. Growth experiments were conducted in triplicate with 0.4 or 0.2% (w/v) substrate for each of the following carbon sources: cellobiose, Avicel-PH101, Whatman #1 filter paper and switchgrass[21] (Figure 1). The latter was pretreated with dilute sulfuric acid at the National Renewable Energy Laboratory (Golden, CO),[21] washed several times with deionized water to remove soluble sugars, and dried at 50 °C to obtain the dry weight. A 2% (v/v) inoculum was passed at least 3 times on the same carbon source before starting the growth experiment.

Planktonic cell densities were measured using a Petroff-Hauser microscope counting chamber. For growth experiments cultures were grown at 75 °C until late stationary phase. Cultures for proteomics experiments were harvested at the end of the exponential growth stage as follows. Cell cultures were centrifuged twice (5500× *g*, 4 °C, 30 min), and the supernatants were filtered through a 0.22 μm PES membrane (Millipore) to obtain a cell-free fraction (SN). Aliquots of 50 mL were processed immediately for proteomics analysis. To obtain a substrate-free cell fraction the pellets were resuspended in buffer solution containing 10 mM MOPS (pH 6.8) and 1 mM DTT, vortexed vigorously and vacuum filtered through a 3-μm glass filter. After three washes with 50 mL of buffer, the filtrate containing substrate-free cells was centrifuged (swinging bucket rotor, 3100× *g*, 4 °C, 15 min) and also washed twice to remove any residual supernatant protein. Aliquots containing 50 mg of cells were frozen in liquid nitrogen and stored at −80 °C.

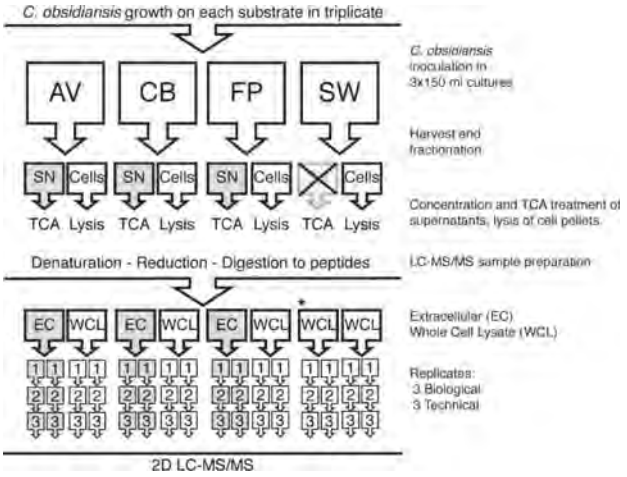


Figure 1. Flowchart illustrating cell culture and sample preparation scheme. *C. obsidiansis* was grown in triplicate with 0.4% (w/v) of each substrate: Avicel (AV), cellobiose (CB), filter paper (FP) and switchgrass (SW). At the late exponential growth stage all cultures were centrifuged and filtered to obtain cell-free supernatant (SN) and substrate-free cells. SN was concentrated by TCA precipitation, and the cells were lysed by sonication. The proteins in both fractions were then prepared for mass spectrometric analysis and extracellular (EC) and whole cell lysate (WCL) samples for each biological replicate were analyzed as technical replicates in 48 instrument runs. * Switchgrass supernatant contained colored solutes that interfered with the sample analysis. A reduction of the substrate load in the cultures to 0.2% (w/v) did not resolve the problem. However, the WCL data correlated well with the one obtained from 0.4% w/v and was therefore included in the statistical analysis.

Sample Preparation for Proteomics Analysis

Cell-free supernatant samples (SN) were prepared for LC−MS analysis as follows (Figure 1). For each biological replicate, three 50-mL aliquots of cell-free culture supernatant were concentrated to approximately 0.5 mL using a Vivaspin centrifugal device containing a noncellulosic PES membrane with a 5000 molecular weight cutoff (Sartorius Stedim). Concentrated samples were then pooled, precipitated with trichloroacetic acid (TCA) as previously described,[7] resuspended in 6 M guanidine-HCl, 50 mM Tris-HCl, 10 mM DTT, at pH 8, and incubated at 60 °C for 1 h. Denatured and reduced protein samples were then diluted to 1 M guanidine-HCl with 50 mM Tris-HCl, pH 8.0 and digested with two separate aliquots of sequencing-grade trypsin (Promega) at a 1:100 enzyme to protein ratio (w:w) successively overnight and 4 h. Following digestion, each sample was desalted and solvent exchanged to acidified water (0.1% formic acid) using a Sep-Pak cartridge (Waters) as detailed in the manufacturer's protocol. Peptide concentrations of the resulting extracellular samples (EC) were measured using the BCA assay (Thermo-Pierce). Cellular fractions were obtained as follows: 50 mg of cell pellet were resuspended in 6 M guanidine-HCl, 10 mM DTT, pH 8.0 and lysed via sonic disruption (Branson). Resulting whole cell lysates (WCLs) were incubated at 60 °C for 1 h and processed to peptides as described above (Figure 1).

Analysis of Peptides by MudPIT LC−MS/MS

Depending on the fraction, an aliquot corresponding to 25 μg (EC) or 100 μg (WCL) of peptides was bomb-loaded onto a biphasic MudPIT back column as previously described.[22,23] Loaded peptides were then placed in-line with an in-house pulled, reversed-phase packed nanospray emitter and analyzed by either a shortened 3-step MudPIT (EC fraction) or full length MudPIT (WCL fraction). Although both LC schemes follow a similar, previously described protocol,[12] they differ in the number of ammonium acetate salt pulses applied to the column with the shortened, mini-MudPIT consisting of 3 salt pulses (25, 50, and 100% of 500 mM ammonium acetate) and the unchanged full length MudPIT consisting of 11 total pulses. LC-separated peptides were analyzed via either a hybrid LTQ/Orbitrap (EC fractions) or LTQ-XL (WCL fractions) mass spectrometer (Thermo Fisher) operating in a data-dependent fashion as previously described. With regard to Avicel (AV), cellobiose (CB), and filter paper (FP), a total of two replicate measurements were obtained for each of three biological replicates, resulting in a total of six LC−MS/MS measurements per substrate per fraction. For switchgrass (SW) however, a total of six biological replicates were measured resulting in 12 measurements (Figure 1).

MS Data Analysis and Evaluation

Acquired MS/MS spectra were assigned to specific peptide sequences using the SEQUEST search algorithm[24] with a FASTA proteome database specific to *C. obsidiansis*.[25] The database contained common contaminant protein entries as well as reversed decoy sequences to assess protein-level false discovery rates. SEQUEST-scored peptide sequence data were filtered and assembled into protein loci using DTASelect[26] with the following conservative criteria: XCorr: +1 = 1.8, +2 = 2.5, +3 = 3.5, DeltCN 0.08, and 2 peptides per protein identification with at least one required to be unique. Prior to semiquantitative analysis, spectral counts were rebalanced to properly distribute nonunique/shared peptides between their potential parent proteins as previously described[7] and raw SpC values for technical replicates (two total) were averaged so that each biological replicate was represented by a single SpC value.

After merging the raw SpC data and assembling all substrates and replicates into an all-inclusive table, each protein (row) was assigned a prevalence value (PV) represented by the "sum of the raw SpC assigned over all substrate replicates" divided by the protein's "count", that is, the number of times the protein was identified across all substrate replicates. PV cut-offs were determined for both the WCL and EC data sets independently by plotting sequential PV cutoff values against the number of proteins that remain after applying said cutoff. The ideal cutoff was defined by the intersection between the two linear tangents that best represent the plotted data.

Once PV cut-offs were determined, proteins with incomplete representation across all substrates and replicates, that is, no raw spectra collected, were assigned nonzero, fractional SpC values that allowed these proteins to be included in the ANOVA analysis. As originally described by Zybailov et al. 2006, this process replaced "blank"/zero values with fractional spectral counts, the magnitude of which were determined by an iterative process that assesses distributional normality on a substrate-by-substrate basis (see below). Once determined, the fractional value was applied to the entire subdata set to maintain the originally measured spectral count differential. These adjusted values were then converted to normalized spectral counts (nSpC), an extension of the widely recognized normalized spectral abundance factor (NSAF)[14] that is calculated by multiplying the NSAF values by an arbitrary number. In this case,

NSAFs were multiplied by either 10 000 or 30 000 for the EC and WCL fractions, respectively. Multiplicative values above represent a rough average of the number of spectra collected per run and present a different but equivalent way of comparing normalized abundance values as theoretical spectral counts rather than the NSAF decimal value.

