ABSTRACT

OZDEMIR, INCI. Role of S-layer Homology Domain Proteins in Plant Biomass Conversion to Biofuels by Extremely Thermophilic *Caldicellulosiruptor* species. (Under the direction of Robert M. Kelly.)

Second-generation biofuels require the conversion of lignocellulose from agricultural waste, non-edible parts of crops, and grasses into fermentable sugars. Lignocellulose is composed of pectin, hemicellulose, cellulose and lignin, all of which form a matrix that must be penetrated and deconstructed by physical, chemical and biological methods to facilitate conversion to biofuels. Certain microorganisms have the capacity to extract the carbohydrate content of lignocellulose through the coordinated action of glycoside hydrolases and other metabolic processes, which together can contribute to biofuel production. Of interest here are extremely thermophilic bacteria from the genus *Caldicellulosiruptor*. One of the intriguing features of these bacteria is that they do not produce cellulosomes but rather produce multi-domain (hemi)cellulolytic enzymes. Some of these biocatalysts contain S-layer homology (SLH) domains, anchoring the enzymes to the cell envelope, and also have carbohydrate-binding domains (CBM), which bind the enzyme to the substrate. This can be a very efficient way of degrading plant biomass since enzymes are localized in close proximity of the cell where control its local microenvironment is possible.

Six different groups of SLH-domain proteins with GH domains could be identified in the genome sequences of *Caldicellulosiruptor* species. Representatives from each group were produced recombinantly in *E. coli* and examined with respect to biochemical and biophysical properties. Csac_0678 (bifunctional

cellulase/xylanase from *C. saccharolyticus* with homologs in all eight genome-sequenced *Caldicellulosiruptor* species), Calkr_2245 (xylanase from *C. kristjanssonii*), Calhy_0060 (β-1,3-1,4-glucanase from *C. hydrothermalis*), Calkro_0111 (bifunctional β-1,3 glucanase/β-1,3-1,4-glucanase from *C. kronotskyensis*), Calhy_1629 (β-xylosidase from *C. hydrothermalis*), and Calhy_2383 (α-1,3 glucanase from *C. hydrothermalis*) were identified in the sequenced genomes of *Caldicellulosiruptor* species as the multi-domain SLH-domain containing enzymes with different conserved domain organizations. These enzymes were characterized biochemically (except Calhy_2383 since no plant biomass-based substrate could be identified) and all exhibited hemicellulolytic activity; Csac_0678 had both hemicellulolytic and cellulolytic activity. The optimum temperature and pH of these enzymes were found to be between 68-75°C and 5-7, respectively. Csac_2722, encoded in a 2593 amino acid open reading frame (ORF) in *C. saccharolyticus*, did not contain a GH domain but had several sugar binding domains and SLH domains, which were also characterized. Csac_2722 and Csac_0678 bound to complex carbohydrates, such as crystalline cellulose and xylan. A truncation mutant of Csac_0678, lacking its CBM28 binding domain, did not bind to the carbohydrates, confirming the necessisity of this motif in anchoring to plant biomass substrates. Polyclonal antibodies raised against Csac_0678 (without the SLH domains) enabled confocal microscopy and fluorescence imaging analysis that confirmed that Csac_0678 was localized on the cell surface of *C. saccharolyticus*. The 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.

The results from this study indicate that SLH-domain proteins with GHs and CBM motifs are important for *Caldicellulosiruptor* species in deconstructing lignocellulose. From the biotechnological perspective, these SLH-domain containing glycoside hydrolases are potentially important in reducing the recalcitrance of crystalline cellulose and facilitating its conversion to fermentable sugars. Furthermore, this work revealed that *Caldicellulosiruptor* species utilize cell envelope-associated enzymes and proteins, in addition to GHs present in the cytoplasm and secretome, to process complex polysaccharides.

Role of S-layer Homology Domain Proteins in Plant Biomass Conversion to Biofuels by Extremely Thermophilic *Caldicellulosiruptor* species

by
Inci Ozdemir

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Chemical Engineering

Raleigh, North Carolina

2012

APPROVED BY:

_____	_____
Robert M. Kelly Chair of Advisory Committee	David F. Ollis
_____	_____
Amy M. Grunden	Balaji M. Rao

DEDICATION

To my parents, my aunt and my uncle

BIOGRAPHY

Inci Ozdemir was born and raised in Izmir, Turkey. She received a Bachelor of Science degree in Chemical Engineering with a concentration in Biology in 2006 at the Middle East Technical University (Ankara, Turkey). After finishing her undergraduate education with high honors she moved to the USA to pursue a Doctorate Degree in the Department of Chemical Engineering at the North Carolina State University with a minor in Biology. After completing her PhD studies in Robert M. Kelly's group, Inci will be joining Pfizer, Inc. as a Senior Scientist to utilize her expertise in Culture Process Development group in June, 2012.

ACKNOWLEDGMENTS

First of all I would like to thank Dr. Robert M. Kelly for giving me the opportunity to work in his lab and for his guidance and support throughout the years. I also would like to thank him for sending me to various international and national scientific conferences to present my work and learn from other scientists. I will never forget the Thermophiles conference and the Yellowstone National Park trip I took with all the Kelly group members. I also would like to thank him for his help and support while I was job hunting; it was very much appreciated.

I truly enjoyed working in the Kelly lab with its recent members: Jaspreet Notey, Andrew Frock, Arpan Mukherjee, Hong Lian, Aaron Hawkins, Andrew Loder, Garrett Wheaton, Jeff Zurawski, and Yejun Han. Special thanks go to Sara Blumer-Schuette - she was a great mentor, teacher, friend, and a great problem solver over the past four years. I also would like to thank the past members of the Kelly lab: Morgan Harris for his friendship and his help in my first years in the lab, Amy VanFossen and Jason Nichols for their mentoring. Thanks to my international friends: Nimish Gera, Mahmud Hussain, Haiyan Liu, Prasenjit Sarkar, Jaspreet Notey (Thank you again!), and Arpan Mukherjee (Thank you again!) for understanding me, for listening to me, and for the parties we organized where we danced and ate wonderful homemade international food. Grad school would not be

the same without you all. I also would like to thank Dr. Saad Khan and Sandra Bailey for their help and support.

Last but not least, I would like to thank my parents for their unfailing love and support. Thanks to my Turkish friends: Hande Ertorer, Funda Gunes, and Nurcan Tezcan Ergin for their endless support and special thanks to Carlos Reyes for his love, support, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

EXTREMOPHILES

Inci Ozdemir and Robert M. Kelly

Department of Chemical and Biomolecular Engineering

North Carolina State University

Raleigh, NC 27695-7905

Published Online: 15 SEP 2010

ENCYCLOPEDIA OF CATALYSIS

I.T. Horváth, Editor-in-Chief

John Wiley & Sons, Inc.

DOI:10.1002/0471227617.eoc089.pub2

ABSTRACT

Extremophiles are microorganisms with the ability to survive under extreme environmental conditions, including geothermal and arctic waters, glacial ices, deserts, saline lakes, and acidic, sulfurous hot springs. Adaptation of extremophiles to harsh conditions and unique stability of the enzymes (extremozymes) from these organisms have recently attracted a great deal of attention. Extremozymes have been replacing enzymes, which do not cope with harsh conditions, in many industries, such as pharmaceutical, food, chemical, laundry detergents, and bioremediation. Furthermore, there is an increasing demand of novel applications for these biocatalysts, working optimally over a range of extreme conditions, such as high temperature, high-salinity, and high alkalinity. Developments in gene discovery and gene expression technologies hopefully will expand the exploitation of these biocatalysts. In this study, different types of extremophiles are introduced and the adaptations to extreme conditions for each condition are indicated. The biocatalysts from overlooked extremophiles and the applications of these biocatalysts are highlighted.

1. Extremophilic Microorganisms and Enzymes

Over the past several decades, it has become clear that certain microorganisms, referred to as **extremophiles** (“extreme-loving”), thrive in environments that are otherwise prohibitive to life (45). Such biotopes can be thermal or cold, highly saline, hyperbaric or hypobaric, have high levels of ionizing radiation, be abnormally dry, contain high levels of heavy metals, or be virtually devoid of nutrients (see Table 1.1) (44, 45). Extremophiles are arguably the earliest forms of life on earth (and, perhaps, on other solar bodies). Indeed, on early Earth, some 4 billion years ago, the upper part and surface of the oceans were not protected from UV radiation due to the absence of an ozone layer (45). Furthermore, the “hot start” scenario for origins of life proposes that thermophiles (“heat-loving” extremophiles) were the original inhabitants of this planet. This theory is supported by 16S rRNA phylogenetic trees, which indicate hyperthermophiles are ancestors to all life (77). A better understanding of molecular mechanisms for adaptation to extreme conditions is crucial to explain evolution and the limits of life on Earth, and, perhaps, beyond this planet. Furthermore, extremophiles have substantial biotechnological potential, a fact that no doubt has been the driving force behind the many genome sequencing projects, completed and now underway, focusing on these microorganisms (29).

There are many existing and potential industrial applications of extremophilic enzymes (some of the important ones are summarized in Table 1.2), which take

advantage of their capacity to thrive in harsh conditions, relative to their mesophilic (i.e., function under moderate conditions) counterparts. For example, the detergent industry makes extensive use of alkaline-stable proteases, accounting for approximately 30% of global enzyme production (43). Enzymes from extremophiles have been used for bioremediation heavy metals and radio-nuclides from waste waters (34), as well as for biodegrading hydrocarbons in polar soils (1). It is likely that extremophilic enzymes will find wide use in applications that strategically use their intrinsic robustness as biocatalysts.

2. Structure, Function, and Stability of Extremophilic Enzymes

While there are many examples of enzymes for conventional microbial sources being used under extreme conditions, the discussion here will focus on those biocatalysts that are intrinsically produced by extremophilic microorganisms.

2.1. Enzymes from Extreme Thermophiles

2.1.1. Extreme thermophiles

Moderately thermophilic ("heat-loving") microorganisms grow optimally at temperatures between 50 and 70^{o}C (89), whereas extreme thermophiles have optimal growth temperature of 70^{o}C and above (10, 89). A subgroup of the extreme thermophiles contains the hyperthermophiles (T_{opt} ≥ 80^{o}C) that are mostly from the Archaea, which along with the Bacteria and Eucarya, represent of the three domains of life. The genera *Thermotoga* and *Aquifex* are the only representative members of

the domain Bacteria that belong to hyperthermophiles (39, 48). Extreme thermophiles, the first of which were discovered in 1967 (14), can be found in diverse habitats on Earth: terrestrial (hot springs, geysers, and solfatares), marine (deep-sea hydrothermal vent sites), subsurface (petroleum reservoirs and geothermally-heated lakes and aquifers), and anthropogenic (industrial process environments and thermal effluent from power plants) (15, 24, 29). Figure 2.5 shows Yellowstone National Park where many extreme thermophiles have been isolated.

Like most microorganisms, extreme thermophiles are typically found in natural communities. Mimicking their natural biotopes is important for establishing growth of extremophiles under laboratory conditions, and this task can present a significant challenge. Consequently, native enzymes from these microorganisms were not obtained in purified form until 1980 (24), given the difficulty of generating sufficient biomass for protein purification efforts (55). Fortunately, there has been significant success with producing these proteins in recombinant forms (40), thus by-passing the need to cultivate extreme thermophiles at large-scale (86).

2.1.2. Adaptations to extreme temperatures

There are several molecular mechanisms used by extreme thermophiles to cope with destabilizing forces at high temperatures. When nucleic acids are heated, their molecular architecture is impacted, and chemical degradation of their molecular building blocks (i.e. sugars, purine and pyrimidine bases) can occur (39). To offset this, thermal stabilization of DNA is accomplished through several complementary

strategies: high concentrations of monovalent cations (Na^{+}, K^{+}, and Mg^{+2}) to protect double-stranded DNA (dsDNA) against chemical degradation of phosphodiester bonds; presence of thermostable histone-like nucleoid-binding proteins that help to preserve the duplex structure of DNA; and, a particular type of DNA topoisomerase, reverse gyrase, which catalyzes the ATP-dependent introduction of positive supercoils into DNA. As a result of positive supercoiling, the number of topological links between the two strands of a closed DNA molecule increases. Hence, this structure is more stable at high temperature (11, 39). RNA in “heat-loving” microorganisms also has been adapted to function at elevated temperatures by increased G+C content, compact self-folding, and, the absence of thermolabile modified nucleotides (39).

Thermostable proteins from extreme thermophiles have some common features that distinguish them from their mesophilic counterparts. For example, these proteins have more charged amino acid (Glu, Arg, and Lys), which provides the framework for ion pairings and extensive thermostabilizing networks (57). Extremely thermophilic proteins also have relatively rigid structures, arising from their apolar cores, smaller surface/volume ratios, and ion bonding on the surface. Moreover, disulfide bridges, aromatic interactions and hydrogen bonds also play a role in protein stability, although these features are not necessarily more pronounced in all proteins from extreme thermophiles (24). Rather than one specific feature, thermostability of extremely thermophilic proteins stems from the cumulative effect of

a large number of subtle stabilizing interactions and forces integrated over the entire macromolecule (72).

2.1.3. Industrial applications of extremely thermophilic enzymes

Due to their extreme thermostability and resistance to chemical denaturants, detergents, chaotropic agents, and organic solvents, enzymes from extreme thermophiles have received a great deal of attention in the biotechnology and chemical industries. At elevated temperatures, viscosity decreases, the diffusion coefficients of substrates increase, and the solubility of substrates and products increase (24). Furthermore, the risk of contamination by mesophilic microorganisms leading to undesired complications and by-products, is reduced. All of these factors lead to potentially higher overall reaction rates, and hence, higher yields. However, it should be mentioned that enzymes from extreme thermophiles do not necessarily have higher intrinsic catalytic efficiencies than their mesophilic counterparts. This is no doubt the result of cellular strategies to regulate metabolic pathway fluxes so that enzyme kinetics are in line with physiological function.

Another advantage of extremely thermophilic enzymes is that when they are produced recombinantly in mesophilic hosts, heat-treatment serves as an efficient method to purify the target enzyme from the host's proteins, which readily denature at higher temperatures (24). This strategy can be used to great advantage for biocatalyst production.

The availability of genome sequence data has created the possibility of identifying an extremely thermophilic counterpart to any mesophilic enzyme or protein with industrial potential. Thus, robust replacement biocatalysts can be considered for a range of current enzyme applications. Of course, new biocatalytic processes that advantageously utilize high temperature enzymes are of considerable interest. Discussed below are several examples of such cases.

Starch processing: Extremely thermophilic enzymes have found strategic use in starch processing (24). Starch from various plant sources (mostly corn) can be converted into higher value products, such as dextrins, glucose, fructose, and trehalose, by starch-hydrolyzing enzymes (α-amylase, α-glucosidase, pullulanase, and cyclodextrin) (90). Figure 1.1 and Figure 1.2 show the types of glycosidic bonds of amylose and amylopectin, respectively, that are cleaved enzymatically during starch hydrolysis. In order to liquefy starch and make it accessible to enzymatic hydrolysis, high processing temperatures are crucial. Thermostable amylases, pullulanases, and α-glucosidases act synergistically for starch conversion at high temperature. For instance, an amylolytic enzyme, with both α-amylase and cyclodextrin-hydrolyzing features, from the hyperthermophile *Pyrococcus furiosus* has an optimum pH and temperature of 4.5 and 90^{o}C, respectively, making it an attractive candidate for starch processing (90). Similarly, another hyperthermophile, *Thermococcus aggregans* produces a novel pullulanase that can hydrolyze α-1,4 as well as α-1,6 glycosidic linkages in pullulan (67). This enzyme is most active at 90^{o}C

and pH 6.5, retains activity after incubation at 90^{o}C for 4 hours, and has a half-life of 2.5 hours at 100^{o}C.

Glucoamylase, an exo-acting starch-degrading enzyme, plays a major role in the production of crystalline glucose or glucose syrup that can be further processed into high fructose corn syrup. A glucoamylase from *Thermoplasma acidophilum* ($T_{growth\ opt}$= 60^{o}C, $pH_{growth\ opt}$= 1-2) is an ideal candidate for this process, since high temperatures and low pH is important for starch processing (25). Trehalose, a dissacharide which stabilizes enzymes, antibodies, vaccines, and hormones can be produced from dextrins using enzymes from hyperthermophilic *Sulfolobus* species in a continuous bioreactor, achieving a final conversion of 90% (26). Given the promise that amylolytic enzymes from extremophiles have for α-glucan bioprocessing, these continue to be evaluated for strategic uses in the starch and related industries.

Cellulolytic enzymes: Lignocellulosic biomass conversion to fermentable sugars, which are then converted into bioethanol or biohydrogen, is of a significant interest given the increasing fossil fuel demand and need for alternative energy sources. Non-edible residues from food crops, as well as agricultural, industrial, and forest by-products containing lignocelluloses, can be converted into fermentable sugars through the action of a range of enzymes, especially cellulases. Cellulose degradation requires the synergetic action of many cellulolytic enzymes, including endoglucanases, exoglucanases and β-glucosidases (Fig 1.3 and Fig 1.4). Due to

the recalcitrance of cellulose to hydrolysis, high temperatures have been proposed to improve biofuel yields (10), making, thermostable and thermoactive cellulases attractive candidates for lignocellulosic biomass deconstruction. Cellulases also have uses beyond bioenergy for food processing, textile manufacturing, laundry detergents, and in the pulp and paper industry (26).

Cellulolytic enzymes have been identified in extremely thermophilic bacteria. *Anaerocellum thermophilum* (T_{opt}= 75^{o}C) and *Caldicellulosiruptor saccharolyticus* (T_{opt}= 70-75^{o}C) are bacteria that each produce a multidomain cellulolytic enzyme, referred to as CelA (190 kDa), that is active towards carboxymethylcellulose (CMC), water-soluble β-glucans, and crystalline cellulose forms, including Avicel (10, 82, 95). Other high temperature microorganisms also produce cellulolytic enzymes with industrial potential. For example, *Thermotoga maritima* MSB8 (T_{opt} = 80^{o}C) produces an extremely thermophilic cellobiohydrolase that was shown to hydrolyze Avicel, CMC, β-glucan, and *p*-nitrophenyl-β-D-cellobioside (80). It is interesting that no cellulosome, a multi-protein complex that catalyzes cellulose deconstruction in mesophilic bacteria, has yet to be identified in extreme thermophiles (10), but this may change as thermal, natural environments are probed for cellulolytic microorganisms.

Other β-glucan polysaccharides can also be processed using extremely thermophilic enzymes. Chitin, a biopolymer consisting of N-acetylglucosamine units, is found in the walls of higher fungi, and in the exoskeleton of insects and arthropods

(85). Oligosaccharides resulting from chitin degradation are valuable products functioning as antibacterial agents, elicitors of lysozyme inducers and immunoenhnancers (20). Chitin-degrading enzymes, or chitinases, are classified either as endo- or exo-acting, depending on whether they degrade chitin internally or from the end of the polymer, respectively (85). The hyperthermophilic archaea *Thermococcus kodakaraensis* KOD1 and *Pyrococcus furiosus* (36) produce multidomain chitinases; these enzymes have several different carbohydrate binding domains which bind to the substrate, and has two different catalytic regions (36, 68).

