AEM
Journals.ASM.org

S-Layer Homology Domain Proteins Csac_0678 and Csac_2722 Are Implicated in Plant Polysaccharide Deconstruction by the Extremely Thermophilic Bacterium *Caldicellulosiruptor saccharolyticus*

Inci Ozdemir, Sara E. Blumer-Schuette, and Robert M. Kelly

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina, USA

The genus *Caldicellulosiruptor* contains extremely thermophilic bacteria that grow on plant polysaccharides. The genomes of *Caldicellulosiruptor* species reveal certain surface layer homology (SLH) domain proteins that have distinguishing features, pointing to a role in lignocellulose deconstruction. Two of these proteins in *Caldicellulosiruptor saccharolyticus* (Csac_0678 and Csac_2722) were examined from this perspective. In addition to three contiguous SLH domains, the Csac_0678 gene encodes a glycoside hydrolase family 5 (GH5) catalytic domain and a family 28 carbohydrate-binding module (CBM); orthologs to Csac_0678 could be identified in all genome-sequenced *Caldicellulosiruptor* species. Recombinant Csac_0678 was optimally active at 75°C and pH 5.0, exhibiting both endoglucanase and xylanase activities. SLH domain removal did not impact Csac_0678 GH activity, but deletion of the CBM28 domain eliminated binding to crystalline cellulose and rendered the enzyme inactive on this substrate. Csac_2722 is the largest open reading frame (ORF) in the *C. saccharolyticus* genome (predicted molecular mass of 286,516 kDa) and contains two putative sugar-binding domains, two Big4 domains (bacterial domains with an immunoglobulin [Ig]-like fold), and a cadherin-like (Cd) domain. Recombinant Csac_2722, lacking the SLH and Cd domains, bound to cellulose and had detectable carboxymethylcellulose (CMC) hydrolytic activity. Antibodies directed against Csac_0678 and Csac_2722 confirmed that these proteins bound to the *C. saccharolyticus* S-layer. Their cellular localization and functional biochemical properties indicate roles for Csac_0678 and Csac_2722 in recruitment and hydrolysis of complex polysaccharides and the deconstruction of lignocellulosic biomass. Furthermore, these results suggest that related SLH domain proteins in other *Caldicellulosiruptor* genomes may also be important contributors to plant biomass utilization.

Members of the extremely thermophilic genus *Caldicellulosiruptor* have potential as consolidated bioprocessing (CBP) microorganisms because of their capacity to convert plant-based polysaccharides directly into a biofuel (i.e., hydrogen) in a growth-associated manner (6, 7, 25, 54). The first member of this genus to be studied in detail, *Caldicellulosiruptor saccharolyticus* (43), has been examined with respect to its genome sequence (53), sugar transport (56), bioenergetics (14, 54), utilization of cellulose in comparison to other members of the genus (7), capacity to degrade plant biomass (55), and biotechnological potential (33). Looking forward, further insights into how *C. saccharolyticus* functions as a CBP microorganism will help in the ultimate goal of designing microbial systems, thermophilic or otherwise, for direct plant biomass conversion to biofuels.

For microorganisms that convert insoluble forms of cellulose and other recalcitrant plant polysaccharides to fermentable sugars, the synergistic action of a variety of glycoside hydrolases (GHs) must be coordinated with overall growth physiology. One strategy is to produce a cellulosome, a novel biological structure that packages many GHs and accessory proteins into a single unit (5, 18); cellulosomes have been described in *Clostridia* (3, 4) and other bacteria and fungi (2). Within the cellulosome are enzymes endowed with carbohydrate-binding modules (CBMs) that serve to anchor the biocatalyst to the substrate surface, thereby placing the active site in close proximity to the substrate (36). The CBMs can also play a role in destabilizing the insoluble substrate, such that enzymatic activity is enhanced or made possible (8, 35). Other cellulose-degrading microorganisms, such as *Trichoderma reesei*, hydrolyze the insoluble substrate through the direct action of several GHs, not associated with a cellulosome (37). This is also the case for the cellulolytic, extremely thermophilic *Caldicellulosiruptor* species. For example, the genome of *C. saccharolyticus* encodes at least a dozen multidomain GHs, 10 of which have identifiable signal peptides (55). These 10 multidomain GHs presumably play key roles in CBP, since they can interact directly with plant polysaccharides. Several of these extracellular GHs (either from *C. saccharolyticus* or related orthologs in *Caldicellulosiruptor bescii*) have been characterized biochemically: Csac_1076 (CelA) (50), Csac_1078 (CelB) (48, 55), Csac_2410 (XynE) (55), and Csac_2411 (XynF) (55). In addition, the secretome of *C. saccharolyticus*, grown on glucose, contained the GHs encoded by Csac_1076 to Csac_1079, suggesting a constitutive role in carbohydrate utilization (1).

C. saccharolyticus has 11 S-layer homology (SLH) domain proteins, presumably to enable binding to the S-layer. SLH domains 50 to 60 amino acids (aa) long have been identified at the amino-terminal region of S-layer proteins from various organisms (45) and at the carboxy-terminal end of cell-associated extracellular enzymes (45, 47). S-layer motifs specifically recognize pyruvylated secondary cell wall polymers (SCWPs) as the anchoring structure

Received 27 September 2011 **Accepted** 21 November 2011
Published ahead of print 2 December 2011
Address correspondence to R. M. Kelly, rmkelly@eos.ncsu.edu.

doi:10.1128/AEM.07031-11

TABLE 1 Primers used for cloning Csac_0678 and Csac_2722 genes

Clone	Primer orientation[a]	Primer sequence
Csac_0678	F	GGCGGCCCATGGATAATACTGCGTATGAAAAGGATAAGTATCCACAC
	R	GGCGCCTCGAGCATCTTTCCTGTAAGTTCTAAAATTTTGTA
Csac_0678-TM1	F	GGCGGCCCATGGATAATACTGCGTATGAAAAGGATAAGTATCCACAC
	R	GGCGGCCTCGAGATTTGTAAATCTTACATTGT
Csac_0678-TM2	F	GGCGGCCCATGGATAATACTGCGTATGAAAAGG
	R	GGCGGCCTCGAGAGCACCAGAAGTCTCATTTT
Csac_2722	F	GGCGGCGCTAGCCTGTTCATAAAGAGTACA
	R	GAAGACCTCGAGCTATTGAGCCTGTCCATAGGTGGC

[a] F, forward; R, reverse.

(17, 47). SLH domains seemingly play a contributing role in plant polysaccharide degradation. For example, cellulosomes contain proteins that have SLH domains (16). The anchoring mechanism of the *Clostridium thermocellum* cellulosome to the cell surface involves several proteins with repeating SLH domains: OlpA, OlpB, open reading frame 2p (ORF2p), and SdbA (34, 45). Many studies, both *in vivo* and *in vitro*, involving extracellular enzymes containing both catalytic domains and SLH domains connected through a linker region, have showed that the SLH motif anchors the enzyme to the cell surface (9, 12, 31, 34, 38, 51). The linker region likely provides a certain degree of flexibility, facilitating attack on the substrate (38). Apparently, SLH domains neither contribute to enzymatic activity nor are required for substrate binding (9). These domains are implicated in the binding of enzymes to the cell surface so that release of hydrolysis products is in close proximity to, and the products can be readily transported into the cell (40, 63).

In the *C. saccharolyticus* genome, two SLH domain-containing proteins (Csac_0678 and Csac_2722) could be distinguished from others by the presence of putative binding domains and, in the case of Csac_0678, a glycoside hydrolase (CAZy [http://www.cazy.org]) (10) belonging to family 5 glycoside hydrolases (GH5), which are reported to have a common $(\beta/\alpha)_8$ TIM barrel fold (24). For Csac_0678, the presence of the SLH domains suggests that this protein associates with the S-layer and may play a specific role in utilization of insoluble substrates by *C. saccharolyticus* and other *Caldicellulosiruptor* species. In addition to the SLH domain, Csac_0678 has a family 28 CBM; this CBM family (type B) has a cleft shape to accommodate cellooligosaccharides (52) and belongs to the β-jellyroll fold subfamily (23). No identifiable GH catalytic domain is present in Csac_2722, the largest ORF in the *C. saccharolyticus* genome. However, this protein contains putative binding motifs, which suggests some involvement in polysaccharide utilization. Furthermore, the cadherin-like domains may play a role in protein-protein interactions, binding to bacterial cell surfaces (19, 20); these domains were shown to bind carbohydrates, such as chitin and pectin (19). In this report, biochemical characteristics of Csac_0678 and Csac_2722 are examined with an eye toward the role of these proteins in plant biomass deconstruction.