Once determined for all proteins in the entire data set, the PV cutoff described above was applied and the nSpC values of the remaining proteins in the data set were log2 transformed to obtain a distribution of values that could be plotted to assess normality, a condition that must be met before applying a parametric statistical test such as ANOVA. Normality was assessed by the Shapiro-Wilk test[27] ($p > 0.05$) with JMP Genomics (SAS Institute, ver. 4.1) on a per substrate basis and was used to evaluate the ideal fractional spectral count value applied to the data (see above).

Log2-transformed nSpC values consisting of proteins that passed the PV cutoff were then imported into the JMP Genomics program for statistical analysis as previously described.[7] ANOVA was used to identify proteins that display significant differences in abundance per substrate ($p < 0.01$). All pairwise comparisons were considered with log2-transformed nSpC values standardized (STD option in JMP) across all substrates on a per protein basis to identify significantly changing proteins in both the WCL and EC fractions. Only those proteins that were found to be significantly different in at least one substrate comparison were displayed in each respective heat map, with those exhibiting similar abundance trending grouped into clusters. Relevant protein groups with spectra counts >5 at least on one substrate were then chosen from each cluster via handpicking and combined to assess abundance trends on the different substrates. Although many comparisons could be made, emphasis was placed on proteins with increased abundance on (i) switchgrass or (ii) cellobiose vs crystalline cellulose, (iii) decreased abundance on cellobiose vs crystalline cellulose and (iv) proteins with abundance differences between the crystalline celluloses Avicel and filter paper.

The predicted protein sequences of *C. obsidiansis*[25] were analyzed for secretory signal peptides and orthologs in other *Caldicellulosiruptor* spp. as described previously.[7] The UniProt database was accessed to improve gene function predictions, using the BLASTP tool.[28]

RESULTS

Growth Characterization

To better characterize the growth strategies of *C. obsidiansis* on complex substrates, this organism was grown in triplicate on four different carbon sources: soluble cellobiose, crystalline cellulose (Avicel and filter paper), and acid-pretreated switchgrass. Planktonic cell densities were monitored and growth rates were calculated as averages (Figure 2). A *t* test (alpha of 0.05) indicates the rates obtained from insoluble substrates do not differ significantly from each other, while the growth rate from cellobiose is significantly higher than the other three. Based on the growth curves shown in Figure 2, the planktonic cell densities for the transition state between exponential and stationary growth were determined to be $1-2 \times 10^8$ cells/mL. Cultures were harvested for proteomics measurements at this point for each specific substrate. *C. obsidiansis* can attach to insoluble substrates, thus planktonic cell counts can only be regarded as an estimate of the total concentration of cells in the cultures.

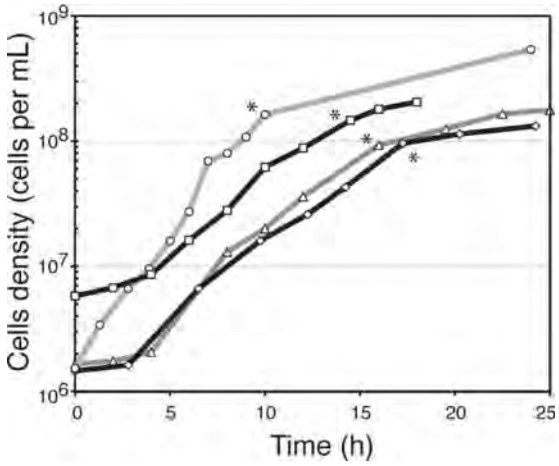


Figure 2. Characterization of *C. obsidiansis* cell growth on different carbon sources. *C. obsidiansis* was grown in triplicate 150-mL cultures on 0.4% (w/v) substrate analyzed over time to assess growth. Growth rates were 0.47 ± 0.01 h^{-1} on cellobiose (circles), 0.28 ± 0.02 h^{-1} on Avicel (squares), 0.30 ± 0.01 h^{-1} on filter paper (triangles) and 0.27 ± 0.02 h^{-1} on switchgrass (diamonds). Asterisks indicate the time points of cell harvest for proteomics analysis at the transition between exponential and stationary growth stages with an approximate cell density of $1-3 \times 10^8$ cells/mL, depending on the substrate.

Table 1. *C. obsidiansis* Proteome Characterization for Whole Cell Lysate (WCL) and Extracellular (EC) Fractions by LC−MS/MS[a]

substrate	AV		CB		FP		SW	
cell fraction	WCL	EC	WCL	EC	WCL	EC	WCL	EC
NRpro (avg per run)	887	333	1086	329	857	354	866	NA
Raw SpC (avg per run)	30035	8414	42638	9028	27191	7825	27827	NA
Total NRpro IDs	1074	427	1260	422	999	465	1128	NA
Core NRpro in all bio reps	713	251	914	249	714	266	631	NA
FDR (avg)	1.73%		2.61%		1.61%		3.11%	

[a] SpC, Spectra counts; NR pro, non-redundant proteins; WCL, whole cell lysate; EC, extracellular; FDR, false discovery rates; NA, not available.

Protein Identification Across Substrates and Fractions

Following LC−MS/MS data acquisition, general proteomic metrics were compared across all substrates. As shown in Table 1, similar numbers of nonredundant proteins (NR pro) were identified in the WCL across all insoluble substrates, while the number of nonredundant proteins identified when *C. obsidiansis* was grown on soluble cellobiose was significantly higher than the other three substrates. The same trend holds true for the number of spectra assigned to peptides (SpC) by the SEQUEST database search algorithm. This trend is less evident in the EC fractions. None of the three replicates of SW EC could be used, due to interference by plant-derived compounds with peptides during reversed-phase sample cleanup or LC−MS/MS data acquisition. Even a decrease in the SW substrate load to 0.2% (w/v) did not remedy the problem. However, WCL protein abundance values obtained from cultures with decreased SW concentrations correlated extremely well with the data set obtained at 0.4% (w/v) and were therefore included in the statistical analysis ($n = 6$).

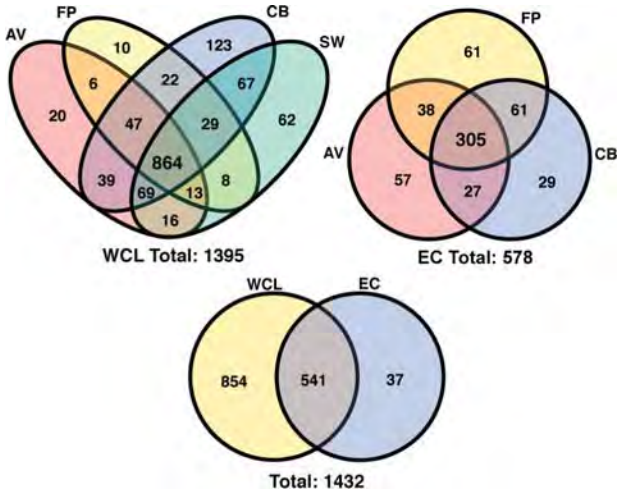


Figure 3. Venn-Diagrams of proteins identified in *C. obsidiansis* grown on different substrates. The numbers of proteins measured across all replicates per substrate are shown while the averages for the replicates are shown in Table 1. Avicel (AV), filter paper (FP), cellobiose (CB), switchgrass (SW).

The total number of proteins identified in each growth condition varied from 1046 to 1283 in the WCL fractions and 290 to 309 in the EC fractions. Searching the MS/MS data against a FASTA protein database concatenated with a reverse-decoy set allowed for the assessment of false-discovery rates (FDR) at the protein-level, which ranged from 1.61 to 3.11% on a per substrate basis. Taken together, a total of 1435 *C. obsidiansis* proteins were identified across all samples (~65% proteome coverage), with 1395 identified across all substrates in the WCL fraction and 578 in the EC fraction (Figure 3). A core set of proteins identified across all substrates included 864 proteins in the WCL fractions and 305 proteins in the EC fractions. A set of 541 highly expressed proteins was identified in both the WCL and EC fractions for all substrates.