Hemicellulases: There are also other thermostable polysaccharide-degrading enzymes produced by extreme thermophiles, of great commercial importance, given that the solubility of hemicelluloses in aqueous solutions is significantly improved at high temperature, enzymes that can hydrolyze these polysaccharides, xylanases in particular, have been a focus of discovery efforts (54).

Xylanases have been widely used in Scandinavia, North America, and China in the animal feed and pulp and paper industries (22, 26). Degradation of xylan, which consists of a poly-β-1,4-linked xylose backbone decorated with various types of side chains, requires the synergistic action of multiple xylanolytic enzymes, including endo-β-1,4-xylanases and β-1,4-xylosidases. To remove the acetyl, arabinofuranosyl and 4-*O*-methylglucuronyl groups that are linked to the xylose backbone, arabinofuranosidases, α-glucuronidases and acetylxylan esterases are

required (26, 58). *T. maritima* MSB8 (T_{opt} = 80^{o}C) produces two extremely thermostable endoxylanases: XynA (120 kDa) and XynB (40 kDa), when this bacterium is grown in xylan or xylose-containing media (58). XynB is found in the cell periplasmic fraction of the cell as well as in the culture supernatant, whereas XynA attaches to the cell wall by means of a hydrophobic peptide (58).

Xylanases offer an environmentally-benign improvement to chemical bleaching in the pulp and paper industry. In the pulping process, lignocellulose fibers are degraded, so that most of the lignin can be removed (47). In the multi-step bleaching process to whiten the ultimate paper product, the lignin that is still retained after the pulping process, is extracted. Enzymatic bleaching (or bio-bleaching) exploits the use of endo-1,4-β-xylanases to improve lignin extraction (47). These xylanases break down the internal glycosidic bonds of xylan, thus reducing the degree of polymerization. Lower degrees of polymerization allow for the access of bleaching reagents into the cellulose fibers, hence, decreasing the need for stronger chlorine-based reagents (22). For pulping and biobleaching, processes at high temperatures are preferred. XynA from *T. maritima,* is a promising bio-bleaching enzyme, shown to be effective in releasing lignin and sugars over a pH range of 5.0-10.0 and optimally activity at pH 10.0 and 90^{o}C (47). A brightness of 90.5 % ISO was obtained for the XynA-treated pulp, compared to a brightness of 86.7% ISO for the untreated pulp. Similarly, the *T. maritima* XynB also has promise for bio-

bleaching, since it is active and stable at 90^{o}C and functions at neutral to alkaline pH values (47).

Enzymes for use in molecular biology: Thermostable enzymes, DNA polymerases in particular, are critical in catalyzing the polymerase chain reaction (PCR), a centrally important research tool in molecular biology (26). The first extremely thermophilic DNA polymerase to be used commercially, Taq polymerase, exhibiting the most activity at 80^{o}C, was isolated from *Thermus aquaticus*, a bacterium initially isolated from a hot spring in Yellowstone National Park (18). Before this discovery, *Escherichia coli* DNA polymerases were used for PCR, making it necessary to add more heat labile DNA polymerase during each cycle of PCR, following denaturation and primer annealing steps. Although *Taq* polymerase has a 5'-3' exonuclease activity, it does not have a 3'-5' exonuclease activity, or, in other words, proof-reading activity (66). Hence, extremely thermophilic DNA polymerases with high fidelity were sought and identified from archaeal hyperthermophiles in the genera *Pyrococcus* and *Thermococcus*; these enzymes have proof-reading activities and are now widely used for PCR (26).

Proteases: Proteolytic enzymes that can work at high temperatures and under alkaline conditions and are also resistant to denaturing agents play a major role in the detergent, food, pharmaceutical and textile industries (26, 42). Extremely thermophilic proteases thus far characterized are mostly produced by archaea in the genera *Pyrococcus*, *Thermoccus*, *Staphylothermus* (42). For instance a 44 kDa

extracellular protease from *Pyrococcus* sp. KOD1 (*Thermococcus kodakarensis* KOD1) was shown to be stable at 90^{o}C and to have a eukaryotic catalytic mechanism similar to thiol proteases (33). Other examples of extremely thermophilic proteases are from *Thermococcus aggreganes*, *Thermococcus celer*, *Thermococcus litoralis*, *Desulfurococcus* sp., and *T. maritima*, all having optimum temperatures of 85^{o}C (42). Proteasomes, intracellular protease complexes with Mr approaching 1 MDa, are common to the archaea and typically exhibit several different proteolytic specificities (61). Their cellular function relates to the need for protein turnover, particularly under stress conditions. Although proteasomes are found in all eukaryotes, all archaea and some bacteria, their composition can vary. For instance, the hyperthermophile *Pyrococcus furiosus* has a 20S proteasome formed by one α-protein and two β-proteins; this proteasome has different amounts of each β-protein depending on assembly temperature (61).

Biomedical uses for high temperature enzymes: Simple biosensing systems are sought in various fields of medical testing. For example, a thermostable glucokinase from *Bacillus stearothermophilus* has been used for the construction of an optical nanosensor for monitoring the levels of blood glucose (23). This glucokinase was labeled with the sulphhydryl-reactive fluorophore 2-(4-(iodoacetamido)aniline)naphthalene-6-sulphonic acid, such that binding of the labeled glucokinase to substrate generated molecular motions that could be quantified by fluorescence spectroscopy (23).One interesting feature of this device

is that no ATP is available so that glucokinase can bind to the substrate but not convert it to product. Thus, this is a non-consuming glucose sensor.

Food processing applications of extremely thermophilic enzymes: Thermostable enzymes are being considered for use as Time-Temperature Indicators (TTI) for evaluation and validation of in-pack sterilization processes. TTIs are heat-responsive sensors that provide information on the thermal history of a food material undergoing the thermal processing, in sterilization for example (73). TTI components are expected to be inexpensive, easily prepared, and not disturb the heat transfer environment in the food; most importantly, the temperature sensitivity of TTI reaction rate and target attribute (e.g., microbial load (contamination), or quality (texture)) must be the same (41). Extremely thermophilic enzymes exhibit a similar elliptical pressure-temperature diagram as for microbial inactivation, thus this property makes these enzymes promising candidates as TTIs (73). The change in activity or residual enthalpy of denaturation of enzymes can be monitored as the TTI responsive property. This approach was demonstrated with an α-amylase from *Bacillus licheniformis* as a TTI sensor. The residual enthalpy of denaturation of these enzymes was monitored and shown to be an applicable TTI for thermal impact evaluation in the sterilization range of 100- 130^{o}C (41).

2.2 Enzymes from Psychrophiles

2.2.1. Psychrophiles

Psychrophiles are microorganisms that thrive at temperatures ranging from -35 to 15^{o}C. They are usually isolated from the deep sea, Antarctic marine environments, marine and glacial ices, polar soils, mountain environments (19), cold deserts, permafrost soils, and glaciers (64). Psychrophiles are mostly bacteria and archaea, however there are also psychrophilic yeast, fungi and algae (64).

2.2.2. Adaptations to cold temperatures

There are many adaptations that enable microorganisms to thrive at very low temperatures. The genomes of ten psychrophiles sequenced to date have revealed that their proteins contain large numbers of non-charged polar amino acids (e.g., glutamine and threonine) (4). Small and/or neutral side chains in the coil regions of secondary structures of psychrophilic proteins provide higher confromational flexibility (64). Cell membranes from psychrophiles have a higher content of unsaturated fatty acids that presumably contribute to membrane fluidity and, hence, allows for the transport of substrates at very cold temperatures (43).

The function of psychrophilic enzymes at low temperatures has been studied to provide a better understanding about biocatalysis. These enzymes have up to ten-fold higher activity than mesophilic enzymes at biologically low temperatures, presumably to compensate for the reduction of chemical reaction rates. Nevertheless, this activity is still lower than that of mesophilic enzymes at 37^{o}C.

Psychrophiles maintain catalytic efficiency, k_{cat}/K_m, necessary for cellular metabolism, by increased turnover rates, k_{cat}, and increased substrate affinity (lower K_m). For instance, a chitobiase from an Antarctic bacterium has a k_{cat} of 8 times higher than that of a related mesophilic chitobiase at 5^{o}C. Similary, K_m for the substrate is 25 times lower at this temperature, such that the k_{cat}/K_m is 200 times that of the mesophilic chitobiase at low temperatures (28). Active site structure also contributes to the function of psychrophilic enzymes at low temperatures, which stems from a larger catalytic cavity and high accessibility to ligands compared to mesophilic enzymes. Better access to the catalytic cavity allows substrates and products to enter and exit faster leading to higher reaction rates (28).

Cold denaturation is an interesting phenomenon that results from hydration of polar and non-polar proteins and weakening of hydrophobic forces, all of which have a significant effect on protein folding and stability. Intracellular enzymes from psychrophiles are protected from cold denaturation by compatible solutes, such as potassium glutamate and trehalose (37). Psychrophiles also have intracellular cold shock proteins, that act as chaperones, cryoprotectors and antifreeze molecules. There is no explanation yet for protection from cold denaturation for extracellular psychrophilic enzymes, but exopolymeric substances (EPS) may be involved (37).

2.2.3. Industrial applications of psychrophilic enzymes

There is an increasing demand for enzymes that can work at lower temperatures to decrease energy consumption in laundry applications, food processing applications, and bioremediation (19).

Lipases: Lipases that function at low temperatures have found use in the detergent industry (45). Cold-adapted lipases have been isolated from many psychrophiles, including *Microbacterium phyllosphaerae* MTCC7530 from Naukuchiatal lake (52), *Pseudoalteromonas* sp. from Antarctic marine waters (59), and *Bacillus sphaeriucus* MTCC7526 from Gangothri Glacier (Western Himalaya) (51). In addition to cold-active lipases from psychrophiles, there are lipases isolated from psychrotrophic species; these are actually mesophiles that can grow at unusually low temperatures (81). For instance, *Acinetobacter* sp. Strain No. 6 is a psychrotrophic bacterium isolated from Alaskan and Sibernian tundra soil samples, has an optimum growth temperature of 28^{o}C, but can grow as low as 4^{o}C. This bacterium produces an extracellular lipase that can degrade triglycerides efficiently at very low temperatures (81).

Feruloyl esterases: Psychrophilic feruloyl esterases are employed in industry for biomass degradation, biotransformation and isolation of ferulic acid derivatives useful as antioxidant, antimicrobial and photoprotectant properties. For instance, one studied feruloyl esterase from the psychrophilic bacterium, *P. haloplanktis* TAC125 (4), is a family 1 carbohydrate esterase and displays significant activity towards

pNP-acetate, α- and β-naphthyl acetate, and 4-methylumbelliferyl p-trimethylammonio cinnamate chloride (a model substrate for determining feruloyl esterase activity) (4).

2.3. Enzymes from Alkaliphiles

2.3.1. Alkaliphiles

Alkaliphiles are microorganisms with the ability to grow at high pH (pH ≥ 10). They are generally found in soda lakes, underground alkaline water, alkaline wastes generated as by-products of food-processing industries, and intestines of insects (92). At one time, alkaliphiles were used for indigo fermentation, in which indigo from crushed fresh plant material (*Polygonum tinctorium* Lour) was reduced by certain bacteria in warm water, through the addition of alkaline ash (7). Through this oxidation reaction insoluble indigo is converted into the soluble form (93). Recently, a novel alkaliphile (T_{opt} = 20-30^{o}C, pH_{opt}= 9-12.3) was isolated from indigo fermentation liquor samples in Shikoku, Japan (93); *Alkalibacterium indicireducens* sp. nov. A11^{T} plays a major role in indigo fermentation under alkaline conditions (93).

Alkaliphiles differ in salt, O_2, and temperature requirements as well as in metabolic characteristics. Haloalkaliphiles require high salinity (up to 33% [wt/vol] NaCl), in addition to alkaline conditions (43). Haloalkaliphiles mostly inhabit the soda deserts and soda lakes, such as the Rift Valley lakes of East Africa and the western soda lakes of the United States. Thermoalkaliphiles grow at high temperatures and

at high pH. For instance, *Thermococcus alcaliphilus* sp. nov. is an archaeon isolated from a shallow marine hydrothermal vent at Vulcano Island, Italy that grows on polysulfide at a pH range of 6.5 to 10.5 and at a temperature range of 56 to 90^{o}C (53). Some alkaliphiles are anaerobic; e.g., *Anaerobranca horikoshi* is an anaerobic, alkalitolerant, thermophilic bacterium isolated from Yellowstone National Park (27). Alkaliphiles also have methanogens, cyanobacteria, and sulfur-oxidizing bacteria among their members (43).

Most intracellular enzymes from alkaliphiles exhibit optimal activity at neutral pH, consistent with the fact that the internal pH of some alkaliphiles is approximately 7 or a little higher (92). For instance, an α-galactosidase from the alkaliphilic bacterium, *Micrococcus* sp. strain 31-2 is active at pH 7.5. Another example is *B. halodurans* C-125, which grows at pH of 10.0, while having an internal pH of 8.2 (43). These enzymes from alkaliphiles are not considered alkaliphilic. There are certain adaptations that allow alkaliphilic microorganisms to grow in alkaline environments, while maintaining their cytoplasm near neutral pH.

2.3.2. Adaptations to high pH

Cell wall components play a major role in protecting the cytoplasm in alkaliphiles from high pH; these include acidic polymers, such as galacturonic acid, gluconic acid, glutamic acid, aspartic acid, and phosphoric acid (56). The negative charges on these polymers can adsorb sodium and hydronium ions and repulse hydroxide ions. There are also many other adaptations to maintain pH homeostasis,

such as having multiple Na^+/H^+ anti-porters in the cell walls compared to that of neutrophilic species (56). Moreover, there are also adaptations in the peptidoglycan layer noted in alkaliphilic bacteria. Comparison of the peptidoglycan layer of a alkaliphilic *Bacillus* spp. with that of neutrophilic *Bacillus subtilis* showed that hexosamine was in excess in alkaliphilic species (43).

2.3.3. Biotechnological applications of alkaliphilic enzymes

Most alkaliphilic enzymes of current commercial importance are isolated from alkaliphilic microorganisms with an alkaline internal pH. For instance, all alkaline amylases are purified from alkaliphiles. No alkaline amylase has been isolated from a neutrophilic microorganism to date. Alkaline proteases have been isolated from alkaliphilic *Bacilllus* strains (43); for example, *Bacillus* sp. strain 221 isolated from soil, produces an extracellular alkaline serine protease that is active at pH 11.5 (43). Enzymes, mostly proteases and amylases, used in alkaline laundry detergents have to be active at high pH (38). Alkaline proteases are also used for de-hairing animal hides and for decomposition of the gelatinous coating of X-ray films (43). Other applications are discussed below.

α-amylases: As mentioned previously, amylolytic enzymes are widely used in the food, detergent and textile industry. There are many examples of such enzymes from alkaliphiles. A saccharifying type α-amylase produced by *Bacillus* sp. strain A-40-2, active at pH 10 to 10.5 (43), hydrolyzed starch to glucose, maltose, and maltotriose with a 70% yield. *Streptomyces gulbargensis* sp. DAS produces an

extracellular α-amylase with the optimum activity at 45^{o}C and optimum pH of 9.0 (83), but retains 70% of its activity at pH 11.0. Due to having high stability and activity at moderately high temperatures and alkaline conditions, this amylase is a promising enzyme for the detergent industry.

Cyclomaltodextrin glucanotransferases (CGTase): Alkaline CGTases have been identified that could be used for the production of crystalline α-, β-, and γ-cyclodextrins. The yield of cyclodextrin from starch was improved when an alkaline CGTase from *Bacillus* sp. strain 38-2 was used, compared to the conversion by a neutral CGTase from *B. macerans*. Due to the low yield with the neutral CGTase, toxic organic solvents (e.g. trichloroethylene, bromobenzene, and toluene) were required to precipitate CD. However, an alkaline CGTase achieved higher yields (85 to 90% from amylose and 70 to 80% from potato starch), without the need of organic solvents. The replacement of a neutral CGTase by an alkaline CGTase decreased the cost of β-CD production from $1,000/ kg to $5/kg (43).

Cellulases: Alkaline enzymes are also important in the textile industry. For bio-polishing denims, alkaline cellulases are preferred since they control back-staining and yield good denim finishing properties. When acidic and neutral cellulases (from *H. insolens* and *T. reesei*, respectively) were used for the enzymatic treatment, back-staining of the white yarn of denim fabric was observed (2).

Among possible sources of fermentable sugar for biofuels production, cotton-based waste textiles have been considered (46). Although this form of cotton does

not contain lignin or hemicelluloses, it has a high crystallinity index (between 0.81-0.95), which needs to be reduced through pretreatment (94). In one study, the addition of 12% NaOH at 0^{o}C for 3 hours was used to disrupt H-bonds between the glucan chains in crystalline cellulose, hence decreasing the crystallinity. This presents an opportunity for alkaline cellulases in the hydrolysis step, so that the washing step to bring the pH to neutral for fermentation can be eliminated, thus improving ethanol yields (46).

2.4. Enzymes from Halophiles

2.4.1. Halophiles

In high salt environments both halophiles (“salt-loving”) and halotolerant species can be found. In general, a halophile grows optimally at 50 g/l salt concentration (equivalent to 0.85 M NaCl) or higher and tolerates 100 g/l salt (equivalent to 1.7 M NaCl) (71). Extreme halophiles grow best in environments containing 2.5-5.2 M salt. A halotolerant microorganism, on the other hand, does not require high salt concentrations for growth, but can tolerate salt concentrations of up to 2.5 M. Halophiles are found in hypersaline niches, including marine environments, alkaline saline waters (low in magnesium and calcium), and magnesium-rich saline bodies of water, as well as in saline soils, salted food products and rock salt deposits (35, 69). The Dead Sea (salt concentration of 340 g/l) and the Great Salt Lake in Utah contain a great diversity in halophilic archaea, bacteria, and eukaryotic unicellular algae (69).

Salt-loving organisms can be found in all three kingdoms of life: Archaea, Bacteria, and Eukarya (71). The order *Halobacteriales* from Archaea has the most halophilic microorganisms. The phyla *Cyanobacteria*, *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Spirochaetes* and *Bacteroidetes* all contain halophilic bacteria. Halophilic eukaryotes can be found in the genus *Dunaliella*, which contains unicellular green algae and brine shrimp (*Artemia salina*, *Artemia franciscana*) (71).

2.4.2. Adaptations to high salt concentration

To cope with high concentrations of salt, halophiles use different strategies. These organisms have to keep their cytoplasm isosmotic with the extracellular environment to avoid losing water by diffusion through their water permeable biological membranes (70). This osmotic adaptation is accomplished by two strategies: “salt-in” and compatible-solute. “Salt-in” refers to the strategy of keeping the intracellular environment at high salt concentrations in addition to accumulating KCl to balance osmolarity. In this case, all intracellular components need to be adapted to high salt concentrations. In the compatible-solute strategy cells maintain low salt concentration intracellularly, but produce or take up small organic molecules, called compatible solutes, from the medium (70). Polyols, such as glycerol and arabitol, sugar and sugar derivatives, amino acids and derivatives, and quaternary amines are examples of compatible solutes. The common features of compatible solutes are low-molecular-weight, high solubility in water, and being uncharged or zwitterionic at the physiological pH (70). Halophilic proteins also have

a greater number of acidic amino acid residues than nonhalophilic counterparts (62). Protein solubility at high salt concentrations is thus accomplished by high surface charge which provides a hydrated ion shell at the protein surface (9).