MATERIALS AND METHODS

Bacterial strains and culture conditions. *Escherichia coli* NovaBlue and *E. coli* Rosetta(DE3) (Novagen, Madison, WI) were used as cloning and expression hosts, respectively. The *E. coli* strains were cultivated in Luria-Bertani (LB) medium, supplemented with kanamycin (Fisher Bioreagents) (30 μg/ml) and/or chloramphenicol (Sigma) (34 μg/ml). *C. saccharolyticus* DSM 8903 cultures were grown anaerobically at 70°C in modified DSM640 medium (DSMZ [http://www.dsmz.de]) containing NH_4Cl (0.9 g/liter), NaCl (0.9 g/liter), $MgCl_2 \cdot 6H_2O$ (0.4 g/liter), KH_2PO_4 (0.75 g/liter), K_2HPO_4 (1.5 g/liter), yeast extract (1 g/ml), trace element solution SL-10 (1 ml/liter), cellobiose (5 g/liter), and $Na_2S \cdot 9H_2O$ (0.05% [wt/vol]). Cells were visualized by epifluorescence microscopy (Intensilight C-HGFI; Nikon, Lewisville, TX) as previously described (55).

Cloning, expression, and purification of Csac_0678, Csac_0678-TM1, Csac_0678-TM2, and Csac_2722. Standard molecular cloning techniques (46) were used in this work. Genomic DNA from *C. saccharolyticus* was isolated as described previously (22). ORFs containing these genes were amplified by PCR from *C. saccharolyticus* genomic DNA by using the primers listed in Table 1. The primers were designed in a way that a C-terminal histidine (His) tag was added to all three versions of Csac_0678 constructs and an N-terminal His-tag was added to Csac_2722. Gene constructs used to express intact proteins and truncation mutants are shown in Fig. 1. The insertions and the plasmid pET-28b(+) (Novagen, Madison, WI), were digested by the restriction endonucleases (New England BioLabs, Beverly, MA). DNA fragments ranging from 990 to 2,134 bp were ligated into the cut vector. The plasmids, which were confirmed to contain the insertion, were isolated from the cloning host by using a Qiagen kit (QIAprep, Valencia, CA). The nucleotide sequences of both strands of the DNA insertions were determined at the Duke University Health System DNA Analysis Facility. *E. coli* Rosetta(DE3) cells, containing pET-28b(+) under the control of T7lac promoter (pET system manual, Novagen, Madison, WI), were used for gene expression. Expression of the target genes was induced by the addition of IPTG (isopropyl-β-D-thiogalactopyranoside) to Rosetta cells when the optical density at 600 nm (OD_{600}) reached 0.8. The cells were then harvested after 4 h of induction by centrifugation at 10,000 $\times$ *g* for 10 min. The cell pellet was resuspended with 5 ml of 50 mM sodium phosphate (pH 8.0), 100 mM NaCl, Nonidet P-40 (0.1% vol/vol), 100 μg/ml lysozyme, and 1 μg/ml DNase for every gram of wet cell pellet. The cells were lysed by sonication (S-4000; Misonix Ultrasonic Liquid Processors, Farmingdale, NY) for 10 min with 10-s off/on pulses. The suspension was heat treated at 60°C for 20 min in order to remove *E. coli* proteins. The cell extract was obtained after centrifuging the lysed and heat-treated cells at 20,000 $\times$ *g* for 20 min. The intact Csac_0678 (without the signal peptide) and the two truncation mutants (Csac_0678-TM1, which lacked the SLH domains, and Csac_0678-TM2, missing the SLH domains and CBM) were present in the cell extract in a soluble form, whereas Csac_2722 was found in the insoluble fraction, forming inclusion bodies. A protein-refolding kit (Novagen, Madison, WI) was used to resolubilize recombinant Csac_2722. HiTrap HP (GE Life Sciences, Piscataway, NJ) and a Resource Q column (GE Life Sciences, Piscataway, NJ) were used to purify the

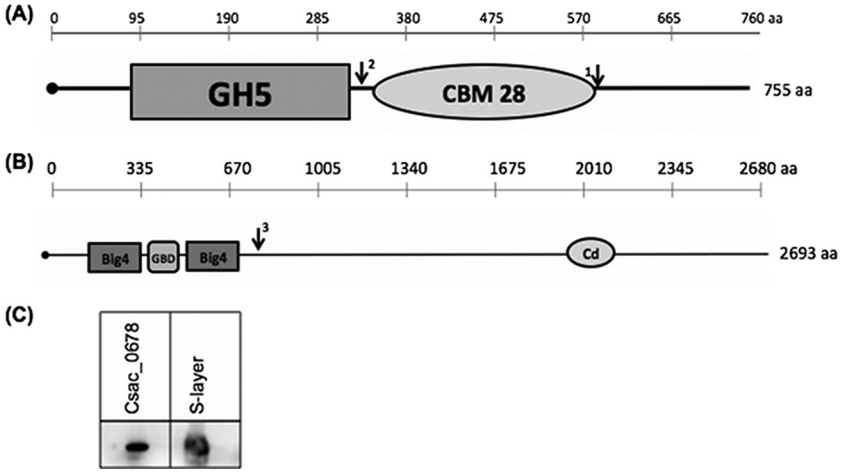

FIG 1 Csac_0678 and Csac_2722 conserved domains and truncation points for deletion mutants. (A) Recombinant intact Csac_0678 (727 aa), lacking only the signal peptide. Csac_0678-TM1 (575 aa) and Csac_0678-TM2 (317 aa) were constructed by truncation at the points 1 and 2, respectively. (B) Recombinant Csac_2722 (810 aa), without the signal peptide, was truncated at point 3. GH5, glycoside hydrolase family 5; CBM28, carbohydrate binding module family 28; SLH, surface layer homology domain; GBD, galactose binding domain-like; Big4, family 4 bacterial immunoglobulin-like domain, Cd, cadherin-like domain. ●, signal peptide. (C) Western blot showing the localization of Csac_0678 in the S-layer of *C. saccharolyticus* grown on cellobiose. The lane labeled as "Csac_0678" had purified recombinant Csac_0678. The lane labeled as the "S-layer" is the S-layer fraction extracted from native *C. saccharolyticus* cells. Negative controls with an unrelated protein did not have any bands (data not shown).

recombinant proteins on a Biologic DuoFlow fast protein liquid chromatograph (FPLC) (Bio-Rad, Hercules, CA). Purity of the recombinant intact Csac_0678, Csac_0678-TM1, Csac_0678-TM2, and Csac_2722 was evaluated by SDS-PAGE. Size exclusion chromatography was carried out to determine the oligomeric state of Csac_0678 using a Superdex-75 HiLoad 16/60 column (GE Healthcare Lifesciences), which was calibrated with the following protein standards: cytochrome *c* (12.4 kDa), carbonic anhydrase (29 kDa), albumin (66 kDa), alcohol dehydrogenase (150 kDa), and β-amylase (200 kDa). The column was equilibrated with 50 mM Tris-HCl (pH 8.0)–100 mM NaCl. The protein sample was applied at 2 mg/ml in 0.5 ml of equilibration buffer. The protein concentration was determined by the Bradford protein assay (Bio-Rad, Hercules, CA), using bovine serum albumin as the standard (Sigma).

Biochemical characterization of Csac_0678. The optimal pH for recombinant Csac_0678 was determined at 70°C in buffers containing 2% carboxymethyl cellulose (CMC) at pH 2.5 to 9 (50 mM sodium acetate buffer, pH 2.5 to 6, and 50 mM sodium phosphate buffer, pH 6 to 9). The optimal temperature for Csac_0678 activity was determined by 3,5-dinitrosalicylic acid (DNS) assay between 40°C and 90°C at the optimal pH. The melting temperature of Csac_0678 was obtained using differential scanning calorimetry (DSC) (Calorimetry Sciences Corporation, Provo, UT). Csac_0678 was prepared at 1 mg/ml in phosphate-buffered saline (PBS) buffer and scanned between 25°C and 125°C at a rate of 1°C min^{-1}. Thermoinactivation of Csac_0678 was examined by incubating the enzyme without the substrate in the reaction buffer at 70, 77, 80, and 85°C for 48 h. Aliquots were taken at different time intervals, and residual activity was assayed on CMC.