Data Processing and Statistical Analysis

To provide a statistically relevant data set, both biological and technical replication were employed (Figure 1). To assess significant differences in protein abundance as a factor of substrate, one-way analysis of variance (ANOVA) was performed. However, before applying this parametric test, several prerequisites must be met, including independent, random sampling of normally distributed data (for data processing see Supplemental Figure 1, Supporting Information). It has been shown previously[14] that both representing "blank" values and converting spectral counts to normalized spectral abundance factors (NSAF) fulfill the necessary prerequisites to substantiate the use of parametric testing once the data have been log-transformed. The maintenance of normality, however, requires that a small fraction of the collected data, specifically those proteins that are spectrally under-represented, are removed from the analysis. Previously, this data trimming has been accomplished by requiring proteins to be identified at least one time per condition.[14] Although effective, this constraint eliminates proteins that may be abundant in one condition but not found in another, which tend to be some of the more interesting observations when considering differential proteomic analyses. To help bring this class of proteins through to the ANOVA analysis, a different cutoff strategy was employed in which each protein was

Table 2. Protein Abundance Correlations in Percent Show Reproducibility of Proteomics Measurements[a]

	CB		AV		FP		SW	
substrate	WCL	EC	WCL	EC	WCL	EC	WCL	EC
CB	97.3	93.4						
AV	86.4	72.3	96.5	91.0				
FP	88.6	54.0	99.4	89.1	93.2	97.2		
SW	86.5	NA	97.6	NA	97.7	NA	98.2	NA

[a] AV, Avicel; CB, cellobiose; FP, filter paper; SW, switchgrass; WCL, whole cell lysate; EC, extracellular fraction; NA, not available.

assigned a prevalence value (PV) as described in the Materials and Methods section. By plotting different PV cutoff values against the number of proteins that remain in the analysis, a curve is generated that allows for the identification of a reasonable cutoff value. As shown in Supplemental Figure 2 (Supporting Information), PV cutoff values of 1.4 and 1.25 were identified and applied to both the WCL and EC data sets, respectively. This process effectively trimmed the data sets from 1395 total NR proteins in the WCL to 956, and from 578 to 344 in the EC fraction, while maintaining 99.2% (WCL) and 98.5% (EC) of the total raw spectra collected for either fraction.

As proteins with spotty, albeit biologically relevant spectral representation are able to pass the PV-filter, for example, those that exhibit conditionally discrete expression patterns, missing values must be represented to properly log-transform the data for subsequent ANOVA analysis. The iteratively determined fractional spectra count was thus applied to all values of the substrate subdata set, effectively shifting the data away from zero while maintaining the originally measured spectral count differential (see Materials and Methods). Values of 0.31 (AV), 0.73 (CB), 0.39 (FP), and 0.16 (SW) were applied to the WCL fraction and 0.33 (AV), 0.40 (CB), and 0.33 (FP) for the EC fraction.

The processed data sets (readjusted, PV-filtered, normalized spectral counts) were then log2-transformed and assessed for normality by the Shapiro-Wilk test.[27] Data distributions are shown in Supplemental Figure 3 (Supporting Information). All data sets achieved *p*-values greater than 0.05, suggesting that the data are normally distributed.

Sample-to-sample variation was assessed by Pearson correlation analysis using all nSpC data points after blank-substitution. As shown in Table 2, the average correlation between biological replicates in both WCL and EC fractions ranged from 91.0 to 98.2%, with an overall average greater than 95% across all biological replicates. Comparing correlation values across substrates, the CB data set varied the most when compared to the other cellulose-based substrates, a trend that was consistent across both the WCL (as low as 86.4% similar) and EC (as low as 54.0% similar) fractions.

Identification of Significantly Changing Proteins

Following data processing, proteins that passed the PV cutoff were moved to JMP Genomics for statistical analysis. In total, 956 WCL and 344 EC proteins were independently analyzed by one-way ANOVA analysis to identify those proteins that displayed differential abundance patterns due to growth on a particular substrate (Supplementary Tables 1 and 2, Supporting Information). Once *p*-values were calculated across multiple pairwise analyses, only those proteins that had a *p*-value <0.01, indicating a significant change in abundance in at least one

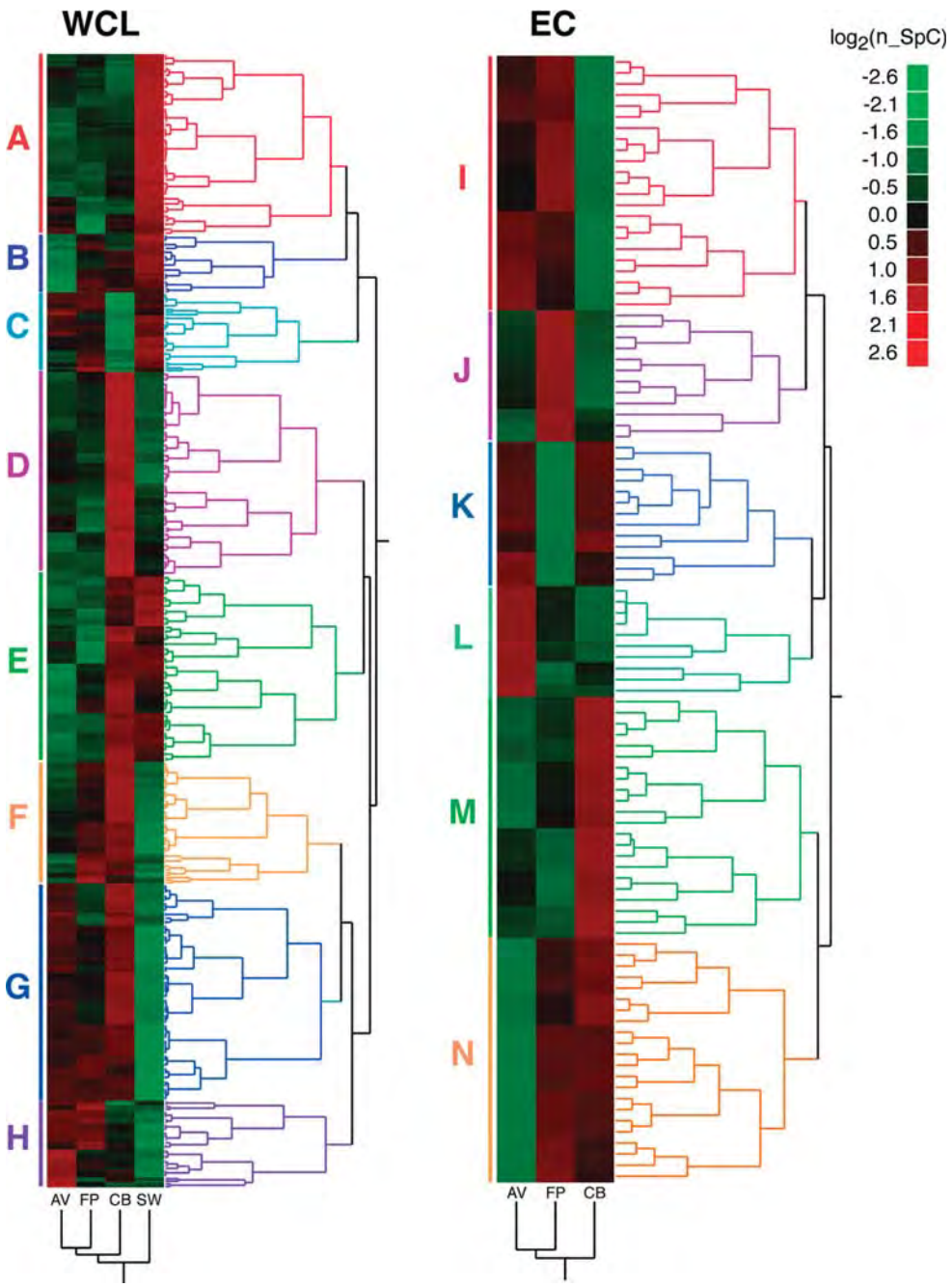


Figure 4. Heat maps show proteins whose abundances changed significantly in the whole cell lysates (WCL) and the cell-free culture supernatant (EC). LC−MS/MS data sets were acquired in biological triplicates and technical duplicates. Peptides were identified using SEQUEST, and NSAF values for each protein were log2 transformed and standardized for one-way ANOVA analysis. Abundances at each time point range from low (green) to high (red). Clusters are colored differently for clarity and labeled WCL-A to -H and EC-I to -N.

substrate comparison, were considered for cluster analysis and heat map generation (Figure 4). Abundance trends were evaluated on a protein-by-protein basis, with each protein's abundance pattern normalized to itself (0−100%, Supplemental Tables 1 and 2, Supporting Information). Proteins exhibiting similar trend profiles were clustered into groups in order to indirectly assess substrate-dependent protein expression. As protein abundance estimates were never compared between proteins (i.e., protein X is 2-fold greater than protein Y), abundance estimate bias inherent to MS-based peptide identification was avoided.