2.4.3. Biotechnological applications of halophilic enzymes

Halophilic enzymes are used commercially to produce compatible solutes and valuable extracellular enzymes (87). Halophilic enzymes are also employed in oil recovery and the degradation of industrial pollutants in high salt concentration environments.

Proteases: *Halogeometricum borinquense* strain TSS-101 is an extremely halophilic archaeon isolated from solar salterns of Tuticorin, Tamilnadu, India (87). An extracellular haloalkaliphilic, thermostable serine protease was purified from this archaeon. This protease showed no activity in the absence of NaCl and was optimally active at 60-70^{o}C and pH 10.0, respectively (87). Moreover, in order to maintain activity, 20% NaCl was required. Azocoll and azocasein were the two substrates against which the protease exhibited the most activity. With its high temperature, high salt, and high pH stability, this protease could find use in the formulation of detergents and in the fermentation of seafood (87).

Lipases and esterases: In one study fifty-four *Halobacteria* strains from waters of the El Golea Sebkha (containing salt concentrations up to 300 g/l) located in Algerian Sahara were screened for halophilic extracellular lipolytic activity (9). Thirty-five of the strains growing optimally at 3-5 M salt concentration, exhibited esterase

and lipase acitivity by hydrolyzing both short and long acyl-chain *p*-nitrophenyl esters (9). In another study 231 moderately halophilic bacteria and 49 extremely halophilic bacteria were isolated from Howz Soltan playa, a hypersaline lake in the central desert region of Iran (75). These strains were identified as members of the genera: *Salicola*, *Halovibrio*, *Halomonas*, *Oceanbacillus*, *Thalassobacillus*, *Halobacillus*, *Virgibacillus*, *Gracilibacillus*, *Salinicoccus*, and *Piscibacillus*. Extracellular hydrolytic activities of these microorganisms were screened. It was shown that gram-negative rods had mainly lipolytic activity; 131 out of 172 gram-positive rods, 45 out of 56 gram-negative rods, and 24 out of 52 gram-positive cocci produced lipases, active at high salinity (75).

2.5. Enzymes from Acidophiles

2.5.1. Acidophiles

Acidophiles grow at low pHs. While moderate acidophiles grow optimally at pH 3 to 5, extreme acidophiles thrive in environments where the pH is lower than 3 (49). Acidic environments are characterized by widely varying physicochemical parameters (49). For instance, extreme acidity in volcanic areas may arise from oxidation of elemental sulfur that is deposited by the condensation of hydrogen sulfide and sulfur dioxide in volcanic gases. Elemental sulfur oxidation by sulfur-oxidizing archaea and bacteria often at high temperature results in production of sulfuric acid, which consequently lowers the pH (12). Most global acidic environments arise by human activity, through the mining of base and precious

metals and coal, in particular. If mining is done without the proper environment controls, extremely acidic and toxic fluids, called acid mine drainage (AMD), are generated, leading to one of the severe forms of environmental pollution worldwide. Sulfide minerals, such as pyrite (FeS_2) present in ore deposits, are converted to sulfuric acid and dissolved iron when they are exposed to water and air, as in the case with abandoned coal mines or active mining areas. Thus, pyrite dissolution results in warm (35 to 57^{o}C) and extremely acidic (pH 0.5-0.9) fluid, acid mine drainage (AMD) which is heavy metal-rich owing to low solubility of metal oxyanions in acidic solutions (12). One interesting example of this situation is the Richmond Mine at Iron Mountain in California, which has a great biodiversity of eukaryotic and prokaryotic microbial communities thriving in AMD (6).

Two examples of eukaryotic extreme acidophiles, isolated from acid mine drainage (AMD), *Acontium velatum* and *Scytalidium acidophilum* are metal-tolerant mitosporic fungi (49, 79). *A. velatum* tolerates copper concentrations up to 14 mM and grows between pH 0.2 and 0.7, while *S. acidophilum* can grow in the presence of 140 mM copper and pH 0 (79). Copper is one of the essential nutritional requirements of filamentous fungi, higher plants, and some animals. However, this requirement is on the order of 1-10 micrograms copper per 100 ml; levels higher than this can be severely toxic (79).

Acidophiles are important microorganisms in biomining. Iron-oxidizing acidophilic bacteria, such as *Leptospirillum ferriphilum* and *Acidithiobacillus*

ferrooxidans are used commercially to accelarate the oxidative dissolution of sulfide minerals (50). Other bacteria have also been considered for biomining. For example, *Sulfobacillus benefaciens* sp. nov, recently isolated from pH- and temperature-controlled stirred tanks where sulfide mineral concentrates were being oxidized to solubilise base metals, is a facultative autotroph capable of growing on S^o, tetrathionate, ferrous iron, and sulfide minerals (50).

Acidophiles are also employed in bioremediation, such as for removing heavy metals from dredged sediments. Conventional approaches for this treatment are chemical extraction and thermo-stabilization (involves extreme temperatures and toxic chemicals), however these methods are not environmentally friendly (8). In bioremediation processes chemo-litho-autotrophic bacteria, such as *Acidithiobacillus* spp., which oxidize sulfur and Fe^{+2} under acidic conditions, can be employed. Oxidation of sulfides located within heavy metal fractions of the sediment results in solubilization and mobilization of heavy metals (17).

2.5.2. Adaptations of acidophiles

The membranes of acidophilic archaea have been investigated in some detail because of their unique biophysical characteristics, because these organisms do not have cell walls, membranes are the only barriers to maintain the osmotic balance and to protect the cell from acidification (60). The cellular membrane of *Ferroplasma acidarmanus* (pH_{opt}=0.0-1.5), an extremely acidophilic archaeon, was characterized by thin layer chromatography (TLC) and matrix-assisted laser desorption/ionization

time of flight mass spectrometry (MALDI-TOF-MS) analysis, to reveal that this microorganism's cytoplasmic membrane has tetraethers and a low mobility core lipid (60). Since ether linkages are less susceptible to disruption by acidic environments than ester linkage found in bacteria (31), having an ether-linked membrane is an advantage for acidophiles. In fact, all acidophilic archaea studied to date have extremely proton-impermeable tetraether-linked monolayer membrane lipids with an isoprenoid core that makes them less permeable to protons (31).

2.5.3. Enzymes from acidophiles

Although most acidophiles grow best in environments where the pH is lower than 4, their intracellular pH is maintained closer to neutral values (60). Therefore, intracellular enzymes function best at neutral pH. For example, the cytoplasmic pH of *F. acidarmanus* (pH_{opt}=0.0-1.5), measured by NMR, showed an intracellular pH of 4.9±0.48. A DNA ligase from *Ferroplasma acidiphilum***,** an extremely acidophilic archaeon from the order *Thermoplasmatales* and phylum *Euryarchaeota* (30) contains two ferric ions that take part in three dimensional structure of the enzyme and are crucial for substrate binding. Unlike DNA ligases from other extremely acidophilic archaea, the *F. acidiphilum* (LigFa) has an *in vitro* pH optimum of 1.5-3.0, although most of the archaea from the same order have an intracellular pH of 4.6-5.6 (30). LigFa was most active on the two short (25- and 35-mer) contiguous synthetic oligonucleotides at 40^{o}C and pH 2.5-3.0, and maintained ~80% of the activity at pH 1.5-2.0, but no activity at pH 5.0 and above. The acidophily of this

enzyme was attributed to the two ferric ions. It was argued that this remarkable LigFa, active at low pH, may have unstable nucleic acids, because of mutagenic activity and depurination of nucleic acids at acidic environments. In addition, the presence of Fe which plays a role in some redox reactions that may yield mutagenic oxygen radicals can cause the instability (30).

2.6. Other extremophiles

There are many types of extremophiles adapted to grow under different extreme conditions (see Table 1.1). The biocatalysts from the ones explained so far are studied extensively, however, the enzymes from the other extremophiles have not been studied to any great extent. Thus, in this section only the microorganisms will be discussed. Barophiles are microorganisms requiring high hydrostatic pressures for their growth and are usually found in the deep ocean or deep-sea hydrothermal vents (below 2000 m), whereas, barotolerant species are found above the 2000 m threshold (35). The barophiles growing in the deep ocean are psychrophilic or psychrotolerant, and the ones growing in the deep-sea hydrothermal vents are thermophilic or thermotolerant. The first barophile, a *Sprillum* sp. (strain CNPT3) was isolated from a decomposing amphipods captured from a depth of 5600 m (91). This strain grew optimally at 2-4^{o}C and at 570 atm. Another example of a barophilic microorganism is strain MT41 isolated from again a dead and decomposing amphipod, but this time it was captured from a depth of 10,476 m which is little less than the depth of Marianas Trench in the Pacific Ocean, having

the maximum depth (10,790 m), corresponding to approximately 110 MPa, on the planet (91). Strain MT41 has an optimum growth temperature and pressure of 2^{o}C and 690 atm, respectively. It did not form colonies after being exposed for several hours at atmospheric pressure (91). An example of eukaryotic barophiles is *Aspergillus sydowii* which is a filamentous fungi, isolated from a depth of 5000 m in the Central Indian Basin. The germination of *A. sydowii* occurred at 490 atm (74). *Methanococcus jannaschii*, a barophilic and hyperthermophilic archaeon is worth mentioning. It was isolated from a deep-sea hot vent (21^{o}N East Pacific Rise at a depth of 2,610 m) and it grows optimally at 3 atm and 85^{o}C (65). According to a study undertaken, the specific growth rate based on methane production of *M. jannaschii* increased from 0.5±0.1 to 2.36±0.01 h^{-1} when the pressure was increased from 7.8 to 750 atm (65).

Adaptations to life at high pressure are seen in the cell membrane. In one of the studies undertaken in this research area it was shown that the total ratio of unsaturated fatty acids to total fatty acids in the cell membrane of barophiles increased from 1.9 to 3 when the incubation pressure was increased from 1 to 680 atm. Another similar study was about the influence of pressure on unsaturated fatty acid composition; a barotolerant strain of *Alteromonas* isolated from 4033 m in the Izu-Ogasawara Trench, Japan, increased the synthesis of phospholipids containing unsaturated fatty acids with increased pressure during its growth. These results

showed that under high pressure conditions microorganisms try to maintain the fluidity of the cell membrane for transportation of solutes.

In order to investigate the stability of barophilic proteins, two lactate dehydrogenases (LDH) from the abyssal grenadier *Coryphaenoides armatus* (grows below 2000 m and at 1-4^{o}C) and the Antlantic cod *Gadus morhua* (grows at 50-200 m and at 0-20^{o}C) were studied. The primary structures of the two enzymes were very similar; only 21 amino acids were different. For instance, *C. armatus* enzyme had lysine and methionine substitutions in the pressure stabilizing N-terminal. Owing to these amino acid substitutions, stability of the enzymes from *C. armatus* and *G. morhua* was significantly different at high pressures. It was concluded that amino acid substitutions away from the active site of the enzyme affect the protein stability and enzyme activity (13).

Another group of extremophiles, xerophiles are microorganisms capable of living under very dry conditions. *Aspergillus penicillioides*, an example of a xerophile, is found in extremely dry environments, such as dried foods, binocular lenses, human skin, and house dust (84). There are some extreme microorganisms which are able to withstand high levels of toxic compounds, organic solvents, electric currents, pure CO_2 gas, and redox potential. Some examples are *Electrophorus electricus*, the electric eel, which can produce and withstand strong electric currents and was isolated from Northern areas of South America (21), *Cyanidium caldarium* which grows in a media ventilated with pure CO_2 (78), *Ferroplasma acidarmanus*, an

arsenic-hypertolerant acidophilic archaeon isolated from the Iron Mountain mine, California and capable of tolerating 10 g arsenate per liter (5). Withstanding high levels of radiation is also an interesting ability, since radiation disrupts nucleic acids. *Deinococcus radiodurans* is an extremophile with the ability to thrive under ionizing radiation (up to 20 kGy of gamma radiation) and UV radiation (up to 1,000 J m^{-2}) (76).

3. Conclusion

Extremophiles have received a great deal of attention. Their commercial potential has been recognized and enzymes from extremophiles, “extremozymes”, made their way to multibillion-dollar industries, such as agriculture, chemical synthesis, laundry detergents, and pharmaceuticals (76). Extremozymes are intriguing because they present a model for stability to modify mesophile-derived enzymes and they can cope with harsh industrial process conditions. Potentially valuable enzymes can be explored by functional genomics approaches and bioinformatics analysis. The enzymes identified have to be produced in sufficient amounts for assessment of biocatalytic properties. These enzymes can be produced either by increasing biomass production of the extremophile or cloning the biocatalyst genes and expressing them in suitable hosts. However, either optimizing the biomass production or cloning those genes is trivial. Extremophiles usually have low biomass yields and very low cell densities to extract DNA for cloning (29). This is one of the bottle necks of studying extremophiles and extremozymes.

Research about extremophiles also provides better hypotheses regarding the origin of life. The extreme conditions were inevitable during the early stages of the Earth; the start of life might have occurred at a very hot environment, a very cold one, a very dry one or at an environment with high radiation, but presumably this was an extreme condition compared to what we call normal conditions now. Moreover, stability studies of extremophiles can improve adaptations and survival of terrestrial organisms in different extreme conditions, such as in space stations, space flights, and laboratory simulations of space conditions (45). In addition, the ability of extremophiles to thrive in harsh conditions shows that there could be life on planets beyond Earth, such as Mars or Europa. Hence, the study of extremophiles is essential to define the limits of life.

REFERENCES

1. Aislabie, J., D. J. Saul, and J. M. Foght. 2006. Bioremediation of hydrocarbon-contaminated polar soils. Extremophiles 10:171-9.
2. Anish, R., M. S. Rahman, and M. Rao. 2007. Application of cellulases from an alkalothermophilic *Thermomonospora* sp. in biopolishing of denims. Biotechnol Bioeng 96:48-56.
3. Asker, D., T. S. Awad, T. Beppu, and K. Ueda. 2009. *Deinococcus aquiradiocola* sp. nov., isolated from a radioactive site in Japan. Int J Syst Evol Microbiol 59:144-9.
4. Aurilia, V., A. Parracino, and S. D'Auria. 2008. Microbial carbohydrate esterases in cold adapted environments. Gene 410:234-40.
5. Baker-Austin, C., M. Dopson, M. Wexler, R. G. Sawers, A. Stemmler, B. P. Rosen, and P. L. Bond. 2007. Extreme arsenic resistance by the acidophilic archaeon '*Ferroplasma acidarmanus'* Fer1. Extremophiles 11:425-34.
6. Baker, B. J., M. A. Lutz, S. C. Dawson, P. L. Bond, and J. F. Banfield. 2004. Metabolically active eukaryotic communities in extremely acidic mine drainage. Appl Environ Microbiol 70:6264-71.
7. Balfour-Paul, J. *Indigo in the Arab world*. pp 84-86. Routledge, 1997.

8. Beolchini, F. et al., Auto- and heterotrophic acidophilic bacteria enhance the bioremediation efficiency of sediments contaminated by heavy metals. Chemosphere (2008), doi:10.1016/j.chemosphere.2008.11.057

9. Bhatnagar, T., S. Boutaiba, H. Hacene, J. L. Cayol, M. L. Fardeau, B. Ollivier, and J. C. Baratti. 2005. Lipolytic activity from Halobacteria: screening and hydrolase production. FEMS Microbiol Lett 248:133-40.

10. Blumer-Schuette, S. E., I. Kataeva, J. Westpheling, M. W. Adams, and R. M. Kelly. 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr Opin Biotechnol 19:210-7.

11. Borges, K. M., A. Bergerat, A. M. Bogert, J. DiRuggiero, P. Forterre, and F. T. Robb. 1997. Characterization of the reverse gyrase from the hyperthermophilic archaeon *Pyrococcus furiosus*. J Bacteriol 179:1721-6.

12. Bridge, T. A. M., and D. B. Johnson. 1998. Reduction of soluble iron and reductive dissolution of ferric iron-containing minerals by moderately thermophilic iron-oxidizing bacteria. Appl Environ Microbiol 64:2181-6.

13. Brindley, A. A., R. W. Pickersgill, J. C. Partridge, D. J. Dunstan, D. M. Hunt, and M. J. Warren. 2008. Enzyme sequence and its relationship to hyperbaric stability of artificial and natural fish lactate dehydrogenases. PLoS ONE 3:e2042.

14. Brock, T. D. 1967. Life at high temperatures. Evolutionary, ecological, and biochemical significance of organisms living in hot springs is discussed. Science 158:1012-9.

15. Burgess, EA, Wagner, ID, and Wiegel, J. Thermal environments and biodiversity. pp 13-23 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

16. Cavicchioli, R., K. S. Siddiqui, D. Andrews, and K. R. Sowers. 2002. Low-temperature extremophiles and their applications. Curr Opin Biotechnol 13:253-61.

17. Effect of substrate concentration on biobleaching of metal contaminated sediment. 2001. J Hazard Mater B (82): 77-89.

18. Chien, A., D. B. Edgar, and J. M. Trela. 1976. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. J Bacteriol 127:1550-7.

19. Cowan, DA, Casanueva, A, and Stafford, W. Ecology and biodiversity of cold-adapted microorganisms. pp 119-129 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

20. Dahiya, N., R. Tewari, and G. S. Hoondal. 2006. Biotechnological aspects of chitinolytic enzymes: a review. Appl Microbiol Biotechnol 71:773-82.

21. Daniel Pauly. *Darwin's Fishes: An encyclopedia of ichthyology, ecology, and evolution*. Cambridge University Press, 2004.

22. de Carvalho Peixoto-Nogueira, S., M. Michelin, J. H. Betini, J. A. Jorge, H. F. Terenzi, and M. de Lourdes Teixeira de Moraes Polizeli. 2009. Production of xylanase by *Aspergilli* using alternative carbon sources: application of the crude extract on cellulose pulp biobleaching. J Ind Microbiol Biotechnol 36:149-55.

23. de Champdore, M., M. Staiano, M. Rossi, and S. D'Auria. 2007. Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest. J R Soc Interface 4:183-91.

24. de Miguel Bouzas, T., J. Barros-Velazquez, and T. G. Villa. 2006. Industrial applications of hyperthermophilic enzymes: a review. Protein Pept Lett 13:645-51.

25. Dock, C., M. Hess, and G. Antranikian. 2008. A thermoactive glucoamylase with biotechnological relevance from the thermoacidophilic Euryarchaeon *Thermoplasma acidophilum*. Appl Microbiol Biotechnol 78:105-14.

26. Egorova, K., and G. Antranikian. 2005. Industrial relevance of thermophilic Archaea. Curr Opin Microbiol 8:649-55.

27. Engle, M., Y. Li, C. Woese, and J. Wiegel. 1995. Isolation and characterization of a novel alkalitolerant thermophile, *Anaerobranca horikoshii* gen. nov., sp. nov. Int J Syst Bacteriol 45:454-61.

28. Feller, G. 2003. Molecular adaptations to cold in psychrophilic enzymes. Cell Mol Life Sci 60:648-62.

29. Ferrer, M., O. Golyshina, A. Beloqui, and P. N. Golyshin. 2007. Mining enzymes from extreme environments. Curr Opin Microbiol 10:207-14.