The DNS reducing sugar assay, which was adapted to utilize a miniaturized 96-well microplate format (61, 62), was used to detect both five- and six-carbon reducing sugars. DNS reagent was prepared as previously described (39). Avicel (PH-102; 100-μm average particle size) (Sigma), oat spelt xylan (Sigma, St. Louis, MO), birchwood xylan (Sigma), barley glucan (Megazyme International Ireland Ltd., Wicklow, Ireland), konjac root glucomannan (Jarrow Formulas, Los Angeles, CA), amyloid xyloglucan (Megazyme), ivory nut mannan (Megazyme), laminarin (from *Laminaria digita*) (Sigma, St. Louis, MO), switchgrass (*Panicum virgatum* −20/+80 mesh fraction; with dilute acid pretreatment performed at the National Renewable Energy Laboratory [44]), lichenan (Megazyme), arabinoxylan (Megazyme), and carboxymethylcellulose (CMC; Sigma) were prepared at 1% (wt/vol) in 50 mM sodium acetate (pH 5). Single discs of filter paper with a radius of 5 mm and total weight of 3 mg (Whatman no. 1; Whatman, Kent, United Kingdom) were used in the same buffer. Phosphoric acid-swollen cellulose (PASC) was prepared from Sigmacell (type 20; 20-μm average particle size) (Sigma), as described previously (65), and was used at a 1% (wt/vol) final concentration in each reaction. Bacterial microcrystalline cellulose (BMCC) (provided by David B. Wilson, Cornell University, Ithaca, NY) at 0.5% (wt/vol) suspended in 50 mM sodium acetate (pH 5) was used for activity analysis. The reactions took place at 75°C and 500 rpm in a thermomixer (Eppendorf) with 2.7 mg/ml Csac_0678. Xylose and glucose standards were used to convert absorbance readings to reducing sugar concentrations. One unit of enzyme activity is defined as the amount of enzyme required for production of 1 μmol of glucose or xylose per minute. The kinetic parameters (V_{max} and K_m) were calculated by using *p*-nitrophenyl-β-D-cellobioside (pNPC) (Sigma). Substrate concentrations were varied between 0.5 and 10 mM. After adding 2 M Na_2CO_3 (pH 10), the absorbance was measured at 420 nm. The concentration of the product was calculated by using a standard curve of 4-nitrophenol (Sigma) in the same buffer. For V_{max} and K_m calculation, 1 U of enzyme activity is defined as the amount of enzyme required for production of 1 μmol of 4-nitrophenol per minute.

Hydrolysis products were determined by high-pressure liquid chromatography (HPLC) refractive index (RI) detection (Waters 2414 RI detector). Shodex KS-801 and KS-802 (Showa Denko K.K., Kanagawa, Japan) columns operated at 80°C with a flow rate of 0.6 ml/min. Hydrolysis products were also viewed on thin-layer chromatography (TLC) by using silica gel 60 TLC sheets (Fisher), following protocols described previously (27). EtOAc-CH_3COOH-H_2O (3:2:1 by volume) was used to develop the TLC plates; reducing sugars were detected by the orcinol reagent (1% orcinol in 10% H_2SO_4 dissolved in ethanol). Cellooligosaccharides and xylooligosaccharides (Sigma) were used as standards.

Activity of Csac_0678 on filter paper. The processivity was determined by using a modified protocol of Irwin et al. (29). Single discs of filter paper with a radius of 5 mm and total weight of 3 mg were incubated with 9 μM purified Csac_0678 at 75°C in 50 mM sodium acetate buffer (pH 5) for 16 h. After the incubation, the supernatant (100 μl), which contained the soluble reducing sugars, was removed. The filter papers were washed with 1 ml of 50 mM sodium acetate buffer three times. After the washing step, 100 μl of the same buffer was added to the filter paper.

DNS reagent (200 μl) was added to the supernatant and to the filter paper tubes. The DNS reaction was run as described above. Reducing sugars were estimated using a glucose standard curve. The ratio of soluble sugars (found in the supernatant fraction) to insoluble sugars (found on the filter paper) was calculated. All of the reactions were run in triplicates.

CMC viscosity reduction assay. CMC viscosity reduction was carried out by following a modified version of a protocol described elsewhere (64). Medium-viscosity CMC (1%) (Sigma) in 50 mM sodium acetate buffer (pH 5) was incubated with 1.5 μM purified Csac_0678 for 1 h at 75°C and 300 rpm. After incubation, the samples were boiled for 5 min and then diluted 10-fold with distilled water. The viscosity of each sample was measured in triplicates by an AR-G2 rheometer (TA Instruments, New Castle, DE).

Carbohydrate affinity assay. Affinity of recombinant Csac_0678, Csac_0678-TM2, and Csac_2722 to insoluble polysaccharides was assessed by following a protocol described elsewhere (32). Avicel cellulose gel PH-102 (20 mg; FMC Corporation), with a 100-μm average particle size, was washed with distilled water and with 50 mM sodium acetate (pH 5.0) three times. Protein (40 μg) was added to the insoluble polysaccharide in the buffer at a final volume of 250 μl. The tubes were incubated at room temperature for 45 min. The sample-polysaccharide mixtures were centrifuged at 15,800 × *g* for 5 min. The supernatant contained the unbound fraction. The pellet was washed with 1 ml of 50 mM sodium acetate (pH 5.0) three times to remove the remaining unbound protein. The bound protein was separated from the insoluble polysaccharides in 100 μl of elution buffer (2% [wt/vol] SDS, 5% [vol/vol] β-mercaptoethanol) by incubation in a boiling water bath for 10 min. These samples were then centrifuged at 15,800 × *g* for 5 min such that the supernatant contained the bound proteins. The bound and unbound fractions were concentrated to 20 μl and compared by SDS-PAGE. Densitometry analysis was done using a Gel Logic 212 Pro densitometer with Carestream MI software v5 (Carestream Molecular Imaging, Woodbridge, CT).

Quantitative binding assays were done following a previously described protocol (57). Protein (40 μg) was incubated in a 1% (wt/vol) Avicel solution in 50 mM sodium acetate buffer at pH 5 and at 4°C for 1 h on a Labquake tube shaker/rotator (Barnstead/Thermolyne). After the incubation, tubes were centrifuged. The optical density of the supernatant at 280 nm was measured by using the Synergy Mx microplate reader (Biotek, Winooski, VT) to determine the concentration of the unbound protein. The extinction coefficients of the proteins were estimated by Expasy ProtParam tool (ExPASy Bioinformatics Resource Portal [http://web.expasy.org/protparam/]). The bound protein fraction was determined by subtracting the unbound protein from the initial amount of protein added.

Localization of Csac_0678 and Csac_2722 in *C. saccharolyticus*: extraction of the S-layer. Four liters of *C. saccharolyticus* cells grown on cellobiose was harvested at early stationary phase by centrifugation at 12,100 × *g* for 10 min. Cell wall fractions were prepared by following the steps described earlier (34), but using a different buffer. Cell pellets were washed twice with 100 ml of 50 mM Tris-HCl (pH 7.4), and cells were lysed by sonication in 30 ml of the same buffer. Intact cells were separated by centrifugation twice at a lower speed (1,940 × *g*) for 5 min. The resulting supernatant was centrifuged at 39,200 × *g* for 20 min. The pellet, resuspended in 5 ml of 50 mM Tris-HCl (pH 7.4), contained the cell wall fraction. The S-layer was purified from the cell wall fraction using a modification of a previously described method (49). The cell wall fraction (24 mg/ml) was treated with lysozyme (150 μg/ml) for 3 h at room temperature on a shaker at 100 rpm. The mixture was centrifuged for 30 min at 25,000 × *g*. The pellet, which contained the S-layer fraction, was washed three times with 50 mM Tris-HCl (pH 7.4) and resuspended in distilled water.

Protein immunoblotting. Proteins in the S-layer fraction and recombinant Csac_0678 and Csac_2722 were resolved by SDS-PAGE using NuPage 4 to 12% Bis-Tris (Invitrogen) precast gels. The proteins were transferred to a nitrocellulose membrane (Whatman) by using Fisher Biotech semidry blotting unit FB-SDB-2020. The blot was incubated with nonfat dry milk to block nonspecific binding of the antibodies. Polyclonal rabbit antibodies (GeneTel Laboratories LLC, Madison, WI), generated against recombinant Csac_0678-TM1 or Csac_2722, both lacking the SLH domains, were incubated with the blot for 2 h. After rinsing to remove the unbound rabbit antibodies, the blot was exposed for 1 h to goat anti-rabbit antibodies conjugated to horseradish peroxide (HRP) (goat anti-rabbit IgG [H+L]) (Invitrogen). Then, the blot was washed to remove any unbound secondary antibody. SuperSignal West Femto Max sensitivity chemiluminescent substrate (Thermoscientific) was used for detection of peroxide activity from HRP-conjugated secondary antibody. A Kodak 1500 Gel Logic system (Carestream Molecular Imaging) was used to image the blot.