Identification of the Most Abundant Proteins Across the Substrates

The clustering in Figure 4 is based on standardized values rather than the absolute abundance of proteins, which bins proteins by their trending similarity. Proteins with high average abundance levels across all substrates are shown in Table 3 for the whole-cell fraction (average nSpC >200) and in Table 4 for the extracellular fraction (average nSpC > 100). Although cell lysis likely released abundant cytosolic proteins into the supernatant, the known, abundant extracellular *C. obsidiansis* proteins were

Table 3. Most Abundant *C. obsidiansis* Proteins in Whole Cell Lysates Based on Normalized Spectra Counts[a]

locus	annotation	SigP	CB	AV	FP	SW	mean	fold change
0164	DNA-binding protein, AbrB family	N	1639	4228	4236	3931	3508	2.6
0604	DNA-binding protein, histone family	N	718	1175	988	1083	991	1.6
1147	Glyceraldehyde-3-phosphate dehydrogenase	N	868	894	988	803	888	1.2
1150	Enolase	N	631	590	785	645	663	1.3
0549	Extracellular solute-binding protein, family 1	Y	330	352	435	1083	550	3.3
1918	Chaperonin GroEL	N	522	453	522	495	498	1.2
0739	Translation elongation factor Tu	N	484	432	515	478	477	1.2
1368	Stage V sporulation protein S	N	288	486	556	519	463	1.9
1026	Ribosomal protein L7/L12	N	177	529	524	294	381	3.0
1145	Pyruvate, phosphate dikinase	N	345	299	348	339	333	1.2
2069	S-layer domain-containing protein	Y	317	274	339	386	329	1.4
0154	Rubrerythrin	N	315	292	263	217	272	1.3
2305	Stage V sporulation protein S	N	359	238	182	239	254	1.3
1616	ATPase	N	394	161	225	171	238	1.4
0807	Pyruvate ferredoxin oxidoreductase subunit	N	313	208	281	133	234	2.1
0576	Fructose-1,6-bisphosphate aldolase, class II	N	189	209	161	270	207	1.7

[a] Normalized spectra counts; CB, cellobiose; AV, Avicel; FP, filter paper; SW, switchgrass; SigP, SignalP predictions.

Table 4. Most Abundant *C. obsidiansis* Proteins in Cell-free Supernatant Based on Normalized Spectra Counts[a]

locus	annotation	SigP	CB	AV	FP	mean	fold change
0440	Unknown protein	Y	282	877	1374	844	4.9
1150	Enolase	N	513	471	391	458	1.3
0918	Flagellin domain protein	N	87	528	667	427	7.7
0549	Extracellular solute-binding protein, family 1	Y	510	417	95	340	5.4
2069	S-layer domain-containing protein	Y	319	396	215	310	1.8
1145	Pyruvate, phosphate dikinase	N	270	283	230	261	1.2
1673	Bifunctional multidomain glycosidase (CelA)	Y	465	94	130	230	5.0
2294	Unknown protein	N	313	286	90	230	3.5
1377	Malonyl CoA-acyl carrier protein transacylase	N	216	211	131	186	1.6
0739	Translation elongation factor Tu	N	258	105	154	172	2.5
1918	Chaperonin GroEL	N	251	126	135	171	2.0
1687	Unknown protein	Y	34	207	229	157	6.7
1147	Glyceraldehyde-3-phosphate dehydrogenase	N	53	196	166	138	3.7
1919	Chaperonin Cpn10	N	160	59	100	106	2.7
1907	Aminotransferase class V	N	130	76	97	101	1.7

[a] Normalized spectra counts; CB, cellobiose; AV, Avicel; FP, filter paper; SW, switchgrass; SigP, SignalP predictions.

still detected within the top 10 extracellular proteins (Table 4).[7] This set includes CelA, the major cellulase secreted by *Caldicellulosiruptor* spp.

Protein Abundance Variation on the Different Carbon Sources

To visualize the distribution and signal intensity for selected proteins from different WCL and EC clusters, the nSpC values were plotted as bar graphs (Figure 5). Plotted proteins required at least five normalized spectra counts (nSpC) in one of the measurements and belong to one of the following groups: glycosidase (GH), pentose hexose metabolism (PHM), ABC-transporter component (ABC), extracellular binding protein (EBP), flagellar protein (Fla), pilus protein (Pil), or chemotaxis protein (CT). Proteins in these sets differed by several orders of magnitude in abundance, therefore a log scale was chosen to illustrate these nSpC values in Figure 5. Figure 5A shows proteins with increased abundance on cellobiose compared to crystalline cellulose (clusters: WCL-D and -E and EC-M), while Figure 5B shows proteins with decreased abundance on cellobiose (clusters: WCL-C and EC-I). Proteins with abundance differences observed between the two crystalline celluloses are represented in Figure 5C, with those higher on Avicel originating from clusters WCL-G and -H and EC-K and -L and those higher on filter paper from WCL-B and -F and EC-J and -N. Figure 5D shows proteins with abundance increases in the cell lysates obtained from switchgrass cultures compared to crystalline cellulose, including clusters WCL-A and -B.

DISCUSSION

Sixty-five percent of *C. obsidiansis*' 2192 open reading frames were represented by a protein product, identified by at least two raw SpC across both cellular (WCL) and extracellular (EC)

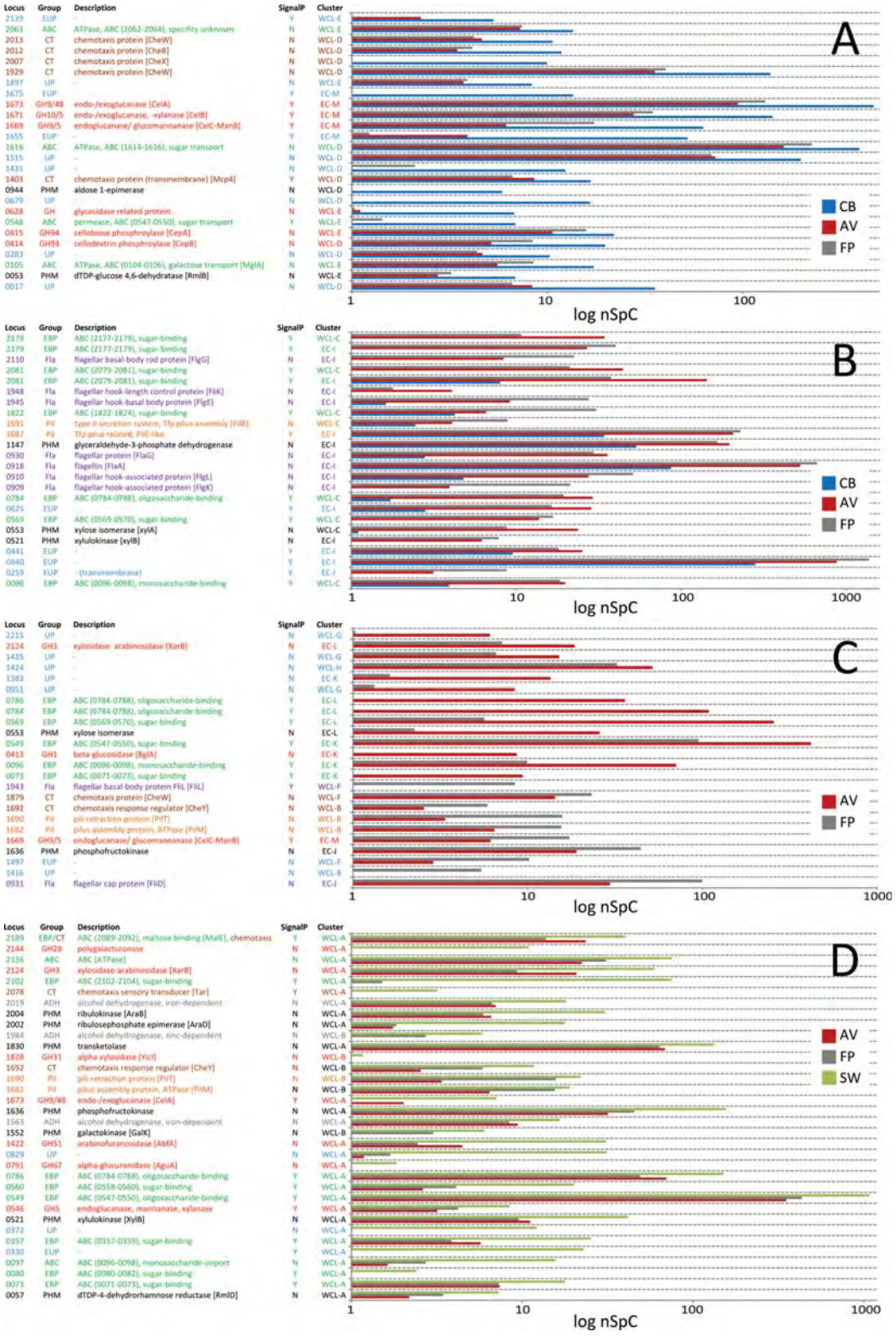


Figure 5. Abundance changes of selected proteins. (A) Proteins with higher abundances in cellobiose-grown cells (CB, blue) than in crystalline cellulose-grown cells: Avicel (AV, red) and filter paper (FP, gray). (B) Proteins that were more abundant in crystalline cellulose-grown cells than in cellobiose-grown cells. (C) Proteins more abundant in either Avicel or filter paper-grown cells. (D) Proteins that were more abundant in switchgrass-grown cells (SW, green) than in Avicel or filter paper-grown cells. Proteins with more than 5 spectral counts on any of the substrates were included and grouped as following: glycosidases (GH) including family number, proteins involved in pentose and hexose metabolism (PHM), ABC-transporter components (ABC) and extracellular binding proteins (EBP) that can be part of an ABC transporter, proteins containing S-layer homology domains (SLH), proteins involved in cell motility or adhesion such as flagella (motility (Fla) or pili (Pil) and chemotaxis proteins (CT)).