30. Ferrer, M., O. V. Golyshina, A. Beloqui, L. H. Bottger, J. M. Andreu, J. Polaina, A. L. De Lacey, A. X. Trautwein, K. N. Timmis, and P. N. Golyshin. 2008. A purple acidophilic di-ferric DNA ligase from *Ferroplasma*. Proc Natl Acad Sci U S A 105:8878-83.

31. Franke, S and Rensing, C. Acidophiles: Mechanisms to tolerate metal and acid toxicity. pp 257- 269 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

32. Fry, N. K., J. K. Fredrickson, S. Fishbain, M. Wagner, and D. A. Stahl. 1997. Population structure of microbial communities associated with two deep, anaerobic, alkaline aquifers. Appl Environ Microbiol 63:1498-504.

33. Fujiwara, S., S. Okuyama, and T. Imanaka. 1996. The world of archaea: genome analysis, evolution and thermostable enzymes. Gene 179:165-70.

34. Gadd, GM. Molecular biology and biotechnology of microbial interactions with organic and inorganic heavy metal compounds. pp 225-233 in Herbert, RA and Sharp, RJ, eds. Molecular biology and biotechnology of extremophiles.1992. Chapman and Hall, New York.

35. Galinski, EA and Tindall, BJ. Biotechnological prospects for halophiles and halotolerant micro-organisms. pp 76-108 in Herbert, RA and Sharp, RJ, eds. Molecular biology and biotechnology of extremophiles.1992. Chapman and Hall, New York.

36. Gao, J., M. W. Bauer, K. R. Shockley, M. A. Pysz, and R. M. Kelly. 2003. Growth of hyperthermophilic archaeon *Pyrococcus furiosus* on chitin involves two family 18 chitinases. Appl Environ Microbiol 69:3119-28.

37. Georlette, D., V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J. C. Marx, G. Sonan, G. Feller, and C. Gerday. 2004. Some like it cold: biocatalysis at low temperatures. FEMS Microbiol Rev 28:25-42.

38. Grant, WD and Horikoshi, K. Alkaliphiles: ecology and biotechnological applications. pp 143-160 in Herbert, RA and Sharp, RJ, eds. Molecular biology and biotechnology of extremophiles.1992. Chapman and Hall, New York.

39. Grosjean, H and Oshima, T. How nucleic acids cope with high temperature. pp 39-56 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

40. Grunden, A.M., D.A. Comfort, E.L. Malotky, and R.M. Kelly. 2004. Expression of Extremophilic Proteins. Expression Technologies: Current Status and Future Trends, F. Baneyx, Ed., Horizon Scientific.

41. Guiavarc'h, Y., A. Van Loey, F. Zuber, and M. Hendrickx. 2004. Development characterization and use of a high-performance enzymatic time-temperature integrator for the control of sterilization process' impacts. Biotechnol Bioeng 88:15-25.

42. Haki, G. D., and S. K. Rakshit. 2003. Developments in industrially important thermostable enzymes: a review. Bioresour Technol 89:17-34.

43. Horikoshi, K. 1999. Alkaliphiles: some applications of their products for biotechnology. Microbiol Mol Biol Rev 63:735-50, table of contents.

44. Islas, S, Velasco, M, Becerra, A, Delaye, L, and Lazcano, A. Extremophiles and the origin of life. pp: 3-10 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

45. Javaux, E. J. 2006. Extreme life on Earth--past, present and possibly beyond. Res Microbiol 157:37-48.

46. Jeihanipour, A., and M. J. Taherzadeh. 2009. Ethanol production from cotton-based waste textiles. Bioresour Technol 100:1007-10.

47. Jiang, Z. Q., X. T. Li, S. Q. Yang, L. T. Li, Y. Li, and W. Y. Feng. 2006. Biobleach boosting effect of recombinant xylanase B from the hyperthermophilic *Thermotoga maritima* on wheat straw pulp. Appl Microbiol Biotechnol 70:65-71.

48. Joan, M and Ernst, G. *Volcanoes and the environment*. pp 183-185. Cambridge University Press, 2005.

49. Johnson, DB. Physiology and ecology of acidophilic microorgansisms. pp 257-270 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

50. Johnson, D. B., C. Joulian, P. d'Hugues, and K. B. Hallberg. 2008. *Sulfobacillus benefaciens* sp. nov., an acidophilic facultative anaerobic *Firmicute* isolated from mineral bioleaching operations. Extremophiles 12:789-98.

51. Joseph, B., P. W. Ramteke, and G. Thomas. 2008. Cold active microbial lipases: some hot issues and recent developments. Biotechnol Adv 26:457-70.

52. Joshi, G. K., S. Kumar, B. N. Tripathi, and V. Sharma. 2006. Production of alkaline lipase by *Corynebacterium paurometabolum*, MTCC 6841 isolated from Lake Naukuchiatal, Uttaranchal State, India. Curr Microbiol 52:354-8.

53. Keller, M., F. J. Braun, R. Dirmeier, D. Hafenbradl, S. Burggraf, R. Rachel, and K. O. Stetter. 1995. *Thermococcus alcaliphilus* sp. nov., a new hyperthermophilic archaeum growing on polysulfide at alkaline pH. Arch Microbiol 164:390-5.

54. Kelly, R.M., and K.R. Shockley. Applications of Genomic Data: Enzyme Discovery and Microbial Genomics. 2004. In Microbial Genomics. C.M. Fraser, T. Read, and K.E. Nelson, Eds., Humana Press, Inc., Totowa, NJ.

55. Kelly, R.M., S.H. Brown, I.I. Blumentals, and M.W.W. Adams. 1992. Characterization of Enzymes from High Temperature Bacteria, in Biocatalysis at Extreme Temperatures, M.W.W. Adams and R.M. Kelly, eds., ACS Symposium Series No. 498, pp. 23-42.

56. Krulwich, TA, Hicks, DB, Swartz, T, and Ito, M. Bioenergetic adaptations that support alkaliphily. pp 311-329 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

57. Kumar, S, Arya, S, and Nussinov, R. Temperature-dependent molecular adaptation features in proteins. pp 75-85 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

58. Liebl, W., C. Winterhalter, W. Baumeister, M. Armbrecht, and M. Valdez. 2008. Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. J Bacteriol 190:1350-8.

59. Lo Giudice, A., L. Michaud, D. de Pascale, M. De Domenico, G. di Prisco, R. Fani, and V. Bruni. 2006. Lipolytic activity of Antarctic cold-adapted marine bacteria (Terra Nova Bay, Ross Sea). J Appl Microbiol 101:1039-48.

60. Macalady, J. L., M. M. Vestling, D. Baumler, N. Boekelheide, C. W. Kaspar, and J. F. Banfield. 2004. Tetraether-linked membrane monolayers in *Ferroplasma* spp: a key to survival in acid. Extremophiles 8:411-9.

61. Madding, L. S., J. K. Michel, K. R. Shockley, S. B. Conners, K. L. Epting, M. R. Johnson, and R. M. Kelly. 2007. Role of the beta1 subunit in the function and stability of the 20S proteasome in the hyperthermophilic archaeon *Pyrococcus furiosus*. J Bacteriol 189:583-90.

62. Madern, D., C. Ebel, and G. Zaccai. 2000. Halophilic adaptation of enzymes. Extremophiles 4:91-8.

63. Mason, O. U., U. Stingl, L. J. Wilhelm, M. M. Moeseneder, C. A. Di Meo-Savoie, M. R. Fisk, and S. J. Giovannoni. 2007. The phylogeny of endolithic microbes associated with marine basalts. Environ Microbiol 9:2539-50.

64. Metpally, R. P., and B. V. Reddy. 2009. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. BMC Genomics 10:11.

65. Miller, J. F., N. N. Shah, C. M. Nelson, J. M. Ludlow, and D. S. Clark. 1988. Pressure and Temperature Effects on Growth and Methane Production of the Extreme Thermophile *Methanococcus jannaschii*. Appl Environ Microbiol 54:3039-3042.

66. Niehaus, F., C. Bertoldo, M. Kahler, and G. Antranikian. 1999. Extremophiles as a source of novel enzymes for industrial application. Appl Microbiol Biotechnol 51:711-29.

67. Niehaus, F., A. Peters, T. Groudieva, and G. Antranikian. 2000. Cloning, expression and biochemical characterisation of a unique thermostable pullulan-hydrolysing enzyme from the hyperthermophilic archaeon *Thermococcus aggregans*. FEMS Microbiol Lett 190:223-9.

68. Oku, T., and K. Ishikawa. 2006. Analysis of the hyperthermophilic chitinase from *Pyrococcus furiosus*: activity toward crystalline chitin. Biosci Biotechnol Biochem 70:1696-701.

69. Oren, A. Biodiversity in highly saline environments. pp 224-231 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

70. Oren, A. 1999. Bioenergetic aspects of halophilism. Microbiol Mol Biol Rev 63:334-48.

71. Oren, A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Systems 4:2.

72. Petsko, G. A. 2001. Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". Methods Enzymol 334:469-78.

73. der Plancken, VI, Grauwet, T, Oey, I, Loey AV, and Hendrickx, M. Impact evaluation of high pressure treatment on foods: considerations on the development of pressure-temperature-time integrators (pTTIs). Trends in Food Science & Technology. 2008; (19): 337-348.

74. Raghukumar, C, Raghukumar, Sheelu, G, Sheelu, G, Gupta, SM, Nath, BN, and Rao, BR. Buried in time: culturable fungi in a deep-sea sediment core from the Chagos Trench, Indian Ocean. Deep-Sea Research. 2004. (51): 1759-1768.

75. Rohban, R., M. A. Amoozegar, and A. Ventosa. 2009. Screening and isolation of halophilic bacteria producing extracellular hydrolyses from Howz Soltan Lake, Iran. J Ind Microbiol Biotechnol 36(3):333-340.

76. Rothschild, L. J., and R. L. Mancinelli. 2001. Life in extreme environments. Nature 409:1092-101.

77. Schwartzman, D. W., and C. H. Lineweaver. 2004. The hyperthermophilic origin of life revisited. Biochem Soc Trans 32:168-71.

78. Seckbach, J., F. A. Baker, and P. M. Shugarman. 1970. Algae thrive under pure CO2. Nature 227:744-5.

79. Starkey, R. L., and S. A. Waksman. 1943. Fungi Tolerant to Extreme Acidity and High Concentrations of Copper Sulfate. J Bacteriol 45:509-19.

80. Sunna, A., M. Moracci, M. Rossi, and G. Antranikian. 1997. Glycosyl hydrolases from hyperthermophiles. Extremophiles 1:2-13.

81. Suzuki, T., T. Nakayama, T. Kurihara, T. Nishino, and N. Esaki. 2001. Cold-active lipolytic activity of psychrotrophic *Acinetobacter* sp. strain no. 6. J Biosci Bioeng 92:144-8.

82. Svetlichnyi, V., Svetlichnaya, TP, Chernykh, NA, Zavarzin, GA. 1990. *Anaerocellum thermophilum* gen. nov. sep. nov.: an extremely thermophilic cellulolytic eubacterium isolated from hot springs in the valley of geysers. Microbiology 59:598-603.

83. Syed, D. G., D. Agasar, and A. Pandey. 2008. Production and partial purification of alpha-amylase from a novel isolate *Streptomyces gulbargensis*. J Ind Microbiol Biotechnol 36(2):189-194.

84. Tamura, M., H. Kawasaki, and J. Sugiyama. 1999. Identity of the xerophilic species *Aspergillus penicillioides*: Integrated analysis of the genotypic and phenotypic characters. J Gen Appl Microbiol 45:29-37.

85. Tanaka, T., S. Fujiwara, S. Nishikori, T. Fukui, M. Takagi, and T. Imanaka. 1999. A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. Appl Environ Microbiol 65:5338-44.

86. Verhagen, M. F., A. L. Menon, G. J. Schut, and M. W. Adams. 2001. *Pyrococcus furiosus*: large-scale cultivation and enzyme purification. Methods Enzymol 330:25-30.

87. Vidyasagar, M., S. Prakash, C. Litchfield, and K. Sreeramulu. 2006. Purification and characterization of a thermostable, haloalkaliphilic extracellular serine protease from the extreme halophilic archaeon *Halogeometricum borinquense* strain TSS101. Archaea 2:51-7.

88. Warren-Rhodes, K. A., K. L. Rhodes, S. B. Pointing, S. A. Ewing, D. C. Lacap, B. Gomez-Silva, R. Amundson, E. I. Friedmann, and C. P. McKay. 2006. Hypolithic cyanobacteria, dry limit of photosynthesis, and microbial ecology in the hyperarid Atacama Desert. Microb Ecol 52:389-98.

89. Wiegel, J. 1998. Anaerobic alkalithermophiles, a novel group of extremophiles. Extremophiles 2:257-67.

90. Yang, S. J., H. S. Lee, C. S. Park, Y. R. Kim, T. W. Moon, and K. H. Park. 2004. Enzymatic analysis of an amylolytic enzyme from the hyperthermophilic archaeon *Pyrococcus furiosus* reveals its novel catalytic properties as both an alpha-amylase and a cyclodextrin-hydrolyzing enzyme. Appl Environ Microbiol 70:5988-95.

91. Yayanos, A. A., A. S. Dietz, and R. Van Boxtel. 1981. Obligately barophilic bacterium from the Mariana trench. Proc Natl Acad Sci U S A 78:5212-5.

92. Yumoto, I. Environmental and taxonomic biodiversity of gram-positive alkaliphiles. pp 295-309 in Gerday, C and Glansdorff, N, eds. Physiology and biochemistry of extremophiles. 2007. ASM Press, Washington, D.C.

93. Yumoto, I., K. Hirota, Y. Nodasaka, Y. Tokiwa, and K. Nakajima. 2008. *Alkalibacterium indicireducens* sp. nov., an obligate alkaliphile that reduces indigo dye. Int J Syst Evol Microbiol 58:901-5.

94. Zhang, Y. H., and L. R. Lynd. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng 88:797-824.

95. Zverlov, V. V., G. A. Velikodvorskaya, and W. H. Schwarz. 2003. Two new cellulosome components encoded downstream of celI in the genome of *Clostridium thermocellum*: the non-processive endoglucanase CelN and the possibly structural protein CseP. Microbiology 149:515-24.

TABLES

Table 1.1 Types of extremophiles		
Extremophile	**Extreme growth conditions**	**Representative habitat**
Acidophiles	pH ≤ 3.0	Acid mine drainage (6), volcanic areas (12)
Alkaliphiles	pH ≥ 10.0	Saline-alkaline lakes and basaltic aquifers (32, 43)
Extreme thermophiles	T ≥ 70^{o}C	Hot springs, geysers, and solfatares(15)
Psychrophiles	T ≤ 15^{o}C	Aquatic environments in polar and alpine regions (16)
Halophiles	Salt concentration ≥ 0.2 M	Salt lakes and saline soils (69)
Endolithic	Lives within or penetrates deeply into stony substances	Ocean crust (63)
Hypolithic	Colonizes the underside of stony substances in arid deserts	Translucent stones in deserts (88)
Metallotolerant	High levels of heavy metals	Acidic mine drainage sites (5)
Oligotrophic	Nutritionally limited environments	Lakes
Barophiles	High pressure	Deep ocean and deep see hydrothermal vents (35)
Radio-resistant	Ionizing radiation	Fresh water and soil (3)
Toxitolerant	Damaging agents	Acidic mine drainage sites (5)
Xerophiles	Dry conditions	Dried foods and house dust (84)

Table 1.2 Examples of industrial extremozymes		
Source extremophile	**Extremozyme**	**Industrial application**
Extreme thermophiles	α-amylases	Starch degradation (90)
	Taq DNA polymerase	Molecular biology (18)
	Lipases and esterases	Biopolymer synthesis (24)
	Xylanases	Pulp and paper industry (22, 26)
Psychrophiles	Lipases	Laundry detergents (45)
	Feruloylesterases	Antimicrobial and antioxidant formulation (4)
Alkaliphiles	Cellulases	Biopolishing of denims (2)
	CGTase	Cyclodextrin production (44)
Halophiles	Proteases	Fermentation of seafood (87)

FIGURES

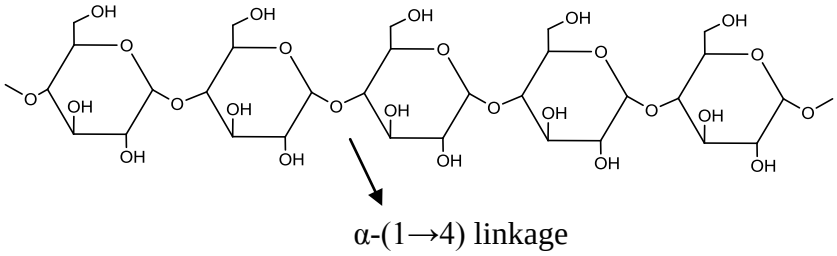

Figure 1.1. Molecular structure of amylose and glycosidic bonds for enzymatic cleavage

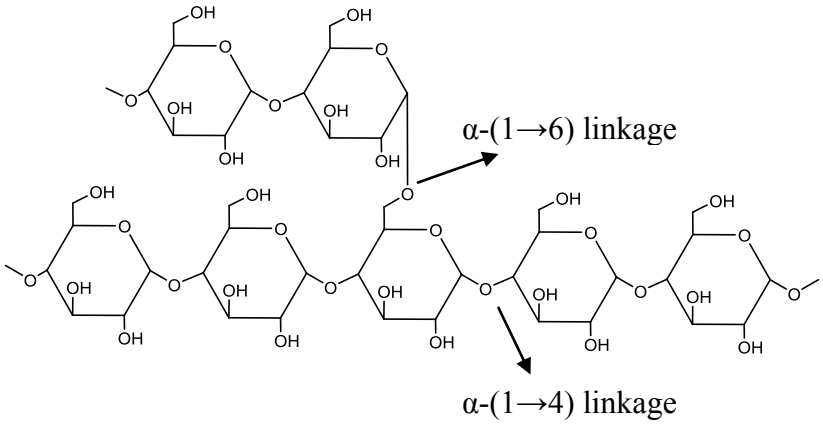

Figure 1.2. Molecular structure of amylopectin and glycosidic bonds for enzymatic cleavage

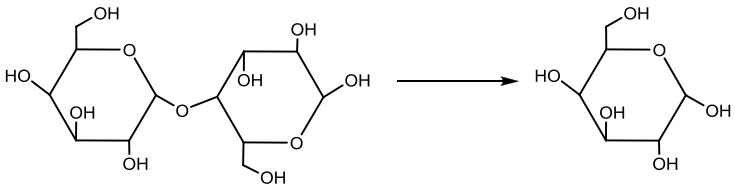

Figure 1.3. Schematic representation of degradation of cellobiose by the attack of β-glucosidase

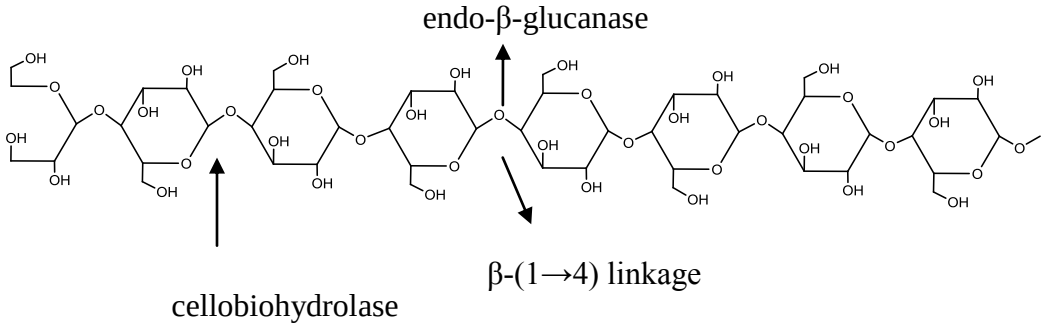

Figure 1.4. Schematic representation of cellulase attack to cellulose chain

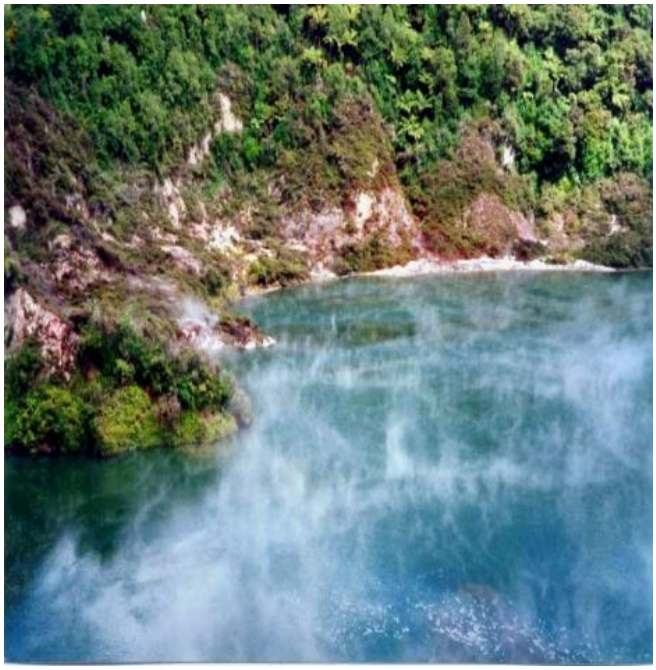

Figure 1.5. A hot spring view from Yellowstone National Park, USA

CHAPTER 2

S-LAYER HOMOLOGY DOMAIN PROTEINS IN *CALDICELLULOSIRUPTOR SACCHAROLYTICUS*

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

Department of Chemical and Biomolecular Engineering,
North Carolina State University
Raleigh, NC 27695-7905

Published in Applied Environmental Microbiology

S-layer homology (SLH) domain proteins Csac_0678 and Csac_2722 implicated in plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*
2012 Feb;78(3)-768-77

ABSTRACT

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, Csac_0678 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 activity. 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 encodes the largest ORF in the *C. saccharolyticus* genome (predicted M_r 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 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.