RESULTS AND DISCUSSION

Csac_0678 is a bifunctional endoglucanase/xylanase that binds to the S-layer. The sequenced *Caldicellulosiruptor* genomes contain ORFs representing putative glycoside hydrolases that are associated with S-layer homology domains (Table 2). The *C. saccharolyticus* DSM 8903 genome, in particular, encodes at least 57 identifiable glycoside hydrolases (GHs), 16 of which have one or more catalytic domains linked to one or more noncatalytic domains (53, 55), but only Csac_0678 has SLH domains, in addition to the glycoside hydrolase (GH5) and sugar-binding (CBM28) domains.

Orthologs of Csac_0678 exist (Table 2) in all eight sequenced *Caldicellulosiruptor* species; in 6 species, including *C. sacharolyticus*, the ORF encodes a 755- to 756-amino-acid (aa) protein (85 kDa), while in *C. owensensis* and *C. obsidiansis*, the ortholog is 566 to 567 aa (64 kDa), differentiated from the longer version by a truncated CBM28 domain. Since Csac_0678 and its orthologs are so highly conserved (Csac_0678 is at least 65% identical at the amino acid sequence level to each of the orthologs), they are a defining feature of *Caldicellulosiruptor* species.

The Csac_0678 gene encodes a signal peptide at the N terminus (SignalP 3.0 Server, [http://www.cbs.dtu.dk/services/SignalP/]) and potentially associates with the cell surface via the SLH domains (Fig. 1A). The molecular assembly of recombinant Csac_0678 (molecular mass of 81 kDa without the signal peptide), as determined by size exclusion chromatography, was found to be monomeric. The optimum temperature and pH of Csac_0678 were determined to be 75°C and 5.0, respectively. Csac_0678 demonstrated high thermostability, retaining 50% of its original activity after 48 h of incubation at 75°C (Fig. 2). The half-lives were found to be ~19 h at 77°C, less than 3 h at 80°C, and less than 30 min at 85°C. Differential scanning calorimetry (DSC) of Csac_0678 with the SLH domains showed a single unfolding transition with a peak at 81°C. To examine cellular localization, polyclonal rabbit antibodies raised against Csac_0678 lacking the SLH domains (Csac_0678-TM1) were used to probe for possible Csac_0678 attachment to the S-layer in cells grown on cellobiose. By using anti-Csac_0678-TM1 antibodies, hybridization to SLH domains was avoided. Immunoblot analysis (Fig. 1C) showed that anti-Csac_0678-TM1 antibodies bound specifically to the S-layer fraction. Purified recombinant Csac_0678 was used as the positive control. An unrelated protein sample from *Sulfolobus solfataricus* and Benchmark protein ladder (Invitrogen) were used as negative controls. Anti-Csac_0678 antibodies recognized the enzyme localized on the cell surface.

Biochemical characterization of Csac_0678. Recombinant Csac_0678 was tested for activity toward a range of complex car-

TABLE 2 SLH domain proteins in *Caldicellulosiruptor* genomes that contain sugar binding domains and GH catalytic domains

Domains[a]	ORF	No. of amino acids
GH5-CBM28-SLH-SLH-SLH	Csac_0678 Calkro_2036	755
	Calla_0352 Calhy_2064 Calkr_2007 Athe_0594	756
	COB47_0546	568
	Calow_0459	567
(CBM22)-CBM22-CBM22-GH10-CBM9-CBM9-CBM9-(CE15)-SLH-SLH	Calow_1924	1,625
	Calkro_0402	1,672
	Calla_0206	1,593
	Calkr_2245	2,159
(SLH-SLH-SLH-CBM54-GH16)[b]-CBM4-CBM4-CBM6-CBM4-CBM4-CBM4	Calkro_0072 Calhy_0060 COB47_0076	1,732
	Csac_2548-2549[b]	1,648[c]
SLH-SLH-SLH-CBM54-FN3-GH16-FLD-FN3-FN3-FLD-GH55-CBM32-(CBM32)	Calkro_0111	2,435
	Calkro_0121	2,229
SLH-SLH-SLH-GH43-CBM54	Calhy_1629	1,440
CBM35-GH87-FN3-SLH-SLH-SLH	Calhy_2383	2,007

[a] SLH, S-layer homology domain; FN3, fibronectin III domain; Big4, Ig-like domain; CBM, cellulose binding domain (followed by family number); FLD, fascin-like domain.
[b] Csac_2548 and Csac_2549 are homologous to the ORFs listed and were most likely created by a deletion event that truncated the GH16 domain and split the ORF into genes coding for two separate proteins.
[c] Total size of Csac_2548 and Csac_2549.

bohydrates (Table 3). Activity on barley glucan and lichenan indicated a preference toward soluble substrates with mixed β-1,4 and β-1,3 glycosidic linkages. Csac_0678 exhibited moderate activity on CMC and PASC, which are at least 20-fold more accessible than Avicel (66). Csac_0678 hydrolyzed glucomannan but not mannan, presumably attacking β-(1→4)-linked D-glucose units in glucomannan. It was interesting that Csac_0678 also degraded polymers composed of β-(1→4)-linked xylose units, such as xylan and arabinoxylan. In comparison with XynA (306 U·mg^{-1}) and XynB (4,600 U·mg^{-1}) from *Thermotoga maritima* (60), Csac_0678 xylanase activity was very low. However, it was comparable to those of thermophilic xylanases from *Bacillus* sp. strain NCIM59 (0.0172 and 0.742 μM·min^{-1}·mg^{-1}) (15). Csac_0678

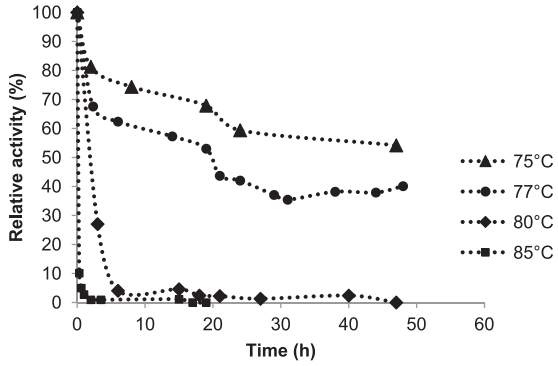

FIG 2 Comparison of the thermostabilities of Csac_0678 at different incubation temperatures. The half-life at 75°C was ~48 h, and that at 77°C was 19 h. Csac_0678 lost 50% of initial activity in less than 3 min at 80°C and in less than 30 min at 85°C.

hydrolyzed pNPC, indicating that the active site of Csac_0678 can accommodate cellobiose moieties. In addition, low but measurable activity was detected on insoluble substrates, such as BMCC, filter paper, Avicel, and switchgrass. Although specific activity on switchgrass could not be directly measured, HPLC analysis showed small amounts of glucose and xylose as hydrolysis products.

Table 3 shows that Csac_0678 exhibited bifunctional xylanase and endoglucanase activity, despite the fact that it contains only a single catalytic domain (GH5). A *Thermotoga maritima* endoglucanase, Cel5A, comprised of only the GH5 domain, had dual activity against both glucan- and mannan-based polysaccharides (41). The specific activity of Cel5A on CMC was 616 U·mg^{-1}, but no activity was detected on xylan (birchwood) (13). Cel5A activity on Avicel was not reported. Another endoglucanase, Cel5B from *Thermobifida fusca*, containing both GH5 and CBM3 domains, had specific activities of 121.4 and 3.9 U·mg^{-1} against CMC and Avicel, respectively

TABLE 3 Hydrolytic activity of Csac_0678 on complex carbohydrates

Substrate[a]	Linkage	Sp act (U·mg^{-1} enzyme)[b]
Barley glucan	β-1,3/4-Glucan	28.2
Lichenan	β-1,3/4-Glucan	17.4
CMC	β-1,4-Glucan	8.94
pNPC	β-1,4-Glucan	5.52
Glucomannan	β-1,4-Glucan/mannan	4.21
PASC	β-1,4-Glucan	2.43
Oat spelt xylan	β-1,4-Xylan	0.709
Birchwood xylan	β-1,4-Xylan	0.570
Arabinoxylan	β-1,4-Xylan	0.500
BMCC	β-1,4-Glucan	0.227
Filter paper	β-1,4-Glucan	0.0518
Avicel	β-1,4-Glucan	0.0113
Xyloglucan	β-1,4-Glucan/β-1,6-xylan	Low activity
Mannan	β-1,4-Mannan	ND[c]
Laminarin	β-1,3/6-Glucan	ND

[a] CMC, carboxymethyl cellulose; PASC, phosphoric acid-swollen cellulose; BMCC, bacterial microcrystalline cellulose; pNPC, *p*-nitrophenyl-β-D-cellobioside.
[b] Specific activities are based on triplicate analysis for each substrate and account for abiotic hydrolysis. All standard deviations were less than ±10%.
[c] ND, no detectable activity.