dx.doi.org/10.1021/pr200536j |*J. Proteome Res.* 2011, 10, 5302–5314

fractions. Acquiring data sets across four different substrates and two cellular fractions enabled this deep coverage and identified many proteins that were unique to specific growth conditions. In comparison, 286 (11%) intracellular proteins were identified from the 2679 predicted protein in *C. saccharolyticus*[13] and 1429 (54%) of 2666 predicted protein-coding sequences were confirmed by proteomic analyses of *C. bescii*.[12] A large core group of proteins was identified across all substrates in the present study of *C. obsidiansis*. WCL and EC fractions shared many of the same proteins which most likely results from (i) intracellular production of proteins in the process of being secreted, (ii) secreted proteins that bind to the cell wall, and/or (iii) cytosolic proteins released into the supernatant by cell lysis. For these reasons our previous analysis of 418 EC proteins (19%) secreted by *C. obsidiansis* over the course of its growth provided a more focused picture of the cells' secreted cellulolytic apparatus.[7] In the present study, cultures were sampled only at the transition from exponential to stationary growth phases. It was previously shown that the abundance of the both *C. obsidiansis*' and *C. bescii*'s cellulose-degrading glycosidases sharply increased at this transition point, despite increased cell lysis and protein adsorption to insoluble substrates.[7] Our previous experiments demonstrated a strong correlation between extracellular cellulase activity and NSAF-derived protein abundance,[7] supporting the robustness of this semiquantitative approach to determine how *C. obsidiansis*' proteome adapts to degrade cellulosic substrates of varying constitution and complexity. Clearly, future experiments could use internal protein or peptide standards to calibrate NSAF-derived protein abundances for individual glycosidases.[29]

Cellulose versus Cellobiose

The cellulolytic strategy employed by *C. obsidiansis* includes the secretion of three major multidomain, multifunctional enzymes (CelA, COB47_1673; CelB, COB47_1671; and CelC-ManB, COB47_1669).[7] Although these proteins function as both endo- and exoglycosidases, effectively allowing them to act on substrates with a high degree of polymerization, their highest levels were measured in the secretome of disaccharide-grown cultures. These results could be explained by glycosidase depletion in the cellulosic media, caused by enzymes binding to insoluble substrates. All three enzymes contain multiple CBM3 (carbohydrate-binding module, family 3) domains that bind cellulose, and they were readily purified by cellulose affinity digestion.[7] Alternatively, cellobiose or its degradation products could induce glycosidase expression: cellobiohydrolases produce cellobiose through cellulose hydrolysis, making this molecule a good proxy for cellulose abundance. Cellobiose induces cellulase expression in the fungus *Trichoderma reesei*,[30,31] the aerobic bacterium *Acidothermus cellulolyticus*,[30] and the actinomycete *Thermobifida fusca*.[31] In contrast, *C. thermocellum* grown on crystalline cellulose demonstrated much higher levels of cellulase production than cellobiose-grown cells, implicating alternative modes of cellulase regulation.[32,33] *Caldicellulosiruptor* spp. likely use multiple levels of cellulase regulation, because these enzymes are expressed regardless of carbon source, even in glucose or xylose-grown cultures.[12,13,34]

Our previous analysis of *C. obsidiansis* extracellular proteins identified the highly abundant flagellin protein (COB47_0918).[7] The present study identified more proteins from the flagellar apparatus (COB47_0909, 0910, 0918, 1945, and 1948) as well as putative chemotaxis proteins (COB47_2013, 2012, 2007, 1929 and 1403) in cells grown on all four substrates. The production of flagellar proteins appears to increase with growth on the crystalline celluloses compared to cellobiose (Figure 5B), suggesting that high concentrations of the disaccharide may repress flagellar biosynthesis. In contrast, chemotaxis protein abundance was highest in cellobiose-grown cells (Figure 5A). We have not observed motility or chemotaxis in these cells, and no flagella have been observed by microscopic observation of planktonic cells. It is possible that the cells have undetected, fine flagella that are used for motility in the absence of cellobiose, or that the flagella facilitate cellular attachment to insoluble substrates, as observed in *Clostridium difficile*.[35] The substrate adhesion of *Caldicellulosiruptor* cells, as well as their formation of a biofilm, could increase enzyme efficacy by tethering cells close to their substrate and limiting the diffusion of hydrolysis products.[36] *C. obsidiansis* also produced the components of Type IV pili, indicated by the increased abundance of COB47_1682, 1687, 1690, and 1691 proteins during growth on insoluble substrates (Figure 5). In addition to their role in cellular attachment, Type IV pili can confer gliding motility in some bacteria, such as *Clostridium perfringens*.[37]

Due to the energetic cost of importing substrates, as well as ATP gains realized by phosphorolytic cleavage of oligosaccharides as opposed to hydrolysis, it is energetically more favorable to take up oligosaccharides with higher degrees of oligomerization.[38] Two GH94 proteins homologous to cytosolic cellobiose/cellodextrin phosphorylases (COB47_0414 and 0415) were slightly more abundant in cellobiose-grown cells (Figure 5A). Phylogenetic analysis identified a cluster of orthologous proteins that includes COB47_0415 and the CepA cellobiose phosphorylase from *Clostridium stercorarium* (data not shown). Another cluster includes COB47_0414 and the CepB cellodextrin phosphorylase from *C. stercorarium* that catalyzes the phosphorolysis of cellooligosaccharides with degrees of polymerization of 3 to 6.[39] Gene expression data from *C. saccharolyticus*[40] and protein abundance data from *C. obsidiansis* indicate the cells may use highly abundant ATP-binding cassette (ABC)-transporters (COB47_0547−0550 and perhaps the ATPase subunit COB47_1616) to import cellobiose or short-chain cellooligosaccharides (Figure 6).

Avicel versus Filter Paper

Both forms of crystalline cellulose share similar crystallinity indices, and the protein abundance correlation between cells grown on each substrate is high (Table 2). Filter paper, which has a higher degree of polymerization than Avicel and four-times higher accessibility to cellulose,[41] elicits an increased abundance of potential adhesion structures like flagella and pili. It is possible that the hydrolysis of Avicel releases more monosaccharides or short chain oligosaccharides that might suppress flagellar formation as described above. Six EBPs were more abundant in Avicel-grown cells (Figure 5C), including the COB47_0784 and 0788 proteins that we observed previously.[7] The orthologous *C. saccharolyticus* transporter transcripts were not identified in a previous study when the organism was grown on monosaccharides.[42]

C. obsidiansis also imports monosaccharides.[20] Glucose is a major reaction product from the hydrolysis of crystalline cellulose by CelA,[43] the most abundant glycosidase observed across several species. Additionally, a GH1 enzyme (COB47_0413) similar to the divalent cation-dependent β-glucosidase BglA from *Clostridium cellulovorans*[44] was identified in cellobiose and Avicel cultures but seems to be less abundant upon growth on filter paper. Like its orthologs in *C. bescii* and *Clostridium* spp., COB47_0413 was found exclusively in the extracellular fraction