1. Introduction

Members of the extremely thermophilic genus *Caldicellulosiruptor* have potential as consolidated bio-processing (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 for its 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 (GH) 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 the *Clostridia* (3, 4) and other bacteria and fungi (2). Within the cellulosome are enzymes endowed with carbohydrate binding modules (CBM) that serve to anchor the biocatalyst to the substrate surface, thereby placing the active site in close proximity to the substrate (36). The CBM 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 *Trichiderma reeseii*, 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 multi-domain GHs, ten of which have identifiable signal peptides (55). These ten multi-domain 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 *C. 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_1076-1079, suggesting a constitutive role in carbohydrate utilization (1).

C. saccharolyticus has eleven S-layer homology (SLH) domain proteins, presumably to enable binding to the S-layer. SLH domains, 50-60 amino acids 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 (SCWP) as the anchoring structure (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, 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 they required for substrate binding (9). These domains are implicated in the binding of enzymes to the cell surface so that release of hydrolysis products are in close proximity to, and 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 also 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 towards the role of these proteins in plant biomass deconstruction.

2. Materials and Methods

2.1. Bacterial strains and culture conditions

E. coli NovaBlue and *E. coli* Rosetta^{TM} (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/l), NaCl (0.9 g/l), $MgCl_2.6H_2O$ (0.4 g/l), KH_2PO_4 (0.75 g/l), K_2HPO_4 (1.5 g/l), yeast extract (1 g/ml), trace element solutions SL-10 (1 ml/l), cellobiose (5 g/l), and $Na_2S.9H_20$ (0.05% w/v). Cells were visualized by epifluorescence microscopy (Nikon Intensilight C-HGFI, Lewisville, TX) as previously described (55).

2.2. 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 encoding these genes were amplified by PCR from *C. saccharolyticus* genomic DNA by using the primers listed in Table 2.1. The primers were designed in a way that a C-terminal histidine (His)-tag was added to all three versions of Csac_0678 constructions and an N-terminal His-tag was added to Csac_2722. Gene constructions used to express intact proteins and truncation mutants are shown in Figure 2.1. 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 990-2134 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^{TM} (DE3) cells, containing pET-28b(+) under the control of T7lac promoter (pET System Manual), were used for gene expression. Expression of the target genes was induced by the addition of IPTG to Rosetta cells when OD_{600} reached 0.8. The cells were then harvested after 4 hours of induction by centrifugation at 10,000 *x g* for 10 minutes. The cell pellet was re-suspended with 5 ml of 50 mM sodium phosphate pH 8.0, 100

mM NaCl, Nonidet P40 (0.1% (v/v), 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 minutes with 10 sec off/on pulses. The suspension was heat-treated at 60^{o}C for 20 minutes in order to remove *E. coli* proteins. The cell extract was obtained after centrifuging the lysed and heat-treated cells at 20,000 *x g* for 20 minutes. 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 re-solubilize recombinant Csac_2722. HiTrap HP (GE Life Sciences, Piscataway, NJ), and a Resource Q (GE Life Sciences, Piscataway, NJ) column were used to purify the recombinant proteins on a Biologic DuoFlow FPLC (BioRad, 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. Protein concentration was determined by the Bradford protein assay (Biorad, Hercules, CA), using bovine serum albumin as the standard (Sigma).

2.3. Biochemical characterization of Csac_0678

The optimal pH for recombinant Csac_0678 was determined at 70°C in buffers containing 2% CMC at pH 2.5 to 9 (50 mM sodium acetate buffer, pH 2.5-6, and 50 mM sodium phosphate buffer, pH 6-9). 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 (Calorimetry Sciences Corporation, Provo, UT). Csac_0678 was prepared at 1 mg/ml in PBS buffer and scanned between 25°C and 125°C, using 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 hours. 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), xylan oat spelt (Sigma, St. Louis, MO), xylan birchwood (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; dilute acid pretreatment was performed at the National Renewable Energy Laboratory (44)), lichenan (Megazyme), arabinoxylan (Megazyme) and carboxymethylcellulose (CMC, Sigma) were prepared at 1% (w/v) 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, Kent, UK) 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 1% (w/v) final concentration in each reaction. Bacterial microcrystalline cellulose (BMCC) (provided by David B. Wilson, Cornell University (Ithaca, NY)) at 0.5% (w/v) 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 (U) of enzyme activity is defined as the amount of enzyme required for producing 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, one unit of

enzyme activity is defined as the amount of enzyme required for producing 1 μmol of 4-nitrophenol per minute.

Hydrolysis products were determined by HPLC refractive index detection (2414 RI detector, Waters). Shodex KS-801 and KS-802 (Showa Denko K.K., Kanagawa, Japan) columns operated at 80^{o}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.

2.4. 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 hours. 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 the reactions were run in triplicates.

2.5. 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 one hour at 75°C and 300 rpm. After incubation, the samples were boiled for 5 minutes, and then diluted ten-fold with distilled water. The viscosity of each sample was measured in triplicates by an AR-G2 rheometer (TA instruments, New Castle, DE).

2.6. 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 (20 mg), PH-102 (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 minutes. The sample-polysaccharide mixtures were centrifuged at 15,800 *x g* for 5 minutes. 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% (w/v) SDS and 5%

(v/v) β-mercaptoethanol) by incubating in a boiling water bath for 10 minutes. These samples were then centrifuged at 15,800 *x g* for 5 minutes 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 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% (w/v) Avicel solution in 50 mM sodium acetate buffer at pH 5 and at 4°C for one hour 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 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.

2.7. Localization of Csac_0678 and Csac_2722 in *C. saccharolyticus* - extraction of the S-layer

Four liters of *C. saccharolyticus* cells grown on cellobiose were harvested at early stationary phase by centrifugation at 12,100 *x g* for 10 minutes. 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 centrifuging twice at a lower speed, 1,940 *x g* for 5 minutes. The resulting supernatant was centrifuged at 39,200 *x g* for 20 min. The pellet, re-suspended 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 hours at room temperature on a shaker at 100 rpm. The mixture was centrifuged for 30 minutes at 25,000 *x g*. The pellet, which contained the S-layer fraction, was washed three times with 50 mM Tris-HCl (pH 7.4) and re-suspended in distilled water.

2.8. Protein immunoblotting

Proteins in the S-layer fraction and recombinant Csac_0678 and Csac_2722 were resolved by SDS-PAGE using Nu-Page 4-12% Bis-Tris (Invitrogen) pre-cast gels. The proteins were transferred to a nitrocellulose membrane (Whatman) by using Fisher Biotech Semi-Dry blotting unit FB-SDB-2020. The blot was incubated with nonfat dry milk to block non-specific 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 two hours. After rinsing to remove the unbound rabbit antibodies, the blot was exposed for an hour 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.

3. Results and Discussion

3.1. Csac_0678 is a bifunctional endoglucanase/xylanase that binds to the S-layer

The sequenced *Caldicellulosiruptor* genomes encode ORFs representing putative glycoside hydrolases that are associated with S-layer homology domains (see 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 non-catalytic domains (53, 55), but only Csac_0678 has SLH domains, in addition to glycoside hydrolase (GH5) and sugar-binding (CBM28) domain.

Orthologs of Csac_0678 exist (Table 2.2) in all eight sequenced *Caldicellulosiruptor* species; in 6 species, including *C. sacharolyticus*, the ORF encodes a 755-756 aa protein (85 kDa), while in *C. owensensis* and *C. obsidiansis*, the ortholog is 566-567 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.

Csac_0678 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 (see Figure 2.1A). The molecular assembly of recombinant Csac_0678 (M_r of 81 kDa without the signal peptide), as determined by size exclusion chromatography, was found to be monomeric. 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 hours of incubation at 75°C (Figure 2.2). The half-life at 77°C was found to be approximately 19 hours at 77°C, less than 3 hours 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 (Figure 2.1C) 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.

3.2. Biochemical characterization of Csac_0678

Recombinant Csac_0678 was tested for activity towards a range of complex carbohydrates (Table 2.3). Activity on barley glucan and lichenan indicated a preference towards 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. But, in comparison with XynA (306 $U \cdot mg^{-1}$) and XynB (4,600 $U \cdot mg^{-1}$) from *Thermotoga maritima* (60), Csac_0678 xylanase activity was very low. However, it was comparable to thermophilic xylanases from *Bacillus* sp. (NCIM59) (0.0172 and 0.742 $\mu M \cdot min^{-1} \cdot mg^{-1}$) (15). Csac_0678 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 2.3 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 GH5 domain, had dual activity against both glucan- and mannan-based polysaccharides (41). The specific activity of Cel5A on CMC was 616 $U \cdot mg^{-1}$, but no activity was detected on xylan (birch wood) (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 $U \cdot mg^{-1}$ against CMC and Avicel, respectively (42), but no activity was observed either on xylan or on mannan. RuCelA, cloned from a metagenomic library of yak rumen microorganisms, encoded a GH5 enzyme possessing xylanase (264.1 $U \cdot mg^{-1}$) and endoglucanase (54.3 $U \cdot mg^{-1}$) activity, but has 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 towards Avicel ($6.6 \ 10^{-4}$ $U \cdot mg^{-1}$) and filter paper ($6.26 \ 10^{-4}$ $U \cdot mg^{-1}$) (58). Hydrolysis products from Csac_0678 on Avicel, determined by TLC and HPLC, indicated that the enzyme generated predominantly cellobiose and some glucose (Figure 2.3). 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 detected (data not shown). Interestingly, Csac_0678 also cleaved xylan at multiple sites, generating xylobiose, xylotriose, and xylotetraose (see Figure 4) and, 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 or 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±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 0.096-1.42 (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/insoluble reducing sugars. For example, 96% of the reducing sugars produced by an exocellulase, Cel48A from *Thermobifida fusca* were found to be soluble (soluble to insoluble reducing sugar ratio is 24) (30).

CMC viscosity before and after GH addition can also be used to distinguish between endo- and exo-glucanase activity (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 CMC (29, 30). Processive endoglucanases decrease the viscosity of CMC solution as they attack the cellulose chain in a similar way to classical endoglucanases. But they differ from other endoglucanases and exoglucanases since they processively cleave cellobiose units from the cellulose chain (58, 59). Cel48, an 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% (w/v) CMC solution by 40% in one hour, further supporting the contention that it is an endoglucanase.

3.3. 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 (refer to Figure 2.1A): 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 2.5A and 2.5C 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 Figure 2.5C. Eliminating SLH domain did not affect the activity of

Csac_0678 (data not shown) indicating that SLH domains do not play a role in activity.

3.4. Csac_2722 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 (see Table 2.4) (26). These proteins potentially represent an element of the complex carbohydrate recruitment strategy used by these bacteria. For example, the *C. kristjanssonii* and *C. 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 *C. kronotskyensis* genomes also encode SLH + FN3 domain proteins, Athe_0012 and Calkro_0014, respectively; these are each approximately 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 proteins 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 a 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 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). Pull-down

experiments with Avicel as bait showed that Csac_2722 binds to Avicel (~32% bound) (Figure 2.5B). Csac_2722 also exhibited activity on CMC (2.8 mM glucose equivalent of reducing sugars were released in 3 hours), 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 2.4, contribute to lignocellulose conversion in *Caldicellulosiruptor* species.

4. Conclusion

Consolidated bioprocessing (CBP) describes the comprehensive capacity of a microbial system to not only significantly deconstruct lignocellulose but also to convert the hydrolysis products to a biofuel. This metabolic capability no doubt involves 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 2.6A shows the attachment of *C. saccharolyticus* cells to switchgrass and Figure 2.6B illustrates how Csac_0678 and Csac_2722 contribute to the deconstruction of plant biomass. Contributions from non-catalytic, 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 US National Science Foundation (CBT0617272) and the Bioenergy Science Center (BESC), a U.S. DOE Bioenergy Research Center supported by the Office of Biological and Environmental Research. The authors also wish to thank Allen Foegeding (Department of Food Science, NCSU), Sunkyu Park (Department of Forest Biomaterials, NCSU), Christina Tang and Saad Khan (Department of Chemical and Biomolecular Engineering, NCSU) for their help with experimental aspects of this effort. The gift of bacterial microcrystalline cellulose from David Wilson and Maxim Kostylev (Cornell University) was also appreciated.

REFERENCES

1. Andrews, G., D. Lewis, J. Notey, R. Kelly, and D. Muddiman. 2010. Part I: characterization of the extracellular proteome of the extreme thermophile *Caldicellulosiruptor saccharolyticus* by GeLC-MS2. Anal. Bioanal. Chem. 398:377-389.
2. Bayer, E. A., H. Chanzy, R. Lamed, and Y. Shoham. 1998. Cellulose, cellulases and cellulosomes. Curr. Opin. Struct. Biol. 8:548-557.
3. Bayer, E. A., R. Kenig, and R. Lamed. 1983. Adherence of *Clostridium thermocellum* to cellulose. J. Bacteriol. 156:818-827.
4. Bayer, E. A., E. Setter, and R. Lamed. 1985. Organization and distribution of the cellulosome in *Clostridium thermocellum*. J. Bacteriol. 163:552-559.
5. Bayer, E. A., L. J. Shimon, Y. Shoham, and R. Lamed. 1998. Cellulosomes-structure and ultrastructure. J. Struct. Biol. 124:221-234.
6. Blumer-Schuette, S. E., I. Kataeva, J. Westpheling, M. W. Adams, and R. M. Kelly. 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr. Opin. Biotechnol. 19:210-217.
7. Blumer-Schuette, S. E., D. L. Lewis, and R. M. Kelly. 2010. Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. Appl. Environ. Microbiol. 76:8084-8092.

8. Boraston, A. B., D. N. Bolam, H. J. Gilbert, and G. J. Davies. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382:769-781.

9. Cann, I. K., S. Kocherginskaya, M. R. King, B. A. White, and R. I. Mackie. 1999. Molecular cloning, sequencing, and expression of a novel multidomain mannanase gene from *Thermoanaerobacterium polysaccharolyticum*. J. Bacteriol. 181:1643-1651.

10. Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37:D233-238.

11. Chang, L., M. Ding, L. Bao, Y. Chen, J. Zhou, and H. Lu. 2011. Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. Appl. Microbiol. Biotechnol. 90:1933-1942.

12. Chauvaux, S., M. Matuschek, and P. Beguin. 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, S. R., K. R. Shockley, D. E. Ward, and R. M. Kelly. 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., A. E. Mars, M. A. Budde, M. H. Lai, C. Dijkema, P. de Waard, and P. A. Claassen. 2007. Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus*. Appl. Microbiol. Biotechnol. 74:1358-1367.

15. Dey D., J. H., Abhay Shendye, and Mala Rao. 1992. Purification and properties of extracellular endoxylanase from alkalophilic thermophilic *Bacillus* sp. Can. J. Microbiol. 38:436-442.

16. Engelhardt, H., and J. Peters. 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., C. Huber-Gries, T. Pavkov, W. Keller, C. Mader, N. Ilk, U. B. Sleytr, and E. M. Egelseer. 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, C. M., and H. J. Gilbert. 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu. Rev. Biochem. 79:655-681.

19. Fraiberg, M., I. Borovok, E. A. Bayer, R. M. Weiner, and R. Lamed. 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., I. Borovok, R. M. Weiner, and R. Lamed. 2010. Discovery and characterization of cadherin domains in *Saccharophagus degradans* 2-40. J Bacteriol 192:1066-1074.
21. Fraser, J. S., Z. Yu, K. L. Maxwell, and A. R. Davidson. 2006. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. J. Mol. Biol. 359:496-507.
22. Geslin, C., M. Le Romancer, G. Erauso, M. Gaillard, G. Perrot, and D. Prieur. 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., I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon, and G. Davies. 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., G. Rakhely, Z. Bagi, G. Ivanova, N. Acs, E. Kovacs, and K. L. Kovacs. 2010. Exploitation of the extremely thermophilic *Caldicellulosiruptor saccharolyticus* in hydrogen and biogas production from biomasses. Environ. Technol. 31:1017-1024.

26. Hunter, S., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, and C. Yeats. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res. 37:D211-215.

27. Igarashi, K., T. Ishida, C. Hori, and M. Samejima. 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., D. H. Shin, S. Zhang, B. K. Barr, J. Sakon, P. A. Karplus, and D. B. Wilson. 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, D. C., M. Spezio, L. P. Walker, and D. B. Wilson. 1993. Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. Biotechnol. Bioeng. 42:1002-1013.

30. Irwin, D. C., S. Zhang, and D. B. Wilson. 2000. Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. Eur. J. Biochem. 267:4988-4997.