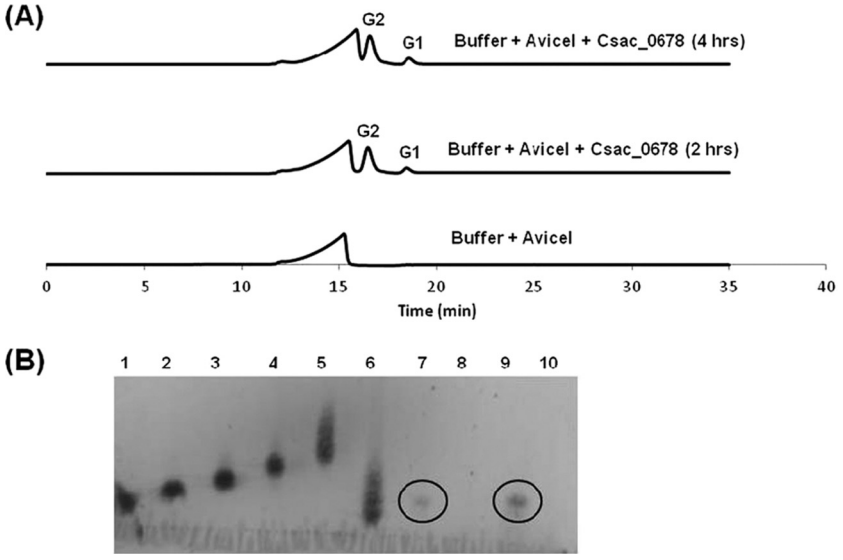

FIG 3 Hydrolysis of cellulose by Csac_0678 at 70°C. (A) HPLC (RI) analysis shows glucose and cellobiose produced after 2- and 4-h incubations of enzyme with Avicel (100 μm). (B) TLC for incubation of Csac_0678 for 4 h on two different particle sizes of cellulose (20 and 100 μm). TLC standards: lane 1, glucose; lane 2, cellobiose; lane 3, cellotriose; lane 4, cellotetraose; lane 5, cellopentaose; lane 6, xylooligosaccharides. For incubation with Csac_0678, the following standards were used: lane 7, 100-μm cellulose; and lane 9, 20-μm cellulose. Controls were 100-μm cellulose (lane 8) and 20-μm cellulose (lane 8), in each case with no added enzyme.

(42), but no activity was observed on either xylan or mannan. RuCelA, cloned from a metagenomic library of yak rumen microorganisms, encoded a GH5 enzyme possessing xylanase (264.1 U·mg^{-1}) and endoglucanase (54.3 U·mg^{-1}) activity, but with no CBM (11). Additionally, RuCelA had no activity on Avicel but measurable activity on filter paper.

The crystalline cellulose-hydrolyzing capability of endoglucanases such as *T. maritima* Cel5A (41) and Csac_0678 likely arises from the presence of a CBM. However, Cel5F, an endoglucanase from *Saccharophagus degradans* contains only a GH5 domain, yet has measurable activity toward Avicel (6.6×10^{-4} U·mg^{-1}) and filter paper (6.26×10^{-4} U·mg^{-1}) (58). Hydrolysis products from Csac_0678 on Avicel, determined by TLC and HPLC, indicated that the enzyme generated predominantly cellobiose and some glucose (Fig. 3). Although cleaving cellobiose from crystalline cellulose is a cellobiohydrolase trait, longer oligosaccharides may have been initially released by Csac_0678, which were then converted to the disaccharide. X-ray diffraction analysis of Avicel incubated with Csac_0678 indicated that crystallinity of Avicel decreased after the hydrolysis (data not shown). Furthermore, Csac_0678 appears to produce cellotriose, cellobiose, and small amounts of glucose from cellohexaose, but no cellotetraose was detected (data not shown). Interestingly, Csac_0678 also cleaved xylan at multiple sites, generating xylobiose, xylotriose, and xylotetraose (Fig. 4), as well as, after prolonged incubation, small amounts of xylose. This supports the premise that Csac_0678 acts via an endo-type mechanism. Csac_0678 activity followed typical Michaelis-Menten kinetics for *p*-nitrophenyl β-D-cellobioside hydrolysis: at pH 5.0 and 75°C, V_{max} and K_m were 6.1 U mg^{-1} and 0.65 mM, respectively.

The processivity of Csac_0678 on filter paper was examined to determine if the enzyme functions as an exoglucanase, an endoglucanase, or a processive endoglucanase (28, 29). Exoglucanases cleave cellobiose moieties from the ends of the cellulose molecule, whereas endoglucanases attack the cellulose molecule at any accessible point, randomly cutting the β-1,4-linkages, and they dissociate, leaving reducing sugars on the substrate (28, 29). Therefore, the ratio of soluble to insoluble reducing sugars is large for exo-acting enzymes and small for endo-acting enzymes. The distribution of reducing sugars on filter paper compared to the supernatant was measured. The soluble/insoluble reducing sugar ratio of Csac_0678 was 0.6 $\pm$ 0.05, indicating that it is an endoglucanase. In a previous study, soluble/insoluble reducing sugar ratios for classical endo-acting GH5 enzymes from *S. degradans*, Cel5B, Cel5C, CelD, Cel5E, and Cel5F, were reported to be between 0.096 and 1.42 (58). The same study showed that the ratios for processive endo-acting GH5 enzymes Cel5G, Cel5H, and Cel5J were between 4.04 and 4.59. Exocellulases have larger ratios of soluble to insoluble reducing sugars. For example, 96% of the reducing sugars produced by an exocellulase, Cel48A, from *Thermobifida fusca* were found to be soluble (soluble/insoluble reducing sugar ratio of 24) (30).

CMC viscosity before and after GH addition can also be used to distinguish between endo- and exoglucanase activities (29). As a soluble form of cellulose, CMC, is a good substrate for endo-acting glucanases, as they randomly bind CMC, cleaving β-1,4-linkages, and then they dissociate from the cellulose molecule (59). This leads to a reduction in the viscosity of CMC solution. Exoglucanases do not reduce the viscosity of CMC solution, although they have low activity on CMC (59). Cellobiohydrolases Cel7A (formerly CBHI) and Cel6A (formerly CBHII), which are exo-acting enzymes from *Trichoderma reesei*, did not significantly reduce the viscosity of a CMC solution, although they degraded 6% of the CMC (29, 30). Processive endoglucanases decrease the viscosity of CMC solution as they attack the cellulose chain in a similar way to classical endoglucanases. However, they differ from other endoglucanases and exoglucanases since they processively cleave cellobiose units from the cellulose chain (58, 59). Cel48, an

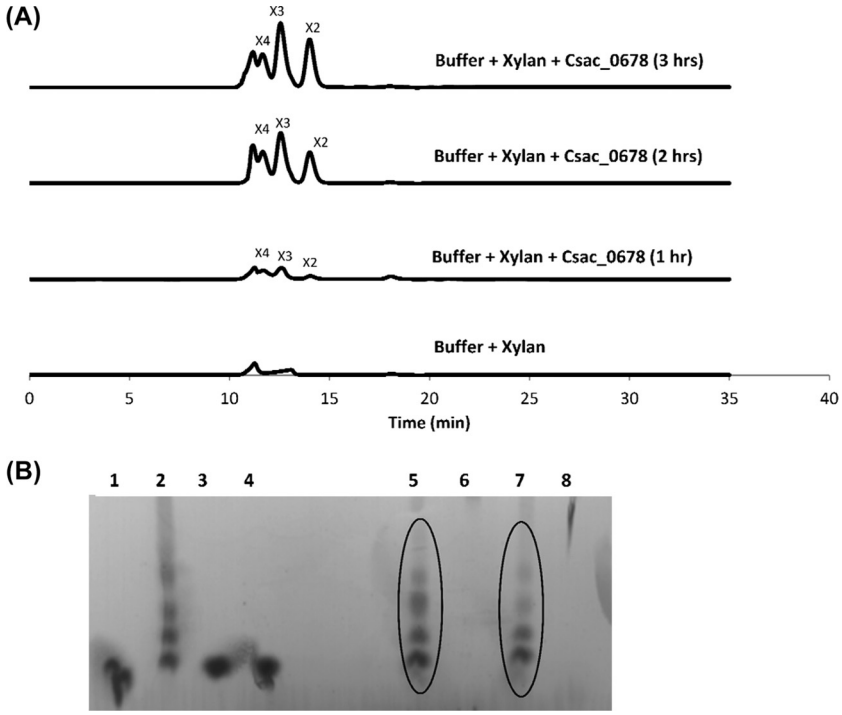

FIG 4 Hydrolysis of xylan by Csac_0678 at 75°C. (A) HPLC (RI) analysis shows xylobiose, xylotriose and xylotetraose released upon incubation with the enzyme. (B) TLC for incubation of xylan with the enzyme for 4 h. Standards: lane 1, xylose; lane 2, xylooligosaccharides; lane 3, arabinose; lane 4, glucose. The standards for reactions with Csac_0678 were as follows: lane 5, birchwood xylan; and lane 7, oat spelt xylan. The controls were birchwood xylan (lane 6) and oat spelt xylan (lane 8), in each case with no enzyme added.