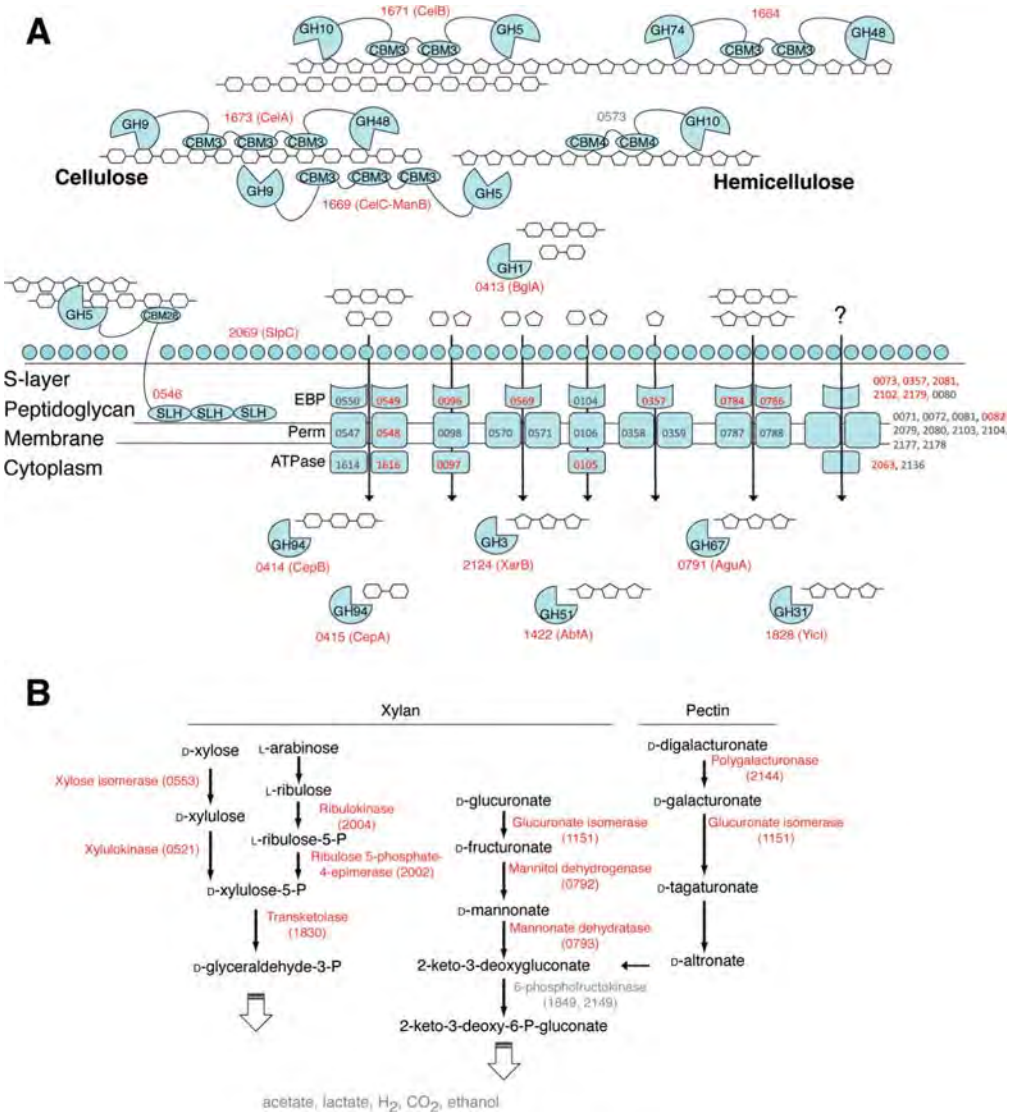


Figure 6. Proposed mechanisms for biomass deconstruction and carbohydrate metabolism in *C. obsidiansis*. Locus numbers are shown without COB47 prefixes in red for mass spectrometrically identified proteins and in gray for genome-encoded but not detected proteins. (A) Degradation of cellulose and hemicellulose by a small number of multifunctional, multidomain glycosidases that are usually cell-free except for one potentially cell wall-associated glycosidase (0546) (sizes are not proportional). ABC-transporters that can transport mono- and oligosaccharides are composed of extracellular binding protein (EBP), permease (Perm) and ATPase subunits. The highly abundant ATPase COB47_1616 may serve the highly abundant COB47_0548/49 transport system (Tables 3 and 4). Sugar import is followed by phosphorolysis and hydrolysis of oligosaccharides by cytosolic enzymes. GH, glycosidase; CBM, carbohydrate-binding module; SLH, S-layer homology domain. (B) Sugars derived from xylan and pectin hydrolysis are enzymatically converted into glyceraldehyde-3-phosphate and 2-keto-3-deoxy-6-phosphogluconate that are subsequently converted into the fermentation products acetate, lactate, CO_2, H_2 and ethanol. Enzymes in red were found to be more abundant in switchgrass-grown cell lysates.

although it contains no Sec-pathway translocation signal[12] and must be secreted using a different pathway. Members of this β-glucosidase family release hexoses from a variety of plant cell wall polymers;[45] therefore, further studies will be required to determine whether this enzyme primarily functions in cellulose degradation.

Switchgrass versus Crystalline Cellulose

Although extracellular proteomic data for switchgrass deconstruction is unavailable at this time, data from the whole-cell analysis as well as extracellular proteomic data from model substrates described here and in previous studies[7,13,34] suggest that *C. obsidiansis* employs a concerted enzymatic strategy to degrade complex biomass, as proposed in Figure 6A. A similar proteomic study using purified C. *thermocellum* cellulosomes showed reduced abundance of xylanases in switchgrass-grown cultures.[33] *C. thermocellum* cannot use C_5 sugars;[38] therefore, it might have a different regulatory system for hemicellulase expression than *Caldicellulosiruptor* spp. Pili proteins, glycosidases and carbohydrate uptake components, numerous hemicellulases, ABC-transporters and enzymes involved in pentose metabolism exhibit increased abundance on switchgrass compared to the crystalline substrates (Figure 5D). Acidic pretreatment partially hydrolyzes switchgrass hemicellulose, although the solids may retain some sugars besides glucose.[6] As a result, *C. obsidiansis* expressed a variety of enzymes to hydrolyze glucuronoarabinoxylan, a polymer

comprised of a β-1,4-xylosyl backbone with arabinose, glucuronate, and acetyl side chains.[5] Xylose and arabinose are the most common monosaccharides obtained from hemicellulose hydrolysis, followed by galactose as well as small amounts of mannose, rhamnose, fucose and uronic acids.[6] The CelB and CelC glycosidases have broad substrate specificities and are actively involved in hemicellulose degradation.[34,42,46,47]

Glycosidases with putative secretion signals and increased abundance in switchgrass-grown cell lysates may represent the extracellular glycosidase complement. The abundance of the putative β-1,4-endoglucanase/mannanase/xylanase COB47_0546 increased in these lysates. This enzyme contains a catalytic GH5 domain, a CBM28 amorphous cellulose binding module, and three S-layer homology (SLH) domains. The homologous enzyme in *C. bescii* (Cbes_0594) has been suggested to be attached to the cell-wall via the SLH modules and could promote cellular attachment to the substrate.[12] The COB47_2124 protein also increased in abundance. This enzyme contains a GH3 catalytic module that is similar to the *Thermoanaerobacter ethanolicus* XarB protein, which mainly functions as a β-xylosidase or α-arabinosidase but also exhibits β-glucosidase activity.[48] The GH31 enzyme COB47_1828 is a conserved α-xylosidase, similar to the *Escherichia coli* YicI protein.[49] Another hemicellulolytic enzyme with significantly increased abundance on switchgrass is the α-L-arabinofuranosidase (COB47_1422), an enzyme with a GH51 domain that cleaves arabinofuranoside side chains of hemicelluloses.[50] A recent microarray analysis of gene expression levels in *C. saccharolyticus* cells grown on switchgrass indicates that species uses a different gene complement for hemicellulose degradation.[34] *C. obsidiansis* has no equivalent to the *xynB-xynF* gene cluster of endoxylanases and β-xylosidases that are transcribed at low levels in *C. saccharolyticus*. Conversely, *C. saccharolyticus* has no homologues of the abundant COB47_0546 and COB47_1422, described above.

COB47_0791 was also most abundant on switchgrass and consists of a GH67 domain similar to the α-glucuronidase AguA from *Thermotoga maritima*, which cleaves the glucuronic acid residues of xylooligosaccharides. In close genomic proximity and similarly clustered is the EBP COB47_0786 that could help import the resulting monosaccharides. Annotated as polygalacturonase, the COB47_2144 protein contains a GH28 catalytic domain that is conserved in all known *Caldicellulosiruptor* spp. and could degrade pectin,[51] a minor component of plant cell walls that contains a galacturonic acid backbone. While *C. saccharolyticus* also expresses a homologue of COB47_2144, that gene is not up-regulated during growth on switchgrass, and those cells have no ortholog of the COB47_0786 protein[34]

The main reaction products of the above listed enzymes are most likely monomeric D-xylose, L-arabinose, and D-galacturonate. This is substantiated by the increased prevalence of enzymes involved in pentose metabolism (Figure 6B) and glucuronate degradation.[40] Xylose isomerase (COB47_0553) converts D-xylose to D-xylulose, which is phosphorylated by a xylulokinase (COB47_0551) resulting in D-xylulose-5-phosphate. L-arabinose is converted to the same metabolite by the concerted action of a ribulokinase (COB47_2004) and ribulose-phosphate-5-epimerase (COB47_2002). A similar pathway of pentose conversion was identified in *C. saccharolyticus*,[42] with mostly orthologous enzymes except for COB47_0553. Transketolase (COB47_1830) and transaldolase proteins convert D-xylulose-5-phosphate to D-glyceraldehyde-3-phosphate, a central metabolite.