31. Ito, Y., T. Tomita, N. Roy, A. Nakano, N. Sugawara-Tomita, S. Watanabe, N. Okai, N. Abe, and Y. Kamio. 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., Q. Xu, R. Kenig, M. Shulman, Y. Shoham, E. A. Bayer, and R. Lamed. 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., T. de Vrije, G. E. van Noorden, M. A. Budde, Z. Szengyel, K. Reczey, and P. A. Claassen. 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., K. Murashima, Y. Tamaru, and R. H. Doi. 2002. Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. J. Bacteriol. 184:884-888.

35. Lehtio, J., J. Sugiyama, M. Gustavsson, L. Fransson, M. Linder, and T. T. Teeri. 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., K. T. Leung, and W. Qin. 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. Int. J. Biol. Sci. 5:500-516.

37. Martinez, D., R. M. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S. E. Baker, J. Chapman, O. Chertkov, P. M. Coutinho, D. Cullen, E. G. Danchin, I. V. Grigoriev, P. Harris, M. Jackson, C. P. Kubicek, C. S. Han, I. Ho, L. F. Larrondo, A. L. de Leon, J. K. Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A. A. Salamov, M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C. L. Schoch, J. Yao, R. Barabote, M. A. Nelson, C. Detter, D. Bruce, C. R. Kuske, G. Xie, P. Richardson, D. S. Rokhsar, S. M. Lucas, E. M. Rubin, N. Dunn-Coleman, M. Ward, and T. S. Brettin. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat. Biotechnol. 26:553-560.

38. Matuschek M, B. G., Sahm K, 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, G. L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. 31:426-428.

40. Mizuno, M., T. Tonozuka, S. Suzuki, R. Uotsu-Tomita, S. Kamitori, A. Nishikawa, and Y. Sakano. 2004. Structural insights into substrate specificity and function of glucodextranase. J. Biol. Chem. 279:10575-10583.

41. Pereira, J. H., Z. Chen, R. P. McAndrew, R. Sapra, S. R. Chhabra, K. L. Sale, B. A. Simmons, and P. D. Adams. 2010. Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*. J. Struct. Biol. 172:372-379.

42. Posta, K., E. Beki, D. B. Wilson, J. Kukolya, and L. Hornok. 2004. Cloning, characterization and phylogenetic relationships of cel5B, a new endoglucanase encoding gene from *Thermobifida fusca*. J. Basic. Microbiol. 44:383-399.

43. Rainey, F. A., A. M. Donnison, P. H. Janssen, D. Saul, A. Rodrigo, P. L. Bergquist, R. M. Daniel, E. Stackebrandt, and H. W. Morgan. 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., C. Pan, G. B. Hurst, M. Rodriguez, Jr., C. K. McKeown, P. K. Lankford, N. F. Samatova, and J. R. Mielenz. 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., C. Hotzy, I. Schocher, U. B. Sleytr, and M. Sara. 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, E.F., Maniatis, T. 1989. Molecular Cloning Laboratory Manual, 2nd ed, vol. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

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, D. J., L. C. Williams, R. A. Grayling, L. W. Chamley, D. R. Love, and P. L. Bergquist. 1990. celB, a gene coding for a bifunctional cellulase from the extreme thermophile "*Caldocellum saccharolyticum*". Appl. Environ. Microbiol. 56:3117-3124.

49. Sleytr, U. B. 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, V. S., D. J. Saul, and P. L. Bergquist. 1995. celA, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. Appl. Microbiol. Biotechnol. 43:291-296.

51. Tschiggerl, H., A. Breitwieser, G. de Roo, T. Verwoerd, C. Schaffer, and U. B. Sleytr. 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., R. Takada, Y. Araki, K. Suzuki, S. Karita, T. Wakagi, H. Shoun, T. Watanabe, and S. Fushinobu. 2010. Recognition of cellooligosaccharides by a family 28 carbohydrate-binding module. FEBS Lett. 584:1205-1211.

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

54. Vanfossen, A. L., D. L. Lewis, J. D. Nichols, and R. M. Kelly. 2008. Polysaccharide degradation and synthesis by extremely thermophilic anaerobes. Ann. N.Y. Acad. Sci. 1125:322-337.

55. Vanfossen, A. L., I. Ozdemir, S. L. Zelin, and R. M. Kelly. 2011. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Biotechnol. Bioeng. 108:1559-1569.

56. Vanfossen, A. L., M. R. Verhaart, S. M. Kengen, and R. M. Kelly. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl. Environ. Microbiol. 75:7718-7724.

57. Vuong, T. V., and D. B. Wilson. 2009. Processivity, synergism, and substrate specificity of *Thermobifida fusca* Cel6B. Appl. Environ. Microbiol. 75:6655-6661.

58. Watson, B. J., H. Zhang, A. G. Longmire, Y. H. Moon, and S. W. Hutcheson. 2009. Processive endoglucanases mediate degradation of cellulose by *Saccharophagus degradans*. J. Bacteriol. 191:5697-5705.

59. Wilson, D. B. 2011. Microbial diversity of cellulose hydrolysis. Curr. Opin. Microbiol. 14:259-263.

60. Winterhalter, C., and W. Liebl. 1995. Two Extremely Thermostable Xylanases of the Hyperthermophilic Bacterium *Thermotoga maritima* MSB8. Appl. Environ. Microbiol. 61:1810-1815.

61. Xiao, Z., R. Storms, and A. Tsang. 2005. Microplate-based carboxymethylcellulose assay for endoglucanase activity. Anal. Biochem. 342:176-178.

62. Xiao, Z., R. Storms, and A. Tsang. 2004. Microplate-based filter paper assay to measure total cellulase activity. Biotechnol. Bioeng. 88:832-837.

63. Xu, Q., E. A. Bayer, M. Goldman, R. Kenig, Y. Shoham, and R. Lamed. 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, X. Z., Z. Zhang, Z. Zhu, N. Sathitsuksanoh, Y. Yang, and Y. H. Zhang. 2010. The noncellulosomal family 48 cellobiohydrolase from *Clostridium phytofermentans* ISDg: heterologous expression, characterization, and processivity. Appl. Microbiol. Biotechnol. 86:525-533.

65. Zhang, Y. H., J. Cui, L. R. Lynd, and L. R. Kuang. 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, Y. H., and L. R. Lynd. 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., D. C. Irwin, J. Escovar-Kousen, and D. B. Wilson. 2004. Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. Biochemistry 43:9655-9663.

TABLES

Table 2.1. Primers used for cloning Csac_0678 and Csac_2722 genes

Clone	F/R	Primer
Csac_0678	F	GGCGGCCCATGGATAATACTGCGTATGAAAAGGATAAGTATCCACAC
	R	GGCGCCTCGAGCATCTTTCCTGTAAGTTCTAAAATTTTGTA
Csac_0678-TM1	F	GGCGGCCCATGGATAATACTGCGTATGAAAAGGATAAGTATCCACAC
	R	GGCGGCCTCGAGATTTGTAAATCTTACATTGT
Csac_0678-TM2	F	GGCGGCCCATGGATAATACTGCGTATGAAAAGG
	R	GGCGGCCTCGAGAGCACCAGAAGTCTCATT TT
Csac_2722	F	GGCGGCGCTAGCCTGTTCATAAAGAGTACA
	R	GAAGACCTCGAGCTATTGAGCCTGTCCATAGGTGGC

Table 2.2. SLH-domain proteins in *Caldicellulosiruptor* genomes that contain sugar binding domains and GH domains

DOMAINS[a]	ORF	# Amino Acids
GH5-CBM28-SLH-SLH-SLH	Csac_0678	755
	Calkro_2036	
	Calla_0352	756
	Calhy_2064	
	Calkr_2007	
	Athe_0594	
	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	1,732
	Calhy_0060	
	COB47_0076	
	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	Calhy_2383	2,007

[a]Abbreviations used are as follows: SLH (S-layer homology domain); FN3 (Fibronectin III domain); Big4 (Ig-like domain); CBM (cellulose binding domain, family #); 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 two separate proteins.

[c]Total size of Csac_2548 and Csac_2549.

Table 2.3. Hydrolytic activity of Csac_0678 on complex carbohydrates

Substrate	**Linkage**	**Specific activity*** ($U \cdot mg^{-1}$enzyme)
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
Xylan Oat Spelt	β-1,4-xylan	0.709
Xylan Birchwood	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
Laminarin	β-1,3/6-glucan	ND

*Specific activities are based on triplicate analysis for each substrate and account for abiotic hydrolysis.

All standard deviations were less than ± 10%

ND = No detectable activity
CMC = Carboxymethyl Cellulose
PASC = Phosphoric Acid Swollen Cellulose
BMCC = Bacterial Microcrystalline Cellulose
pNPC = *p*-nitrophenyl-β-D-cellobioside

Table 2.4. SLH-domain proteins in *Caldicellulosiruptor* genomes that contain sugar binding domains but no GH domains		
DOMAINS[a]	**ORF**	**# 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	575
	Calow_1583	
	Csac_2381	
	Calkro_0875	576
Big4-SLH-SLH-(SLH)	Calkro_0550	1,789
	Calkro_0550	
	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	575
	Calkr_0834	
(Big)-(CBM20-CBM20)-SLH-SLH-SLH-CBM20	Calla_0367	1,097
	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]Abbreviations used are as follows: SLH (S-layer homology domain); FN3 (Fibronectin III domain); vWFA (von Willebrand domain); SH3 (src Homology-3 domain); GBD (Galactose Binding Domain-like); Big1, 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)

FIGURES

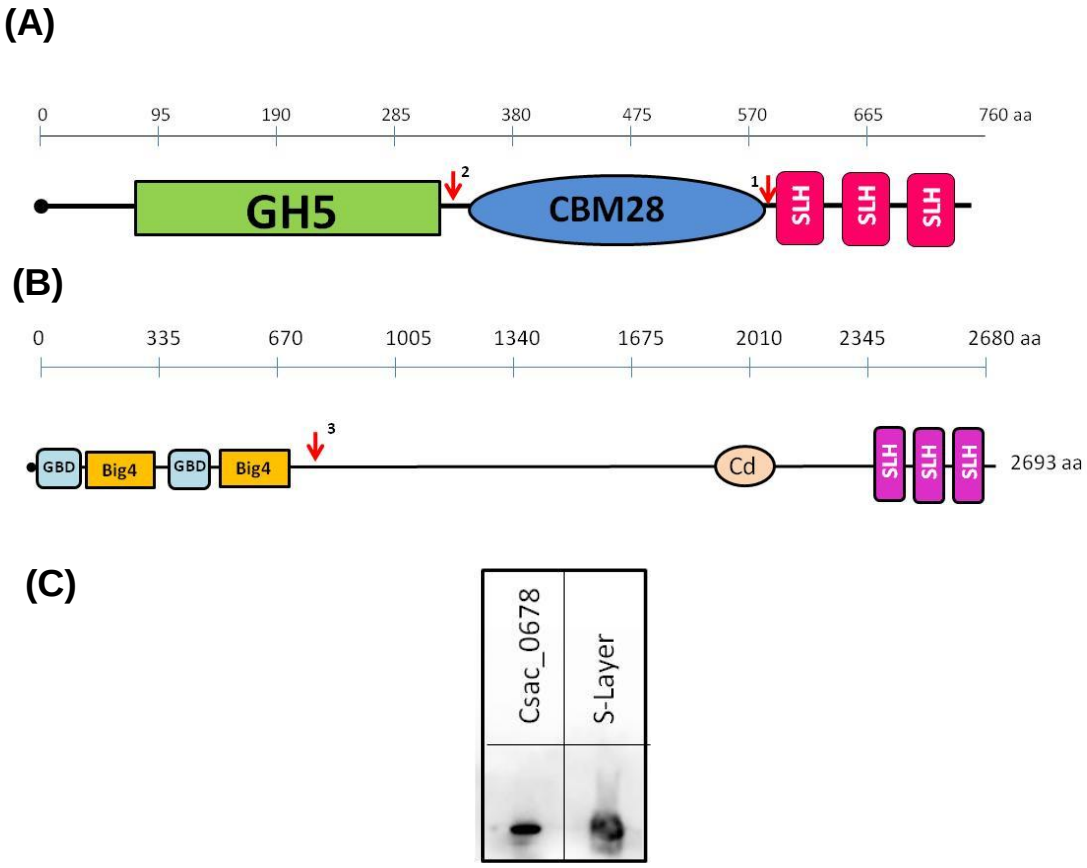

Figure 2.1. Csac_0678 and Csac_2722 conserved domains and truncation points for deletion mutants(A) Recombinant intact Csac_0678 (727 aa), lacking only the signal peptide. Csac_0678-TM1 (575 aa) and Csac_0678-TM2 (317 aa) were constructed by truncating at the points 1 and 2, respectively. (B) Recombinant Csac_2722 (810 aa), without the signal peptide, was truncated at point 3. LEGEND - 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; aa: amino acid. (C) Western Blot showing the localization of Csac_0678 to the S-layer for *C. saccharolyticus* grown on cellobiose. The lane labeled as "Csac_0678" had purified recombinant Csac_0678. The lane labeled as "S-layer" is the S-layer fraction extracted from native C. saccharolyticus cells. Negative controls with an related protein did not have any bands (data not shown).

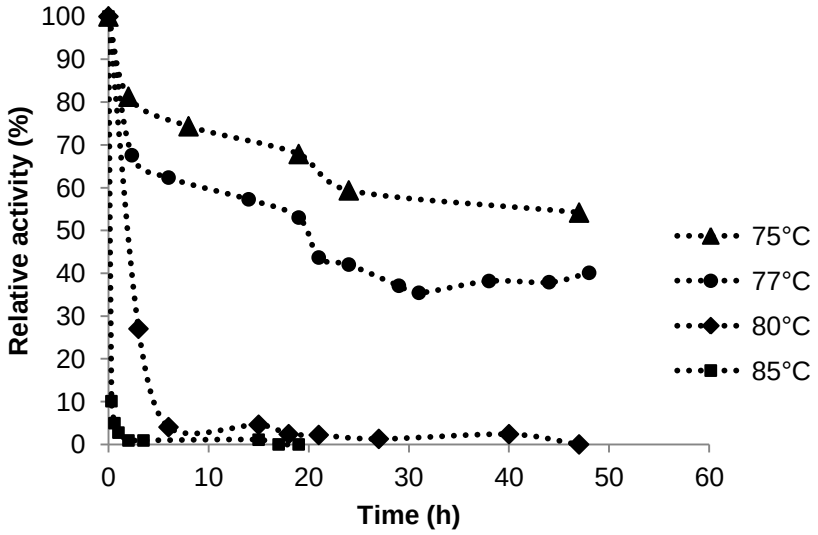

Figure 2.2 Comparison of the thermostabilities of Csac_0678 at different incubation temperatures.The half-life at 75°C was ~48 h, and that at 77°C was 19 h. Csac_0678 lost 50% of initial activity in less than 3 h at 80°C and in less than 30 min at 85°C.

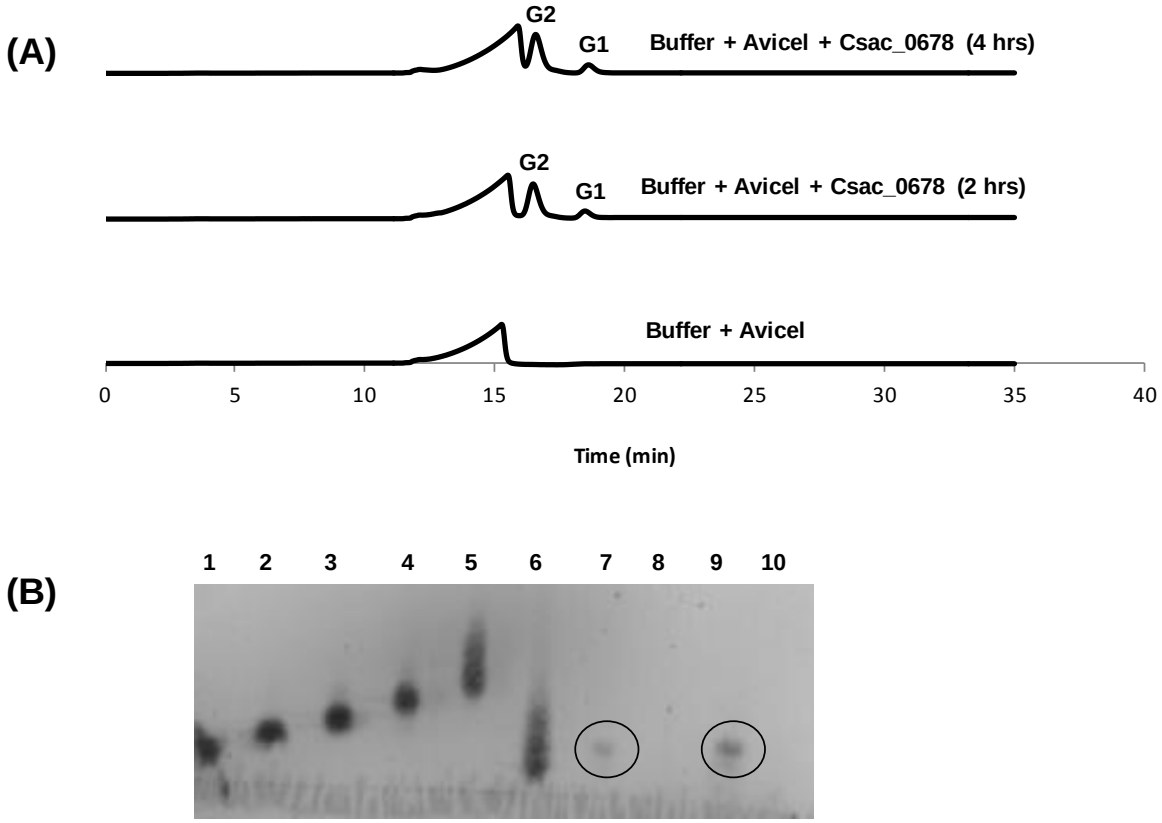

Figure 2.3. Hydrolysis of cellulose by Csac_0678 at 70°C. (A) HPLC (RI) analysis shows glucose and cellobiose produced after 2 and 4 hour incubations of enzyme with Avicel (100 µm). (B) TLC for incubation of Csac_0678 for 4 hrs on two different particle sizes of cellulose (20 and 100 µm). TLC Standards, 1: Glucose, 2: Cellobiose. 3: Cellotriose, 4: Cellotetraose, 5: Cellopentaose. Incubation with Csac_0678 with 6: Cellulose (100 µm) and 8: Cellulose (20 µm). Controls were 7: Cellulose (100 µm) and 9: Cellulose (20 µm), in each case with no added enzyme.