exocellulase from *T. fusca*, reduced the viscosity of a CMC solution only by a few percent (64), while Cel5B from *Clostridium phytofermentas* decreased the viscosity of the CMC ~65%. Here, Csac_0678 reduced the viscosity of 1% (wt/vol) CMC solution by 40% in 1 h, further supporting the contention that it is an endoglucanase.

Biochemical analysis of Csac_0678 truncation mutants. Three different recombinant (C-terminal His tag) versions of Csac_0678, lacking the N-terminal signal peptide, were produced to examine the significance of the SLH and CBM domains (Fig. 1A): an intact version with SLH domains, CBM, and GH5; truncation mutant 1 (TM1) with GH5 and CBM28 domains (60 kDa) but no SLH domains; and truncation mutant 2 (TM2) with only the GH5 domain and neither CBM nor SLH domains (34 kDa). While intact Csac_0678 could hydrolyze Avicel, PASC, BMCC, xylan, and filter paper, the removal of the CBM28 domain virtually eliminated the activity of the enzyme (i.e., ~98% decrease in specific activity) on these substrates as well as on CMC (i.e., ~95% decrease in specific activity). CBMs assist hydrolysis by targeting, binding, and disrupting the cellulosic substrates (8). Binding of Csac_0678 to insoluble polysaccharides was determined by incubating the intact enzyme and Csac_0678-TM2 with Avicel. Figure 5A and C show intact Csac_0678 bound to Avicel to a greater extent than Csac_0678-TM2; by densitometry, 48 wt% and 13 wt% of Csac_0678 and Csac_0678-TM2 bound to Avicel, respectively. Quantitative binding study results agreed with the densitometry (Fig. 5C). Eliminating the SLH domain did not affect the activity of Csac_0678 (data not shown), indicating that SLH domains do not play a role in activity.

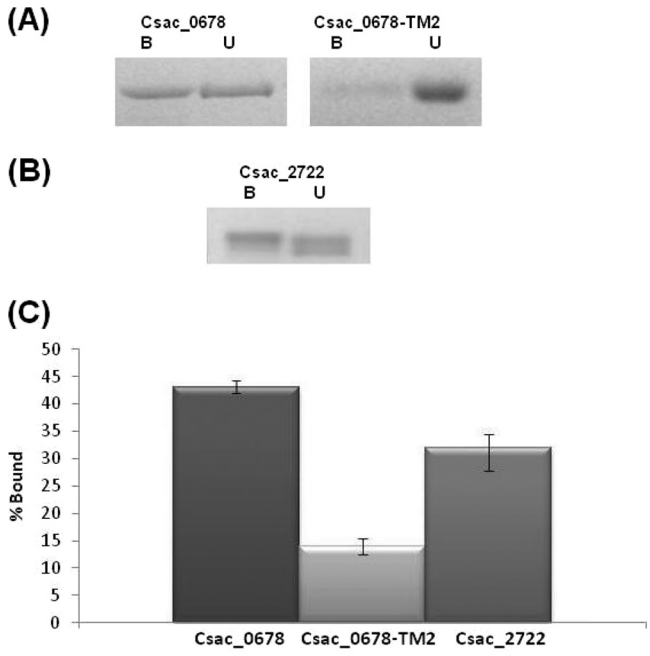

FIG 5 Csac_0678 binding to Avicel was decreased with the deletion of the CBM28 domain. (A and B) Shown are results from SDS-PAGE with bound (B) and unbound (U) fractions of Csac_0678 with and without CBM domains and Csac_2722, respectively, from the carbohydrate (Avicel) affinity experiment. (C) Binding of Csac_0678, Csac_0678 without the binding domain, and Csac_2722 to Avicel.

The Csac_2722 gene encodes a large polypeptide that binds to cellulose. The process by which cellulolytic microorganisms attack and hydrolyze crystalline cellulose is complex and likely involves steps not limited to hydrolytic biocatalysis. While most SLH domains lacking GH domains in *Caldicellulosiruptor* species appear to function as structural components of the S-layer, several contain putative binding domains (Table 4) (26). These proteins potentially represent an element of the complex carbohydrate recruitment strategy used by these bacteria. For example, the *Caldicellulosiruptor kristjanssonii* and *Caldicellulosiruptor lactoaceticus* genomes encode homologous 575-aa proteins (Calkr_0834 and Calla_1498, respectively) with SLH domains at the N terminus and which contain fibronectin type 3 (FN3) at the C terminus (21). *C. bescii* (formerly *Anaerocellum thermophilum*) and *Caldicellulosiruptor kronotskyensis* genomes also encode SLH-plus-FN3 domain proteins, Athe_0012 and Calkro_0014, respectively; these are each ~3,000 aa, the two largest ORFs in the *Caldicellulosiruptor* genomes sequenced to date. FN3 domains have been implicated in lignocellulose degradation. For example, Cel9A-90, a processive endoglucanase from *Thermobifida fusca*, has a GH9 catalytic domain and a CBM3c, followed by a fibronectin 3-like domain and a CBM2 (67). Deletion studies showed that the activity of Cel9A-90 on BMCC decreased (43% reduction in activity compared to the wild type) when the fibronectin domain was deleted (67). There are also SLH domain genes in *Caldicellulosiruptor* species that encode CBMs: Calkr_1989 in the *C. kristjanssonii* genome encodes a 326-aa protein, with three SLH domains at the N terminus followed by a CBM20 domain, while in *C. obsidiansis*, COB47_0167 is an 886-aa protein with three CBM27 domains upstream of the three SLH domains. Whether any of these SLH domain proteins in *Caldicellulosiruptor* species participate directly or indirectly in plant biomass deconstruction has not been determined.

Several *Caldicellulosiruptor* species (*C. kronotskyensis*, *C. lactoaceticus*, and *C. owensensis*) contain SLH domain proteins at the C terminus and also Big domains; these domains may play a role in biomass attachment, since proteins containing Ig-like domains from other microorganisms were reported to play roles in cell-cell adhesion, binding, and extracellular hydrolysis (21). Csac_2722 represents the largest ORF in the *C. saccharolyticus* genome, comprised of 2,593 amino acids, and is one of the largest annotated ORFs in *Caldicellusiruptor* species. Csac_2722 has no discernible catalytic domains but does contain two Big4 domains arranged in tandem with two galactose-binding domain-like domains (GBD), one cadherin-like domain, and three SLH domains. Cadherin-like domains appeared to contribute to protein-protein interactions and carbohydrate binding (19, 20). Polyclonal rabbit antibodies raised against a recombinant version of Csac_2722 lacking the SLH and cadherin-like domains recognized Csac_2722 attachment to the S-layer and cell membrane fraction in cells grown on switchgrass (data not shown). Pulldown experiments with Avicel as bait showed that Csac_2722 binds to Avicel (~32% bound) (Fig. 5B). Csac_2722 also exhibited activity on CMC (2.8 mM glucose equivalent of reducing sugars was released in 3 h), although no identifiable GH domain could be found in the Csac_2722 amino acid sequence. The fact that Csac_2722 bound to cellulose and that it is associated with the outer cell envelope suggested that it and perhaps some or all of the putative proteins listed in Table 4 contribute to lignocellulose conversion in *Caldicellulosiruptor* species.