Highly Abundant Proteins Across All Substrates

Several proteins were identified that do not change significantly but were highly abundant on all substrates. Cytosolic enzymes that metabolize triose phosphate intermediates were consistently detected at high levels, including fructose-1,6-bisphosphate aldolase (COB47_0576), glyceraldehyde-3-phosphate dehydrogenase (COB47_1147), enolase (COB47_1150), phosphoenolpyruvate synthase (COB47_1145) and a pyruvate:ferredoxin oxidoreductase (PFOR) subunit (COB47_0807). Like other iron−sulfur cluster containing enzymes, PFOR is sensitive to oxidative stress. Iron-containing rubrerythrin (COB47_0154) is one of the most abundant proteins identified across all substrates, and could be regulated by the adjacent PerR-type transcriptional regulator (COB47_0155) that senses H_2O_2.[52] Rubrerythrin can act as a peroxidase in *Clostridium acetobutylicum*, and could reduce oxidative damage, together with abundant desulfoferredoxin (COB47_1062), which reduces O_2^- to H_2O_2.[53] The red coloring of the cell pellets could be due to oxidation of ruberythrin (data not shown).

Another group of highly abundant proteins includes DNA-binding proteins and the chaperones GroEL and Cpn10. The most abundant protein is the 80 amino-acid DNA-binding protein AbrB (COB47_0164). The *B. subtilis* homologue binds many sites on the chromosome, contributing to both chromosomal structure and transcriptional regulation. Its binding influences biofilm formation, motility, competence development and extracellular enzyme synthesis, especially during the transition period between exponential and stationary growth phases.[54] Two homologues of the *B. subtilis* SpoVS protein (COB47_1368 and 2305) share 80% identity and are both highly abundant. These putative DNA-binding proteins are widely distributed among bacteria and could function analogously to AbrB.[55]

Several proteins containing SLH domains and secretion signals were also highly abundant in both cellular and extracellular fractions (Supplemental Table 1 and 2, Supporting Information). In many Gram-positive bacteria, SLH domains bind to pyruvate-modified carbohydrates in peptidoglycan, anchoring the proteins to the cell wall.[56] A handful of genes encoding SLH domains were up-regulated and expressed extracellularly when *C. bescii* was grown on filter paper or xylan, suggesting a role in substrate attachment.[12] The most abundant SLH domain protein (COB47_2069) is an ortholog of the *C. saccharolyticus* paracrystalline S-layer protein (Csac_2451);[51] both resemble the S-layer glycoprotein SlpC of *Lysinibacillus sphaericus*.[57] It is possible that *C. obsidiansis* contains an S-layer, which could be glycosylated using modified sugars produced by the dTDP-glucose-4,6 dehydratase (COB47_0053) and the dTDP-4-dehydrorhamnose reductase (COB47_0057) enzymes.[58]

The unknown proteins COB47_0440, 2294, and 1687 are also among the most abundant proteins in the extracellular fraction. While COB47_1687 is possibly involved in pilus formation, the role of the former two is not clear and could potentially be investigated in future studies.

CONCLUSIONS

The quantitative proteomics data collected in this study indicate that primary glycosidase and transporter expression in *C. obsidiansis* occurs constitutively but is increased and augmented in response to varying substrates. Switchgrass elicited increased abundance of hemicellulose-targeted glycosidases and an expanded set of EBPs to import new sugars. To optimize the

microbial or enzymatic biomass conversion with members of the *Caldicellulosiruptor* genus, more in-depth studies will be required to test the suggested hypotheses. Comparative studies employing transcriptomics could be integrated with the proteomics data to give a more complete picture of variations in gene-expression, as well as reveal additional levels of post-transcriptional regulation and rates of protein turnover. Additionally, the isolation or recombinant production of several glycosidases could allow more insight into their substrate specificities. Finally, the establishment of a genetic system for *Caldicellulosiruptor* spp. could enable the production of gene-targeted mutations that would both aid in the understanding of the metabolic protein machinery employed by the organism as well as upgrade its fermentation product mixture to produce alcohols for CBP.

■ ASSOCIATED CONTENT

Supporting Information

Supplemental tables and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*D.E.G.: Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6038. Phone: 865-574-0559. Fax: 865-574-8646. Email: grahamde@ornl.gov.
R.L.H.: Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6131. Phone: 865-574-4986. Fax: 865-576-8559. Email: hettichrl@ornl.gov.

Author Contributions

‖These authors contributed equally to this work.

■ ACKNOWLEDGMENT

We thank Manesh Shah, Barbara Klippel, Andrew Dykstra, Scott Hamilton-Brehm 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 UT-Battelle LLC for the Department of Energy.

■ REFERENCES

(1) Lynd, L. R.; Elander, R. T.; Wyman, C. E. Likely features and costs of mature biomass ethanol technology. *Appl. Biochem. Biotechnol.* **1996**, *57−8*, 741–761.

(2) Lynd, L. R.; van Zyl, W. H.; McBride, J. E.; Laser, M. Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* **2005**, *16* (5), 577–583.

(3) de Groot, M. J.; Daran-Lapujade, P.; van Breukelen, B.; Knijnenburg, T. A.; de Hulster, E. A.; Reinders, M. J.; Pronk, J. T.; Heck, A. J.; Slijper, M. Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. *Microbiology* **2007**, *153* (Pt 11), 3864–3878.

(4) McLaughlin, S. B.; Kszos, L. A. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenergy* **2005**, *28* (6), 515–535.

(5) Vogel, J. Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* **2008**, *11* (3), 301–307.

(6) Dien, B. S.; Jung, H. J. G.; Vogel, K. P.; Casler, M. D.; Lamb, J. F. S.; Iten, L.; Mitchell, R. B.; Sarath, G. Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. *Biomass Bioenergy* **2006**, *30* (10), 880–891.

(7) Lochner, A.; Giannone, R. J.; Rodriguez, M. J.; Shah, M. B.; Mielenz, J. R.; Keller, M.; Antranikian, G.; Graham, D. E.; Hettich, R. L. Label-free quantitative proteomics distinguish the secreted cellulolytic systems of *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis*. *Appl. Environ. Microbiol.* **2011**, *77* (12), 4042–4054.

(8) Park, S.; Johnson, D. K.; Ishizawa, C. I.; Parilla, P. A.; Davis, M. F. Measuring the crystallinity index of cellulose by solid state C-13 nuclear magnetic resonance. *Cellulose* **2009**, *16* (4), 641–647.

(9) Hong, J.; Ye, X. H.; Zhang, Y. H. P. Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. *Langmuir* **2007**, *23* (25), 12535–12540.

(10) Blumer-Schuette, S. E.; Kataeva, I.; Westpheling, J.; Adams, M. W.; Kelly, R. M. Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr. Opin. Biotechnol.* **2008**, *19* (3), 210–217.

(11) Yang, S.; Kataeva, I.; Hamilton-Brehm, S. D.; Engle, N. L.; Tschaplinski, T. J.; Doeppke, C.; Davis, D.; Westpheling, J.; Adams, M. W. W. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe *Anaerocellum thermophilum* DSM 6725. *Appl. Environ. Microbiol.* **2009**, *75* (14), 4762–4769.

(12) Dam, P.; Kataeva, I.; Yang, S. J.; Zhou, F. F.; Yin, Y. B.; Chou, W. C.; Poole, F. L.; Westpheling, J.; Hettich, R.; Giannone, R.; Lewis, D. L.; Kelly, R.; Gilbert, H. J.; Henrissat, B.; Xu, Y.; Adams, M. W. W. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. *Nucleic Acids Res.* **2011**, *39* (8), 3240–3254.

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

(14) Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **2006**, *5* (9), 2339–2347.

(15) Adav, S. S.; Ng, C. S.; Arulmani, M.; Sze, S. K. Quantitative iTRAQ secretome analysis of cellulolytic *Thermobifida fusca*. *J. Proteome Res.* **2010**, *9* (6), 3016–3024.

(16) Tolonen, A. C.; Haas, W.; Chilaka, A. C.; Aach, J.; Gygi, S. P.; Church, G. M. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol. Syst. Biol.* **2011**, *6*, 1–12.

(17) Gold, N. D.; Martin, V. J. J. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J. Bacteriol.* **2007**, *189* (19), 6787–6795.

(18) Ong, S. E.; Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, *1* (5), 252–262.

(19) Paoletti, A. C.; Parmely, T. J.; Tomomori-Sato, C.; Sato, S.; Zhu, D. X.; Conaway, R. C.; Conaway, J. W.; Florens, L.; Washburn, M. P. Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (50), 18928–18933.

(20) Hamilton-Brehm, S. D.; Mosher, J. J.; Vishnivetskaya, T.; Podar, M.; Carroll, S.; Allman, S.; Phelps, T. J.; Keller, M.; Elkins, J. G. *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. *Appl. Environ. Microbiol.* **2010**, *76* (4), 1014–1020.