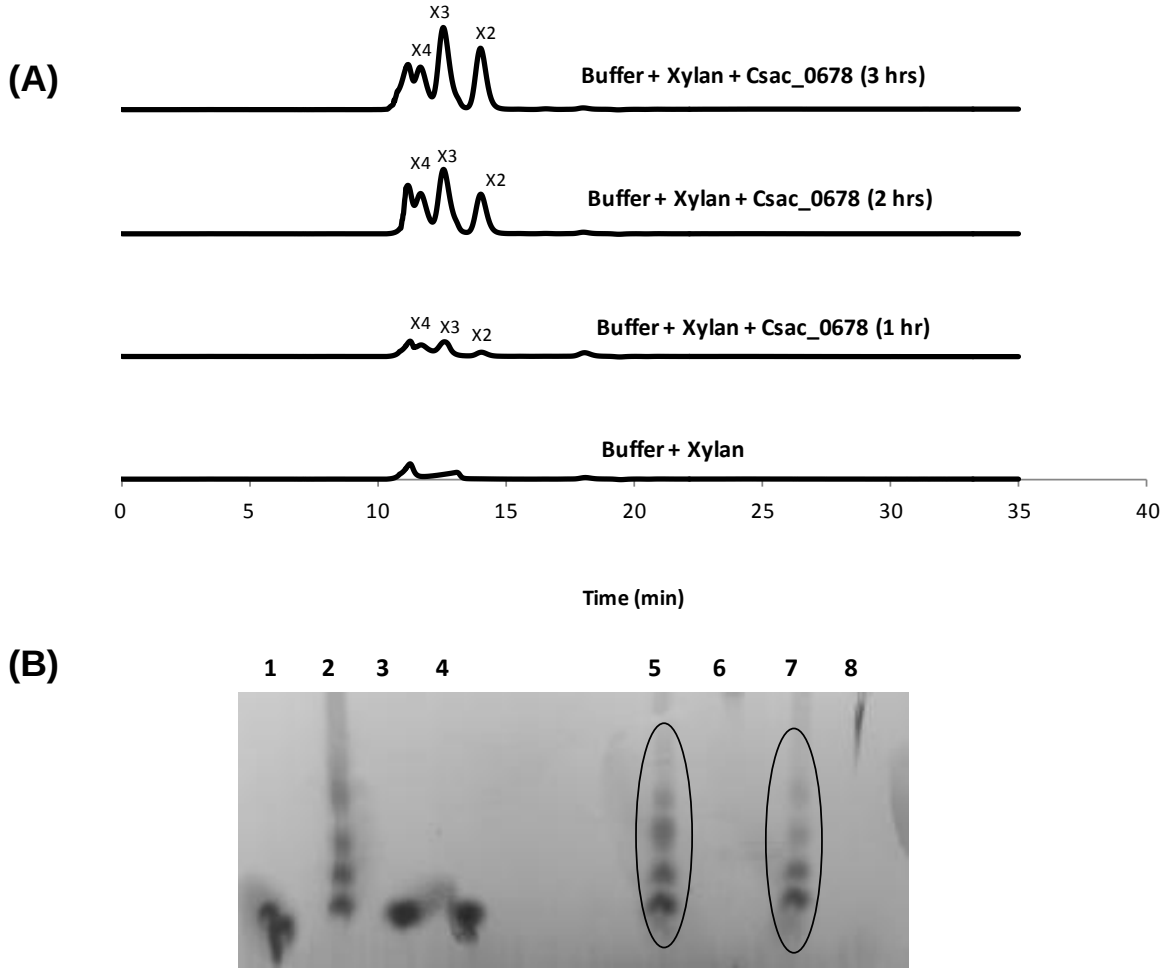

Figure 2.4. Hydrolysis of xylan by Csac_0678 at 70°C. (A) HPLC (RI) analysis shows xylobiose, xylotriose and xylotetraose released upon incubation with the enzyme. (B) TLC for incubation of xylan with the enzyme for 4 hours. Standards, 1: Xylose, 2: Xylooligosaccharides, 3: Arabinose, 4: Glucose. Reactions with Csac_0678 on 5: Xylan birchwood and 7: Xylan oat spelt. Controls were 6: Xylan birchwood and 8: Xylan oat spelt, in each case with no enzyme added.

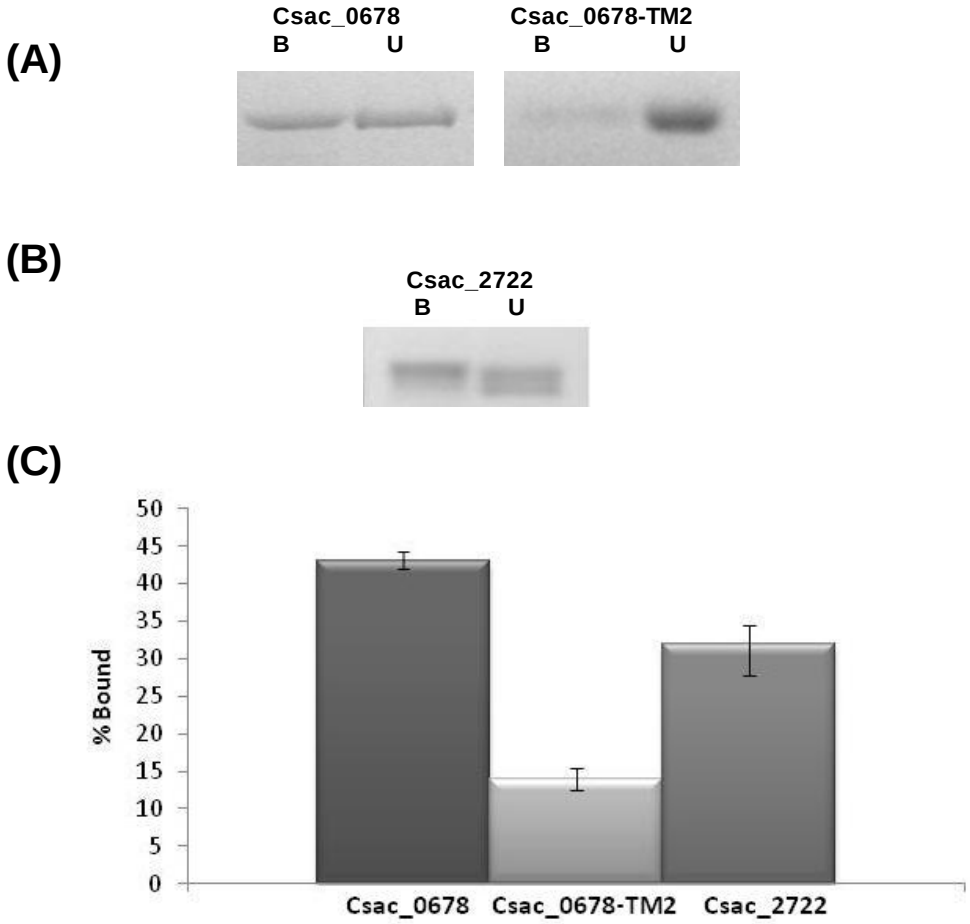

Figure 2.5. Csac_0678 and Csac_2722 binding to Avicel. Csac_0678 binding to Avicel was decreased with the deletion of the CBM28 domain. (A) and (B) Shown are SDS-PAGE with bound (B) and unbound (U) fractions of Csac_0678 with and without CBM domains and Csac_2722, respectively from carbohydrate (Avicel) affinity experiment. (C) Binding of Csac_0678, Csac_0678 without the binding domain, and Csac_2722 to Avicel.

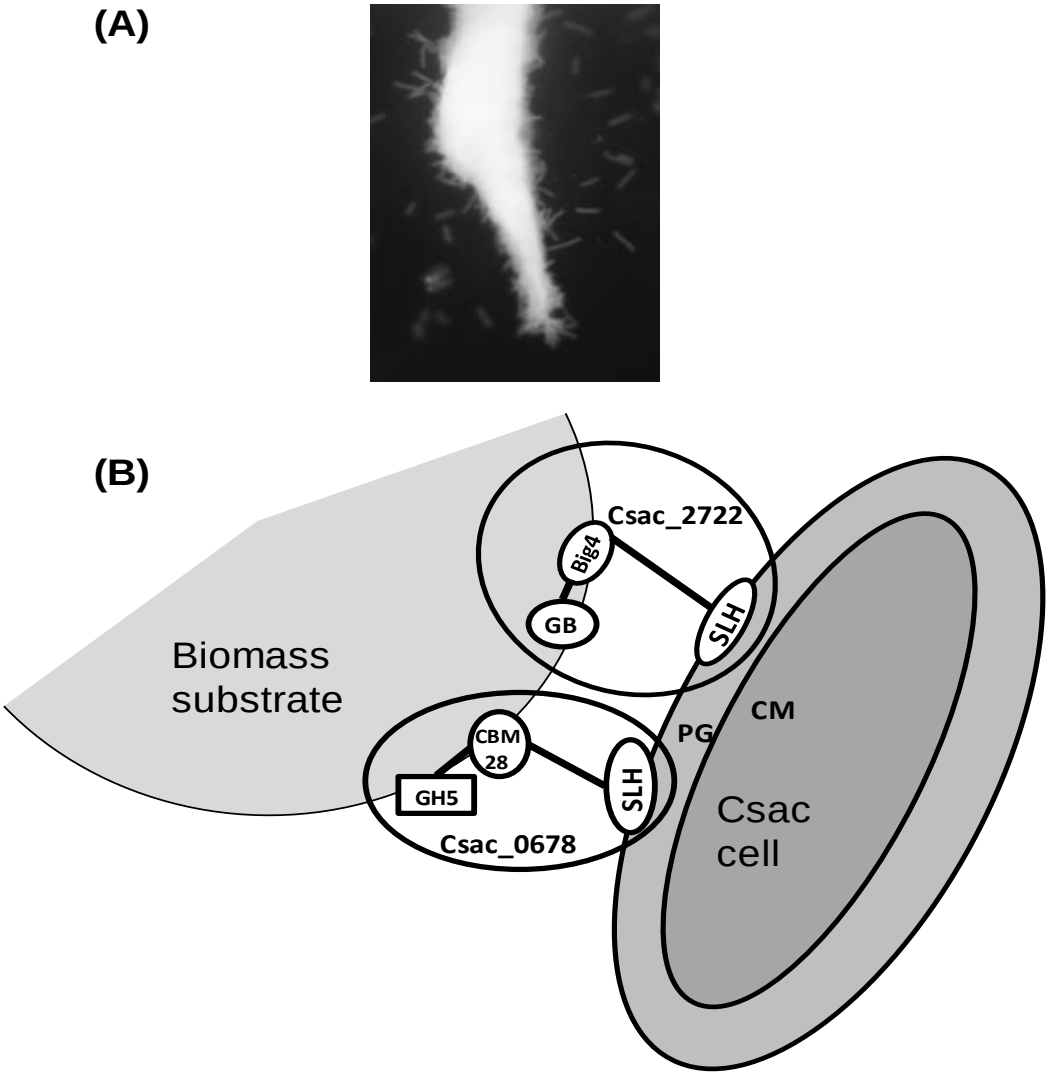

Figure 2.6. Proposed role of Csac_0678 and Csac_2722 in plant biomass deconstruction by *C. saccharolyticus.* (A) Epifluorescence micrograph (acridine orange stain) showing *C. saccharolyticus* (rods) attachment to acid pre-treated switchgrass particle. (B) Schematic representation of cell wall and carbohydrate attachment of Csac_0678 and Csac_2722. LEGEND- 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.

CHAPTER 3

DECONSTRUCTION OF PLANT BIOMASS BY EXTREMELY THERMOPHILIC *CALDICELLULOSIRUPTOR* SPECIES INVOLVES NOVEL MULTI-DOMAIN HEMICELLULASES ASSOCIATED WITH THE S-LAYER

Inci Ozdemir and Robert M. Kelly*

Department of Chemical and Biomolecular Engineering,
North Carolina State University
Raleigh, NC 27695-7905

To be submitted to: *Biotechnology and Bioengineering*

ABSTRACT

Biological deconstruction of lignocellulose requires the synergistic action of a range of glycoside hyrolases with complementary functions to overcome the heterogeneous and recalcitrant nature of lignocellulose. The analysis of genome sequences of eight different extremely thermophilic *Caldicellulosiruptor* species showed that, while they do not contain cellulosomes, they do encode large, extracellular, multi-domain cellulolytic and hemicellulolytic enzymes with surface layer (S-layer) homology domains. To examine the biochemical characteristics of these enzymes, several representatives from this group, including Csac_0678 (GH5+CBM28+SLH+SLH), Calkr_2245 (CBM22-CBM22-CBM22-GH10-CBM9-CBM9-CBM9-CE15-SLH-SLH), Calhy_0060 (SLH-SLH-SLH-CBM54-GH16-CBM4-CBM4-CBM6-CBM4-CBM4-CBM4), Calkro_0111 (SLH-SLH-SLH-CBM54-FN3-GH16-FN3-FN3-GH55-CBM32), Calhy_1629 (SLH-SLH-SLH-GH43-CBM54), and Calhy_2383 (CBM35-GH87-FN3-SLH-SLH) were expressed heterologously in *Escherichia coli* without SLH domains either with all catalytic and binding domains or as truncation mutants (TM). Calkr_2245-TM1 (expressed as CBM22+GH10+CBM9) exhibited xylanase activity, degrading xylan alone and synergistically with recombinant β-xylosidase, Calhy_1629. Calhy_0060 degraded lichenan, laminarin, and β-glucan, exhibiting β-1,3-1,4-glucanase activity. Two truncated versions of of Calkro_0111 were produced: TM1 (CBM54-FN3-GH16-FN3) and TM2

(FN3+GH55+CBM32). Calkro_0111-TM1 degraded the same substrates as Calhy_0060 did, exhibiting lichenase activity. However, Calkro_0111-TM2 showed only laminarinase activity. Calhy_2383 was not characterized but considering the fact that it has a family 87 glycoside hydrolase domain, it is expected to have α-1,3-glucanase activity. Csac_0678's biochemical features were reported previously (Ozdemir et al., 2012, Appl. Environ. Microbiol., 78:768-777). Here, the crystal structure of the GH5 domain from this enzyme was solved and found to have a TIM barrel protein fold. Polyclonal antibodies directed against Csac_0678 were used to image the localization of this enzyme on the cell envelope, revealing that it was cell-surface associated and also proteomics data showed that Csac_0678 was found bound to the substrate and to the whole cell. The *Caldicellulosiruptor* GHs associated with SLH domains examined here were largely hemicellulolytic, and not cellulololytic. This suggests that they play a role in removing hemicellulose from plant biomass substrates, perhaps to facilitate access of *Calidcellulosiruptor* extracellular cellulases to the cellulose fractions.

1. Introduction

The plant cell wall consists of a complex combination of cellulose, hemicellulose, pectin, and lignin (17). Cellulose is a homogeneous linear polymer composed of β-1,4-linked glucose molecules. Individuals chains of repeating glucose molecules are arranged in a very regular and ordered manner by intra-chain and inter-chain hydrogen bonds, conferring the recalcitrant crystalline structure of this complex polysaccharide (37). Hemicelluloses are polysaccharides containing β-1,4-linked backbones substituted with side chains, strengthening the cell wall by interacting with cellulose and lignin. Hemicelluloses include xylans, xyloglucans, mannans, arabinans and glucomannans, and β-(1,3)(1,4)-glucans (15, 47). Pectic polysaccharides, including homogalacturonans, xylogalacturonans, apiogalacturonans, and rhamnogalacturonans, contribute to the mechanical strength, porosity, adhesion, and rigidity of the cell wall (17). Lignin is a large aromatic polyphenolic molecule, which is a major an obstacle in the lignocelluloses degradation (16, 17).

The complete degradation of each carbohydrate component of lignocellulose requires the cooperative action of a variety of glycolytic enzymes specific for the different linkages between building blocks. For cellulose degradation, endoglucanases, cellobiohydrolases, and β-glucosidases are involved. Endoglucanases cleave internal β-1,4-glycosidic bonds randomly in the amorphous region of cellulose. Cellobiohydrolases processively cleave glucose or cellobiose

from the reducing or non-reducing ends of cellulose. β-glucosidases remove glucose from cellobiose and cellodextrins (37). Xylan, the major hemicellulose found in the plant cell wall, has a β-1,4-linked xylose backbone substituted by acetyl, L-arabinofuranosyl, glucuronyl and 4-*O*-methylglucuronyl side chains. Endo-1,4- β-D-xylanase and β-xylosidase are required to cleave the xylose back-bone. Additionally, α-L-arabinofuranosidase and α-D-glucuronidase to cleave the side chains, acetyl xylan esterase to hydrolyze ester bonds between acetyl groups and the backbone of xylan, and p-coumaryl esterase and feruloyl esterase to degrade the linkages between xylan and lignin, are needed (38). β-1,3- and β-1,3-1,4-glucans are constituents of the plant cell wall. β-1,3-glucans are found in the reproductive structures of plants (callose), in marine macro-algae *Laminaria saccharina*, in the cell walls of yeast and fungi, and in the *Graminaceae* (e.g., barley). *Cetraria islandica* (lichenan) also has alternating β-1,3- and β-1,4-glycosidic linkages (24).

Due to its high complexity, degradation of lignocellulosic biomass to fermentable sugars is not trivial. Nonetheless, microorganisms that degrade lignocellulose are associated with variety of habitats and are highly diverse. For example, the genera *Clostridium*, *Ruminococcus*, and *Caldicellulosiruptor* all contain gram-positive cellulolytic anaerobes. *Butyrivibrio*, *Acetivibrio*, and *Fibrobacter* are gram-negative fermentative anaerobes with the ability to degrade cellulose. Cellulolytic bacteria can also be aerobic gram-positive (*Cellulomonas* and *Thermobifida*), and aerobic gliding bacteria (*Cytophaga* and *Sporocytophaga*).

Celllluloytic fungi are found in the genera *Aspergillus*, *Cladosporium*, *Fusarium*, *Penicillium*, and *Trichoderma*. Note that lignocellulose degradation is not restricted to bacteria and fungi, since termites and crayfish can also produce cellulolytic enzymes (55).

There are various strategies through which microorganisms degrade lignocellulose. In general, anaerobic bacteria use complex cellulase systems, including the well-characterized, multi-enzyme complex, referred to as the cellulosome, produced by species such as *Clostridium thermocellum*, and *Acidothermus cellulolyticus* (60). Some anaerobic species do not have cellulosomes, but rather produce discrete cell surface-associated cellulases, which act on lignocellulose synergistically. Some anaerobic gastrointestinal bacteria, such as *Prevotella ruminicola*, use this strategy (57). Other cellulose degraders, including aerobic bacteria and fungi (e.g., thermophilic filamentous bacterium, *Thermobifida fusca*, aerobic filamentous fungus, *Trichoderma reesei*, and thermophilic fungus, *Humicola insolens*), secrete cell-free enzymes, which are present in culture supernatants. Furthermore, some aerobic microorganisms, such as *Thermomonospora curvata*, have cell surface-attached cellulolytic and hemicellulolytic enzyme complexes, similar to the cellulosome (9).

The multi-enzyme complex defining cellulosome represents an efficient way of degrading plant biomass, since many cellulolytic and hemicellulolytic enzymes are presumably organized with the correct ratios for optimal synergism. The sizes of

cellulosomes among different species vary from 600 kDa to 16 MDa. The genome of *C. thermocellum* has been sequenced and the genome analysis showed that there are 70 different dockerin containing-proteins (3). Most of these proteins are glycoside hydrolases with carbohydrate binding modules. The glycoside hydrolases included in the cellulosome from *C. thermocellum* belong to GH families 2, 5, 8, 9, 10, 11, 16, 18, 26, 28, 30, 53, 81, 39, 43, 44, 48, 54, and 74, which collectively represent cellobiohydrolase, endoglucanase, xylanase, lichenase, chitinase, mannanase, laminarinase, and carbohydrate esterase functionalities (3). Many of these GHs have carbohydrate binding modules to facilitate attachment to plant biomass. *C. thermocellum* also produces GHs which are not associated with the cellulosome complex and apparently contribute to plant biomass deconstruction as free enzymes (4).