TABLE 4 SLH domain proteins in *Caldicellulosiruptor* genomes that contain binding domains but no GH catalytic domains

Domains[a]	ORF	No. of amino acids
SLH-SLH-SLH-FN3-FN3-vWFA-vWFA-SH3	Athe_0012	3,027
	Calkro_0014	2,994
SLH-SLH-SLH-Big-(Big)	Athe_1839 Calow_1583 Csac_2381	575
	Calkro_0875	576
Big4-SLH-SLH-(SLH)	Calkro_0550 Calkro_0550	1,789
	Calow_1771	1,790
(SLH-SLH)-SLH-Big1-Tg	Calhy_0047	1,626
	Calla_2324	1,779
	COB47_0063	1,774
	Calkr_2463	1,179
SLH-SLH-SLH-FN3-FN3	Calla_1498 Calkr_0834	575
(Big)-(CBM20-CBM20)-SLH-SLH-SLH-CBM20	Calla_0367 COB47_0564 Calow_0484	1,097
	Calkr_1989	326
(CBM27)-CBM27-CBM27-CBM27-SLH-SLH-SLH	Calla_2176	1,088
	COB47_0167	886
SLH-SLH-SLH-CBM54-Tg	Calow_0034	1,774
Big3-Big3-Big3-Big3-PL9-SLH-SLH-SLH	Calow_2109	1,711
GBD-Big4-GBD-Big4-Cd-SLH-SLH-SLH	Csac_2722	2,593

[a] SLH, S-layer homology domain; FN3, fibronectin III domain; vWFA, von Willebrand factor domain; SH3, Src homology-3 domain; GBD, galactose binding domain-like; Big1 and Big4, Ig-like domains; CBM, cellulose binding domain, family 6, 20, 27; RB, ricin-like domain; PL, pectate lyase fold; Tg, transglutiminase-like domain; Cd, cadherin-like domain.

Conclusion. Consolidated bioprocessing (CBP) describes the comprehensive capacity of a microbial system not only to significantly deconstruct lignocellulose but also to convert the hydrolysis products to a biofuel. This metabolic capability no doubt in-

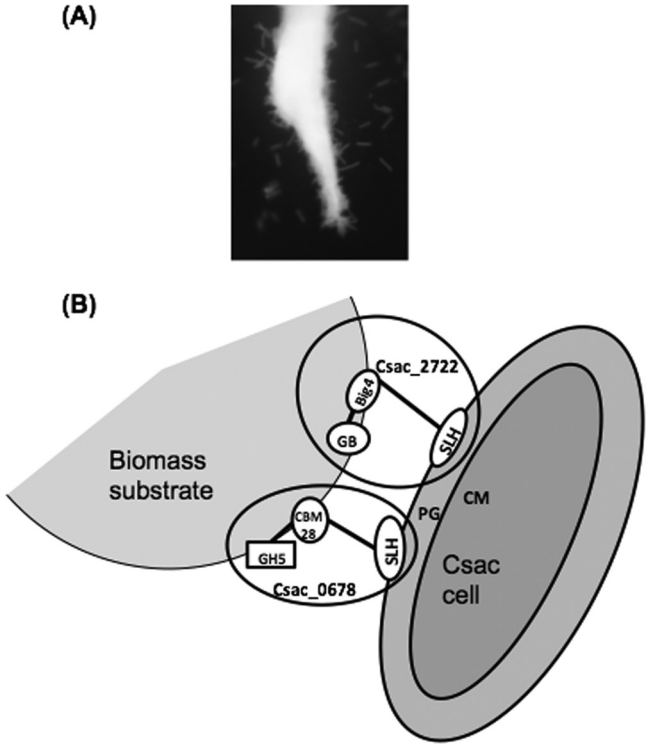

FIG 6 Proposed role of Csac_0678 and Csac_2722 in plant biomass deconstruction by *C. saccharolyticus*. (A) Epifluorescence micrograph (acridine orange stain) showing *C. saccharolyticus* (rods) attachment to acid-pretreated switchgrass particle. (B) Schematic representation of cell wall and carbohydrate attachment of Csac_0678 and Csac_2722. PG, peptidoglycan layer; CM, cell membrane; GB, galactose binding domain like; GH5, glycoside hydrolase family 5; CBM28, carbohydrate binding module family 28; SLH, surface layer homology domain.

volves subtle but important contributions which are more than enzymatic hydrolysis and involves factors not yet fully appreciated in CBP microorganisms. Certain SLH domain-containing proteins likely play a role in lignocellulose-degrading microorganisms that goes beyond providing structural integrity to the cell envelope. Two such proteins in the extremely thermophilic cellulolytic bacterium (Csac_0678 and Csac_2722) have the capacity to bind to crystalline cellulose and, in the case of Csac_0678, hydrolyze this substrate. Figure 6A shows the attachment of *C. saccharolyticus* cells to switchgrass, and Fig. 6B illustrates how Csac_0678 and Csac_2722 contribute to the deconstruction of plant biomass. Contributions from noncatalytic, carbohydrate binding proteins, which can associate with the S-layer, such as Csac_2722, may be important to CBP microorganisms, and this issue merits further attention. There is much still to understand about the complex process by which lignocellulose is degraded and utilized in natural environments that can be translated to current efforts to produce biofuels from renewable feedstocks.

ACKNOWLEDGMENTS

R.M.K. acknowledges support from the U.S. National Science Foundation (CBT0617272) and the Bioenergy Science Center (BESC), a U.S. DOE Bioenergy Research Center supported by the Office of Biological and Environmental Research.

We also wish to thank Allen Foegeding (Department of Food Science, NCSU), Sunkyu Park (Department of Forest Biomaterials, NCSU), and Christina Tang and Saad Khan (Department of Chemical and Biomolecular Engineering, NCSU) for their help with experimental aspects of this effort. Jeffrey Zurawski (Department of Chemical and Biomolecular Engineering, NCSU) helped with the graphics. The gift of bacterial microcrystalline cellulose from David Wilson and Maxim Kostylev (Cornell University) was also appreciated.

REFERENCES

1. **Andrews G, Lewis D, Notey J, Kelly R, Muddiman D.** 2010. Part I. Characterization of the extracellular proteome of the extreme thermophile *Caldicellulosiruptor saccharolyticus* by GeLC-MS2. Anal. Bioanal. Chem. **398**:377–389.
2. **Bayer EA, Chanzy H, Lamed R, Shoham Y.** 1998. Cellulose, cellulases and cellulosomes. Curr. Opin. Struct. Biol. **8**:548–557.
3. **Bayer EA, Kenig R, Lamed R.** 1983. Adherence of *Clostridium thermocellum* to cellulose. J. Bacteriol. **156**:818–827.
4. **Bayer EA, Setter E, Lamed R.** 1985. Organization and distribution of the cellulosome in *Clostridium thermocellum.* J. Bacteriol. **163**:552–559.
5. **Bayer EA, Shimon LJ, Shoham Y, Lamed R.** 1998. Cellulosomes—structure and ultrastructure. J. Struct. Biol. **124**:221–234.
6. **Blumer-Schuette SE, Kataeva I, Westpheling J, Adams MW, Kelly RM.** 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr. Opin. Biotechnol. **19**:210–217.
7. **Blumer-Schuette SE, Lewis DL, Kelly RM.** 2010. Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor.* Appl. Environ. Microbiol. **76**:8084–8092.
8. **Boraston AB, Bolam DN, Gilbert HJ, Davies GJ.** 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. **382**:769–781.
9. **Cann IK, Kocherginskaya S, King MR, White BA, Mackie RI.** 1999. Molecular cloning, sequencing, and expression of a novel multidomain mannanase gene from *Thermoanaerobacterium polysaccharolyticum.* J. Bacteriol. **181**:1643–1651.
10. **Cantarel BL, et al.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. **37**: D233–D238.
11. **Chang L, et al.** 2011. Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. Appl. Microbiol. Biotechnol. **90**:1933–1942.
12. **Chauvaux S, Matuschek M, Beguin P.** 1999. Distinct affinity of binding sites for S-layer homologous domains in *Clostridium thermocellum* and *Bacillus anthracis* cell envelopes. J. Bacteriol. **181**:2455–2458.
13. **Chhabra SR, Shockley KR, Ward DE, Kelly RM.** 2002. Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium *Thermotoga maritima* grown on glucan- and mannan-based polysaccharides. Appl. Environ. Microbiol. **68**:545–554.
14. **de Vrije T, et al.** 2007. Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus.* Appl. Microbiol. Biotechnol. **74**:1358–1367.
15. **Dey D, Hinge J, Shendye A, Rao M.** 1992. Purification and properties of extracellular endoxylanase from alkalophilic thermophilic *Bacillus* sp. Can. J. Microbiol. **38**:436–442.
16. **Engelhardt H, Peters J.** 1998. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions. J. Struct. Biol. **124**:276–302.
17. **Ferner-Ortner-Bleckmann J, et al.** 2009. The high-molecular-mass amylase (HMMA) of *Geobacillus stearothermophilus* ATCC 12980 interacts with the cell wall components by virtue of three specific binding regions. Mol. Microbiol. **72**:1448–1461.
18. **Fontes CM, Gilbert HJ.** 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu. Rev. Biochem. **79**:655–681.
19. **Fraiberg M, Borovok I, Bayer EA, Weiner RM, Lamed R.** 2011. Cadherin domains in the polysaccharide-degrading marine bacterium *Saccharophagus degradans* 2-40 are carbohydrate-binding modules. J. Bacteriol. **193**:283–285.
20. **Fraiberg M, Borovok I, Weiner RM, Lamed R.** 2010. Discovery and characterization of cadherin domains in *Saccharophagus degradans* 2-40. J. Bacteriol. **192**:1066–1074.
21. **Fraser JS, Yu Z, Maxwell KL, Davidson AR.** 2006. Ig-like domains on