(21) Schell, D. J.; Farmer, J.; Newman, M.; McMillan, J. D. Dilute-sulfuric acid pretreatment of corn stover in pilot-scale reactor - Investigation of yields, kinetics, and enzymatic digestibilities of solids. *Appl. Biochem. Biotechnol.* **2003**, *105*, 69–85.

(22) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.

(23) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. 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.* **2002**, *219* (1), 245–251.

(24) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(25) Elkins, J. G.; Lochner, A.; Hamilton-Brehm, S. D.; Davenport, K. W.; Podar, M.; Brown, S. D.; Land, M. L.; Hauser, L. J.; Klingeman, D. M.; Raman, B.; Goodwin, L. A.; Tapia, R.; Meincke, L. J.; Detter, J. C.; Bruce, D. C.; Han, C. S.; Palumbo, A. V.; Cottingham, R. W.; Keller, M.; Graham, D. E. Complete genome sequence of the cellulolytic thermophile *Caldicellulosiruptor obsidiansis* OB47^T. *J. Bacteriol.* **2010**, *192*, 6099–6100.

(26) Tabb, D. L.; McDonald, W. H.; Yates, J. R. DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **2002**, *1* (1), 21–26.

(27) Shapiro, S. S.; Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52* (3−4), 591–611.

(28) UniProt-Consortium, The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2008**, *36*, D190-D195.

(29) Zhang, B.; VerBerkmoes, N. C.; Langston, M. A.; Uberbacher, E.; Hettich, R. L.; Samatova, N. F. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* **2006**, *5* (11), 2909–2918.

(30) Shiang, M.; Linden, J. C.; Mohagheghi, A.; Grohmann, K.; Himmel, M. E. Regulation of Cellulase Synthesis in Acidothermus-Cellulolyticus. *Biotechnol. Prog.* **1991**, *7* (4), 315–322.

(31) Spiridonov, N. A.; Wilson, D. B. A celR mutation affecting transcription of cellulase genes in Thermobifida fusca. *J. Bacteriol.* **2000**, *182* (1), 252–255.

(32) Zhang, Y. H.; Lynd, L. R. Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. *J. Bacteriol.* **2005**, *187* (1), 99–106.

(33) Raman, B.; Pan, C.; Hurst, G. B.; Rodriguez, M., Jr.; McKeown, C. K.; Lankford, P. K.; Samatova, N. F.; Mielenz, J. R. Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: A quantitative proteomic analysis. *PLoS One* **2009**, *4* (4), e5271.

(34) VanFossen, A. L.; Ozdemir, I.; Zelin, S. L.; Kelly, R. M. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Biotechnol. Bioeng.* **2011**, *108*, 1559–1569.

(35) Tasteyre, A.; Barc, M. C.; Collignon, A.; Boureau, H.; Karjalainen, T. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect. Immun.* **2001**, *69* (12), 7937–7940.

(36) Wang, Z. W.; Hamilton-Brehm, S. D.; Lochner, A.; Elkins, J. G.; Morrell-Falvey, J. L. Mathematical modeling of hydrolysate diffusion and utilization in cellulolytic biofilms of the extreme thermophile *Caldicellulosiruptor obsidiansis*. *Bioresour. Technol.* **2010**, *102* (3), 3155–3162.

(37) Varga, J. J.; Nguyen, V.; O'Brien, D. K.; Rodgers, K.; Walker, R. A.; Melville, S. B. Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other *Clostridia*. *Mol. Microbiol.* **2006**, *62* (3), 680–694.

(38) Lynd, L. R.; Weimer, P. J.; van Zyl, W. H.; Pretorius, I. S. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **2002**, *66* (3), 506–739.

(39) Reichenbecher, M.; Lottspeich, F.; Bronnenmeier, K. Purification and properties of a cellobiose phosphorylase (CepA) and a cellodextrin phosphorylase (CepB) from the cellulolytic thermophile *Clostridium stercorarium*. *Eur. J. Biochem.* **1997**, *247* (1), 262–267.

(40) van de Werken, H. J.; Verhaart, M. R.; VanFossen, A. L.; Willquist, K.; Lewis, D. L.; Nichols, J. D.; Goorissen, H. P.; Mongodin, E. F.; Nelson, K. E.; van Niel, E. W.; Stams, A. J.; Ward, D. E.; de Vos, W. M.; van der Oost, J.; Kelly, R. M.; Kengen, S. W. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl. Environ. Microbiol.* **2008**, *74* (21), 6720–6729.

(41) Hong, J.; Ye, X.; Zhang, Y. H. Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. *Langmuir* **2007**, *23* (25), 12535–12540.

(42) VanFossen, A. L.; Verhaart, M. R. A.; Kengen, S. M. W.; Kelly, R. M. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. *Appl. Environ. Microbiol.* **2009**, *75* (24), 7718–7724.

(43) Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K. 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 (Reading, U. K.)* **1998**, *144*, 457–465.

(44) Jeng, W. Y.; Wang, N. C.; Lin, M. H.; Lin, C. T.; Liaw, Y. C.; Chang, W. J.; Liu, C. I.; Liang, P. H.; Wang, A. H. J. Structural and functional analysis of three β-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *J. Struct. Biol.* **2011**, *173* (1), 46–56.

(45) Ketudat Cairns, J.; Esen, A. β-Glucosidases. *Cell. Mol. Life Sci.* **2010**, *67*, 3389–3405.

(46) Gibbs, M. D.; Saul, D. J.; Luthi, E.; Bergquist, P. L. The β-mannanase from *Caldocellum saccharolyticum* is part of a multidomain enzyme. *Appl. Environ. Microbiol.* **1992**, *58* (12), 3864–3867.

(47) Morris, D. D.; Reeves, R. A.; Gibbs, M. D.; Saul, D. J.; Bergquist, P. L. 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.* **1995**, *61* (6), 2262–2269.

(48) Mai, V.; Wiegel, J.; Lorenz, W. W. Cloning, sequencing, and characterization of the bifunctional xylosidase-arabinosidase from the anaerobic thermophile *Thermoanaerobacter ethanolicus*. *Gene* **2000**, *247* (1−2), 137–143.

(49) Okuyama, M.; Mori, H.; Chiba, S.; Kimura, A. Overexpression and characterization of two unknown proteins, YicI and YihQ, originated from *Escherichia coli*. *Protein Expr. Purif.* **2004**, *37* (1), 170–179.

(50) Shallom, D.; Belakhov, V.; Solomon, D.; Shoham, G.; Baasov, T.; Shoham, Y. Detailed kinetic analysis and identification of the nucleophile in α-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *J. Biol. Chem.* **2002**, *277* (46), 43667–43673.

(51) Blumer-Schuette, S. E.; Lewis, D. L.; Kelly, R. M. Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. *Appl. Environ. Microbiol.* **2010**, *76* (24), 8084–8092.

(52) Hillmann, F.; Fischer, R. J.; Saint-Prix, F.; Girbal, L.; Bahl, H. PerR acts as a switch for oxygen tolerance in the strict anaerobe Clostridium acetobutylicum. *Mol. Microbiol.* **2008**, *68* (4), 848–860.

(53) Riebe, O.; Fischer, R. J.; Wampler, D. A.; Kurtz, D. M., Jr.; Bahl, H. Pathway for H2O2 and O2 detoxification in Clostridium acetobutylicum. *Microbiology* **2009**, *155* (Pt 1), 16–24.

(54) Chumsakul, O.; Takahashi, H.; Oshima, T.; Hishimoto, T.; Kanaya, S.; Ogasawara, N.; Ishikawa, S. Genome-wide binding profiles of the Bacillus subtilis transition state regulator AbrB and its homolog Abh reveals their interactive role in transcriptional regulation. *Nucleic Acids Res.* **2011**, *39* (2), 414–428.

(55) Rigden, D. J.; Galperin, M. Y. Sequence analysis of GerM and SpoVS, uncharacterized bacterial 'sporulation' proteins with widespread phylogenetic distribution. *Bioinformatics* **2008**, *24* (16), 1793–1797.

(56) Mesnage, S.; Fontaine, T.; Mignot, T.; Delepierre, M.; Mock, M.; Fouet, A. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *EMBO J.* **2000**, *19* (17), 4473–4484.

(57) Hu, X. M.; Li, J.; Hansen, B. M.; Yuan, Z. M. Phylogenetic analysis and heterologous expression of surface layer protein SlpC of *Bacillus sphaericus* C3−41. *Biosci. Biotechnol. Biochem.* **2008**, *72* (5), 1257–1263.

(58) Zayni, S.; Steiner, K.; Pfostl, A.; Hofinger, A.; Kosma, P.; Schaffer, C.; Messner, P. The dTDP-4-dehydro-6-deoxyglucose reductase encoding *fcd* gene is part of the surface layer glycoprotein glycosylation gene cluster of *Geobacillus tepidamans* GS5−97(T). *Glycobiology* **2007**, *17* (4), 433–443.