bacteriophages: a tale of promiscuity and deceit. J. Mol. Biol. **359**: 496–507.
22. **Geslin C, et al.** 2003. PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "*Pyrococcus abyssi.*" J. Bacteriol. **185**: 3888–3894.
23. **Hashimoto H.** 2006. Recent structural studies of carbohydrate-binding modules. Cell. Mol. Life Sci. **63**:2954–2967.
24. **Henrissat B, et al.** 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc. Natl. Acad. Sci. U. S. A. **92**:7090–7094.
25. **Herbel Z, et al.** 2010. Exploitation of the extremely thermophilic *Caldicellulosiruptor saccharolyticus* in hydrogen and biogas production from biomasses. Environ. Technol. **31**:1017–1024.
26. **Hunter S, et al.** 2009. InterPro: the integrative protein signature database. Nucleic Acids Res. **37**:D211–D215.
27. **Igarashi K, Ishida T, Hori C, Samejima M.** 2008. Characterization of an endoglucanase belonging to a new subfamily of glycoside hydrolase family 45 of the basidiomycete *Phanerochaete chrysosporium.* Appl. Environ. Microbiol. **74**:5628–5634.
28. **Irwin D, et al.** 1998. Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. J. Bacteriol. **180**:1709–1714.
29. **Irwin DC, Spezio M, Walker LP, Wilson DB.** 1993. Activity studies of eight purified cellulases: specificity, synergism, and binding domain effects. Biotechnol. Bioeng. **42**:1002–1013.
30. **Irwin DC, Zhang S, Wilson DB.** 2000. Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca.* Eur. J. Biochem. **267**:4988–4997.
31. **Ito Y, et al.** 2003. Cloning, expression, and cell surface localization of *Paenibacillus* sp. strain W-61 xylanase 5, a multidomain xylanase. Appl. Environ. Microbiol. **69**:6969–6978.
32. **Jindou S, et al.** 2006. Novel architecture of family-9 glycoside hydrolases identified in cellulosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum.* FEMS Microbiol. Lett. **254**:308–316.
33. **Kadar Z, et al.** 2004. Yields from glucose, xylose, and paper sludge hydrolysate during hydrogen production by the extreme thermophile *Caldicellulosiruptor saccharolyticus.* Appl. Biochem. Biotechnol. **113-116**: 497–508.
34. **Kosugi A, Murashima K, Tamaru Y, Doi RH.** 2002. Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. J. Bacteriol. **184**:884–888.
35. **Lehtio J, et al.** 2003. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. Proc. Natl. Acad. Sci. U. S. A. **100**:484–489.
36. **Maki M, Leung KT, Qin W.** 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. Int. J. Biol. Sci. **5**:500–516.
37. **Martinez D, et al.** 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat. Biotechnol. **26**:553–560.
38. **Matuschek M, Sahm BGK, Bahl H.** 1994. Pullulanase of *Thermoanaerobacterium thermosulfurigenes* EM1 (*Clostridium thermosulfurogenes*): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface. J. Bacteriol. **176**: 3295–3302.
39. **Miller GL.** 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. **31**:426–428.
40. **Mizuno M, et al.** 2004. Structural insights into substrate specificity and function of glucodextranase. J. Biol. Chem. **279**:10575–10583.
41. **Pereira JH, et al.** 2010. Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima.* J. Struct. Biol. **172**:372–379.
42. **Posta K, Beki E, Wilson DB, Kukolya J, Hornok L.** 2004. Cloning, characterization and phylogenetic relationships of cel5B, a new endoglucanase encoding gene from *Thermobifida fusca.* J. Basic Microbiol. **44**: 383–399.
43. **Rainey FA, et al.** 1994. Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. FEMS Microbiol. Lett. **120**:263–266.
44. **Raman B, et al.** 2009. Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS One **4**:e5271.
45. **Ries W, Hotzy C, Schocher I, Sleytr UB, Sara M.** 1997. Evidence that the N-terminal part of the S-layer protein from *Bacillus stearothermophilus* PV72/p2 recognizes a secondary cell wall polymer. J. Bacteriol. **179**: 3892–3898.
46. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
47. **Sara M.** 2001. Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria? Trends Microbiol. **9**:47–50.
48. **Saul DJ, et al.** 1990. celB, a gene coding for a bifunctional cellulase from the extreme thermophile "*Caldocellum saccharolyticum.*" Appl. Environ. Microbiol. **56**:3117–3124.
49. **Sleytr UB.** 1976. Self-assembly of the hexagonally and tetragonally arranged subunits of bacterial surface layers and their reattachment to cell walls. J. Ultrastruct. Res. **55**:360–377.
50. **Te'o VS, Saul DJ, Bergquist PL.** 1995. celA, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum.* Appl. Microbiol. Biotechnol. **43**:291–296.
51. **Tschiggerl H, et al.** 2008. Exploitation of the S-layer self-assembly system for site directed immobilization of enzymes demonstrated for an extremophilic laminarinase from *Pyrococcus furiosus.* J. Biotechnol. **133**:403–411.
52. **Tsukimoto K, et al.** 2010. Recognition of cellooligosaccharides by a family 28 carbohydrate-binding module. FEBS Lett. **584**:1205–1211.
53. **van de Werken HJ, et al.** 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus.* Appl. Environ. Microbiol. **74**:6720–6729.
54. **Vanfossen AL, Lewis DL, Nichols JD, Kelly RM.** 2008. Polysaccharide degradation and synthesis by extremely thermophilic anaerobes. Ann. N. Y. Acad. Sci. **1125**:322–337.
55. **Vanfossen AL, Ozdemir I, Zelin SL, Kelly RM.** 2011. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus.* Biotechnol. Bioeng. **108**:1559–1569.
56. **Vanfossen AL, Verhaart MR, Kengen SM, Kelly RM.** 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl. Environ. Microbiol. **75**:7718–7724.
57. **Vuong TV, Wilson DB.** 2009. Processivity, synergism, and substrate specificity of *Thermobifida fusca* Cel6B. Appl. Environ. Microbiol. **75**: 6655–6661.
58. **Watson BJ, Zhang H, Longmire AG, Moon YH, Hutcheson SW.** 2009. Processive endoglucanases mediate degradation of cellulose by *Saccharophagus degradans.* J. Bacteriol. **191**:5697–5705.
59. **Wilson DB.** 2011. Microbial diversity of cellulose hydrolysis. Curr. Opin. Microbiol. **14**:259–263.
60. **Winterhalter C, Liebl W.** 1995. Two extremely thermostable xylanases of the hyperthermophilic bacterium *Thermotoga maritima* MSB8. Appl. Environ. Microbiol. **61**:1810–1815.
61. **Xiao Z, Storms R, Tsang A.** 2005. Microplate-based carboxymethylcellulose assay for endoglucanase activity. Anal. Biochem. **342**:176–178.
62. **Xiao Z, Storms R, Tsang A.** 2004. Microplate-based filter paper assay to measure total cellulase activity. Biotechnol. Bioeng. **88**:832–837.
63. **Xu Q, et al.** 2004. Architecture of the *Bacteroides cellulosolvens* cellulosome: description of a cell surface-anchoring scaffoldin and a family 48 cellulase. J. Bacteriol. **186**:968–977.
64. **Zhang XZ, et al.** 2010. The noncellulosomal family 48 cellobiohydrolase from *Clostridium phytofermentans* ISDg: heterologous expression, characterization, and processivity. Appl. Microbiol. Biotechnol. **86**:525–533.
65. **Zhang YH, Cui J, Lynd LR, Kuang LR.** 2006. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. Biomacromolecules **7**:644–648.
66. **Zhang YH, Lynd LR.** 2005. Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis. Biomacromolecules **6**:1510–1515.
67. **Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB.** 2004. Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. Biochemistry **43**:9655–9663.