Carbohydrate-binding modules (CBMs) are non-catalytic motifs appended to the carbohydrate active enzymes and grouped into at least 64 sequence-based families (27) [CAZy: http://www.cazy.org/]. CBMs improve the activity of plant cell wall degrading glycoside hydrolases (GHs) by increasing the concentration of the enzyme on the substrate surface, by providing prolonged association of the enzyme with its substrate, and by disrupting the carbohydrate structure (10, 19). CBMs are also divided into three groups, based on the topology of the binding sites. Type A CBMs bind to the flat surface of crystalline polysaccharides. Type B modules interact with single polysaccharide chains, present in certain xylan- and cellulose-based

substrates. Type C CBMs which bind to mono-, di-, and tri-saccharides do not contain the extended binding-site groove as type B modules do (10).

The genus *Caldicellulosiruptor* contains bacteria that can degrade complex carbohydrates, such as crystalline cellulose, hemicellulose, starch, pectin, and can also co-ferment hexose and pentose sugars to H_2 (7, 53). To date, the genomes of eight *Caldicellulosiruptor* species have been sequenced: *C. saccharolyticus* (51), *C. bescii* (33), *C. obsidiansis (23)*, *C. hydrothermalis*, *C. kristjanssonii*, *C. kronotskyensis*, *C. owensensis*, and *C. lactoaceticus* (8). These species degrade cellulose to different extents. For example, *C. hydrothermalis*, *C. kristjanssonii*, and *C. owensensis* are less cellulolytic due to lack of family 48 glycoside hydrolases (unpublished data). *Caldicellulosiruptor* species have different glycoside hydrolase and carbohydrate transporter inventories, which underlie the variability in biomass degradation capacity within the genus (8). Regardless of the differences, all *Caldicellulosiruptor* species can grow on model biomass substrates such as, switchgrass (13) and poplar (48). They have non-cellulosomal, multi-domain cellulolytic and hemicellulolytic enzymes, some of which contain surface layer homology (SLH) domains. The SLH domains anchor these enzymes to the cell wall, presumably to make the biomass degradation process more efficient (42). It is likely that the non-cellulosomal multi-domain SLH domain containing cellulolytic and hemicellulolytic enzymes from *Caldicellulosiruptor* species play an essential role in lignocellulose deconstuction. Six different groups of SLH-domain GHs can be

identified in the genomes of *Caldicellulosiruptor* species, each with different sets of conserved binding and catalytic domains. In four cases, homologous enzymes are conserved among other *Caldicellulosiruptor* species. We have previously reported the properties of Csac_0678 (GH5+CBM28+3xSLH) (42). Here, we examine the biochemical characteristics of the five other groups, represented by Calkr_2245, Calhy_0060, Calkro_0111, Calhy_1629, and Calhy_2383.

2. Materials and Methods

2.1. Bacterial strains and culture conditions

For cloning and expression, *E. coli* NovaBlue GigaSingles^{TM} competent cells (cloning host), and *E. coli* BL21(DE3)pLysS (expression host) were purchased from Novagen, Madison, WI. *E. coli* strains were cultivated in Luria-Bertani (LB) medium, supplemented with carbenicilin (Fisher Bioreagents) (50 µg/ml) and/or chloramphenicol (Sigma) (34 µg/ml). *C. hydrothermalis*, *C. kristjanssonii*, and *C. kronotskyensis* cultures were grown anaerobically at 70°C in modified DSM640 medium (DSMZ [http://www.dsmz.de]), containing NH_4Cl (0.9 g/l), NaCl (0.9 g/l), $MgCl_2.6H_2O$ (0.4 g/l), KH_2PO_4 (0.75 g/l), K_2HPO_4 (1.5 g/l), yeast extract (1 g/ml), trace element solutions SL-10 (1 ml/l), cellobiose (5 g/l), and $Na_2S.9H_20$ (0.05% w/v). Cells were visualized by epifluorescence microscopy (Nikon Intensilight C-HGFI, Lewisville, TX), as previously described (52).

2.2. Cloning, expression, and purification

All standard molecular cloning procedures were performed according to protocols in the Molecular Cloning Laboratory Manual (46). Genomic DNA (gDNA) of *C. hydrothermalis*, *C. kristjanssonii*, and *C. kronotskyensis* were extracted, as described previously (25). HotStartTaq DNA polymerase (Qiagen) was used to PCR amplify the genes of interest from the corresponding gDNA with the primers listed in Table 3.1. The primers were designed according to the proposed domain borders in NCBI (http://www.ncbi.nlm.nih.gov/) and EMBL-EBI (http://www.ebi.ac.uk/) public databases. pET-46 Ek/LIC, Novagen's ligation-independent cloning (LIC) vector, was used for cloning with an N-terminal 6xHis-tag. Gene constructions used to express intact proteins and truncation mutants are shown in Figure 3.1. The recombinant 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 by Eton Bioscience (RTP, NC). *E. coli* BL21(DE3)pLysS and Rosetta cells, containing pET-46 vector under the control of T7lac promoter (pET System Manual), were used for gene expression. Expression of Calkr_2245-TM1 was induced by the addition of 1 mM IPTG to the cells when OD_{600} reached 0.8. The cells were then harvested after 4 hours following induction by centrifugation at 10,000 *x g* for 10 minutes. Calhy_0060-TM1, Calkro_0111-TM1, Calkro_0111-TM2, and Calhy_1629 expression were induced when OD_{600} was approximately 1.0 and the cells were further grown at 18°C

for 18 hours. The cell pellet was re-suspended in 5 ml of 50 mM sodium phosphate pH 8.0, 100 mM NaCl, Nonidet P40 (0.1% (v/v), 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 minutes, with 10 sec off/on pulses. The suspension was heat-treated at 60^{o}C for 20 minutes in order to remove *E. coli* proteins. The cell extract was obtained after centrifuging the lysed and heat-treated cells at 20,000 *x g* for 20 minutes. All proteins studied found in the soluble fraction of the cell lysate, except Calkro_0111-TM1 which formed inclusion bodies. A protein refolding kit (Novagen, Madison, WI) was used to re-solubilize recombinant Calkro_0111-TM1. Inclusion bodies were collected and solubilized under mildly denaturing conditions using N-lauroylsarcosine (Novagen, Madison, WI). Then, the solubilized fraction was refolded by dialysis against a neutral pH buffer containing a reducing agent. HiTrap HP (GE Life Sciences, Piscataway, NJ), and a Resource Q (GE Life Sciences, Piscataway, NJ) column were used to purify the recombinant proteins on a Biologic DuoFlow FPLC (BioRad, Hercules, CA). Purity of the recombinant proteins was evaluated by SDS-PAGE.

2.3. Biochemical characterization of SLH-domain enzymes

The DNS reducing sugar assay, which was adapted to utilize a miniaturized 96-well microplate format (58, 59), was used to detect both five- and six-carbon reducing sugars. DNS reagent was prepared, as previously described (39). The optimal pH was determined at 75°C in buffers at pH 2.5 to 10 (50 mM citrate buffer,

pH 2.5-4.5, sodium acetate buffer, pH 5-6, and 50 mM sodium phosphate buffer, pH 7-8, sodium bicarbonate buffer, pH 9-10). The optimal temperature for activity was determined by 3,5-dinitrosalicylic acid (DNS) assay between 20°C and 90°C at the optimal pH. The substrate used for Calkr_2245-TM1 was 1% (w/v) oat spelt xylan, and for Calhy_0060 and Calkro_0111-TM2 1% (w/v) laminarin was used.

Calhy_0060-TM1 was incubated with 1% (w/v) solutions of lichenan (Sigma), laminarin (Sigma), and barley glucan (Megazyme) in 50 mM sodium acetate buffer, pH 6 at 70°C on a thermomixer (Eppendorf) for 5 hours. Calkro_0111-TM2 and Calkr_2245 were incubated with 1% (w/v) solutions of laminarin and xylan, respectively. The reaction was stopped by boiling the mixture for 5 minutes. Each reaction mixture was centrifuged at 10,000*xg* for 5 minutes. The supernatant was applied to thin layer chromatography plates, silica gel 60 F254 (Fisher). The plates were developed in butanol:ethanol:water (5:3:2) mobile phase and reducing sugars were resolved in 1% (w/v) orcinol monohydrate in 10% H_2SO_4 dissolved in ethanol. Glucose (Sigma) and xylooligosaccharides (Sigma) were used as the standards.

2.4. Localization and imaging of Csac_0678 on the cell surface

Localization and imaging experiments were kindly performed by Jennifer Morrell-Falvey and Sarah J. Melton from Oak Ridge National Laboratory (ORNL) (Oak Ridge, TN). Early stationary phase *C. saccharolyticus* cells grown on cellobiose (5 g/L) were sent to ORNL. Primary antibodies generated against the recombinant Csac_0678 (Genetel, LLC, Madison, WI) were used for imaging

analysis. The culture samples (1 ml concentrated down to 100 µl) were incubated with the primary antibody at room temperature for 2 hours with gentle rocking. Then, cells were washed with PBS buffer three times for 3 minutes. Secondary antibody (Alexa Fluor 488 goat anti-rabbit antibody) was added to the sample and incubated at room temperature for 2 hours with gentle rocking. Cells were washed again with PBS buffer three times for 3 minutes. Cells were counter-stained with SYTO61 and then mounted on the glass slides. Cells were imaged with structured illumination microscopy (SIM) (Carl Zeiss MicroImaging Inc., Thornwood, NY)

2.5. Enzyme synergism on complex carbohydrates

5 µM Calkr_2245 and 5 µM Calhy_0678 were incubated with 1% birchwood xylan (Sigma) at 70°C, both separately and in combination. The same enyme molarities were used for the individual reactions and in combination. The reactions were shaken at 500 RPM on a thermoshaker (Eppendorf), to suspend xylan since it is sparingly soluble. After 16 hours, the reactions were stopped by the addition of DNS assay reagent and incubation at 95°C for 5 minutes. DNS reducing sugar assay was performed, as described above. For the synergism experiments, the same total enzyme concentrations were used as for the individual enzymes. The model biomass substrates, switchgrass and poplar, were used for the synergism experiments. Dilute acid pretreated switchgrass was washed with pico-pure water three times to remove soluble sugars that might interfere with the DNS assay. Poplar which was *P. trichocarpa* crossed with *P. deltoids* was also washed with 70°C water

(1 g poplar in 50 ml water) and placed in the 70°C shaking bath at 100 RPM for 16 hours. Then it was washed again with 70°C water (1 g poplar in 100 ml water) and dried at 50°C for 18 hours. Biomass substrates (2.5% (w/v)) mixed 50 mM sodium acetate, pH 6 were incubated with each protein and with protein combinations. Control reactions were included with no added protein. Samples were incubated for either 20 or 48 hours. A DNS assay was performed to quantify any reducing sugars created by the enzymes.

3. Results and Discussion

Table 3.2 lists the SLH-domain GHs identified in the genome sequences of *Caldicellulosiruptor* species and also shows the organization of domains associated with each enzyme. Each protein represents one of six orthologous groups. For example, the set of conserved domains that Csac_0678 represents, Calkro_2036, Calla_0352, Calhy_2064, Calkr_2007, Athe_0594, and COB47_0546. Calkr_2245 and Calhy_0060, are the second most common group. Orthologs of Casc_2245 are found in *C. owensensis* (Calow_1924), *C. kronotskyensis* (Calkro_0402), and *C. lactoaceticus* (Calla_0206), while orthologs of Calhy_0060 are noted in Calkro_0072, COB47_0076, and Csac_2548-2549. Csac_2549 contains the N-terminal domains (SLH-SLH-CBM54), and Csac_2548 contains the remaining domains (GH16-CBM4-CBM4-CBM6-CBM4-CBM4-CBM4). Csac_2549 and Csac_2548 were likely formed by a gene disruption event. A gene duplication event could be responsible for the repetition of Calkro_0111 in the same genome as

Calkro_0121, the latter of which contains exactly the same domains except for the final CBM32 domain. Calkro_0111 and Calkro_0121 are the only members of this group, in the sequenced *Caldicellulosiruptor* genomes. Calhy_1629 and Calhy_2383 do not have any orthologs within the genus. Here, we investigated each orthologous group of enzymes to examine their respective contributions to plant biomass deconstruction within the genus *Caldicellulosiruptor*.

3.1. Calkr_2245

Calkr_2245 is homologous to xylanases in other *Caldicellulosiruptor* species, specifically Calow_1924 (72% amino acid sequence identity over the entire protein), Calkro_0402 (72%), and Calla_0206 (65%), although they all lack the CE15 domain present in Calkr_2245. In fact, Calkr_2245 is the only GH containing a carbohydrate esterase domain in the genome sequenced *Caldicellulosiruptor* spp. (27). Calkr_2245 is related (44% identity) to a xylanase, XynA, from *Thermotoga maritima* MSB8 (35, 56), consisting of two N-terminal CBM22s, followed by GH10, and the two C-terminal CBM9 domains (54). Deletion of the non-catalytic domains increased the activity and stability of XynA (35). This is surprising since the typical role of CBM is to promote a close interaction between the enzyme and the polysaccharide and, thereby, improve polysaccharide degradation (5, 11, 40). The presence of CBM22 in XynA improved activity in the acidic pH range. CBM9 in the *T. maritima* enzyme bound to xylan, β-1,4-glucan, and mixed linked β-1,3-1,4-glucans. CBM22 in XynA

showed highest affinity to insoluble oat spelt xylan and birchwood xylan and low affinity to hydroxyethyl cellulose and carboxymethyl cellulose (35).

Note the *T. maritima* XynA lacks a CE domain while it is the most intriguing feature of Calkr_2245. There is a carbohydrate esterase family 15 (CE15) between the C-terminal CBM9 domain and the two C-terminal SLH domains of Calkr_2245. CAZy classifies CE15 as a 4-*O*-methyl-glucuronoyl methylesterase. In the plant cell wall, lignin and hemicellulose are connected by ester linkages between 4-*O*-methyl-D-glucuronic acid of glucuronoxylan and lignin alcohols (27, 36). Xylan, the predominant hemicellulose component, consists of a β-1,4-linked backbone of xylose units, with single residue side chains of arabinose, glucuronic and *O*-methyl-D-glucuronic acid (38). CE15 cleaves the linkage between xylan and lignin, facilitating the lignin removal from lignocellulose in the deconstruction process. A family 15 CE, Cip2, from *Trichoderma reesei* exhibited glucuronoyl esterase activity (36). The structure of the catalytic domain, CE15, determined at a resolution of 1.9 Å, showed that the active site was on the surface of the protein, indicating that the enzyme can hydrolyze ester bonds of large substrates (45). Another example of an organism encoding carbohydrate esterases is the white-rot fungus, *Phanerochaete chrysosporium*, which degrades lignocellulosic biomass (1). It was also found to produce a family 15 carbohydrate esterase when grown on plant cell walls (27, 30).

Here, three deletion mutants of Calkr_2245 were produced in *E. coli* to elucidate the role of CBMs with respect to the activity and stability of the enzyme,

and the influence of the CE15 domain in degrading biomass substrates. Three TMs were examined: Calkr_2245-TM1 (CBM9+GH10+CBM22), Calkr_2245-TM2 (CBM9+CBM+9+GH10+CBM22+CBM22), and Calkr_2245-TM3 (CBM9+CBM9+CBM+9+GH10+CBM22+CBM22+CBM22+CE15). Calkr_2245-TM1 (79 kDa) exhibited activity towards xylan, with an optimum temperature and pH of 65-70°C and pH 7, respectively. Calkr_2245-TM2 and Calkr_2245-TM3 both showed approximately ten times less activity than Calkr_2245-TM1 on xylan, consistent with previous reports on *T. maritima* XynA (35). However, this result also might be related to the misfolding of the recombinant proteins due to their larger sizes. More optimization has to be done in expression and cloning stages to be able to claim that the deletion mutant exhibits more activity than the mutants with more carbohydrate domains. Calkr_2245-TM2 and Calkr_2245-TM3 deletion mutants also failed to bind to the Ni-column, indicating that the histidine-tags may be hidden in the interior sides of the proteins.

Recombinant Calkr_2245-TM1 (CBM9+GH10+CBM22) exhibited xylanase activity at pH 7 and at 70°C, optimally. It should be noted that the optimum growth temperature of *C. kristjanssonii* is 78°C (14). The reason for the decrease in the optimum temperature of the recombinant enzyme compared to the growth temperature of the host organism could relate to the reduced number of carbohydrate binding motifs. The CBMs could contribute to thermostabilty. *C. kristjanssonii* can grow on switchgrass and xylan but the growth on crystalline

cellulose and filter paper is limited (unpublished data), suggesting that this bacterium uses the hemicellulose component of plant biomass. Calkr_2245 with its CE domain likely removes the linkages between xylan and lignin and anchors to the biomass substrate through SLH domains.

3.2. Calhy_0060

Calhy_0060 is a GH16 glucanase comprised of 1732 amino acids, and has close homologs in other *Caldicellulosiruptor* species: COB47_0076 (95% amino acid sequence identity), Calkro_0072 (91%), and Csac_2548 (76%) (http://www.ncbi.nlm.nih.gov/). Family 16 glycoside hydrolases (GH16) are typically endo-1,3(4)-β-glucanases (18) which cleave β-1,3 or β-1,4 glycosidic bonds when the glucose residue whose reducing group is involved in the linkage to be hydrolyzed is at C-3 (31, 44). 1,3-1,4-β-glucans are linear polysaccharides containing up to 1200 β-D-glucosyl residues linked by β-1,3 (25-30%) and β-1,4 (75-70%) glycosidic bonds. They are found in the cell walls of higher plant families, such as *Poaceae* (grasses) (17, 44). The β-1,3-4-glucans are absent in dicots (47). Endo-1,3(4)-β-glucanases can hydrolyze laminarin, lichenan, and cereal D-glucans (31). For example, *Cm*Unk16A, from *Cellvibrio mixtus*, which contains a GH16 catalytic domain, two CBM 32 domains, one CBM4 and a domain of unknown function, was shown to hydrolyze β-1,3 glucans (19).

Other GH16 enzymes from thermophilic microorganisms have been studied. An extracellular GH16 glycoside hydrolase produced extracellularly by a

thermophilic fungus, *Paecilomyces* sp. FLH30, PsBg16A, optimally active at 70°C and at pH 7, degraded barley glucan (100%), lichenan (68.7%), and laminarin (22.3%), but not CMC-Na, Avicel, or birchwood xylan (31). A Calhy_0060 homolog (36% amino acid identity), LicA (Cthe_2809) from *Clostridium thermocellum* was characterized (22, 24). LicA is not cellulosomal, as opposed to the other family GH16, LicB, produced by *C. thermocellum* (3). LicA contains three SLH domains, one GH54, one GH16, and four CBM4s. However, unlike Calhy_0060, it does not have a CBM6 domain. CBM6 modules were found to be linked to enzymes showing different substrate specificities, such as cellulose, xylan, mixed β-(1,3)(1,4)-glucan and β-1,3-glucan (43). Interestingly, family 6 CBMs contain two distinct clefts, cleft A, found in the loop region connecting the inner and outer β-sheets of the jelly roll fold and, cleft B, located on the concave surface of one β-sheet. A previous study by Henshaw et al. (28) investigated the role of CBM6 on the C-terminal side of a GH 5 from *Cellvibrio mixtus*. The CBM 6 bound to the mixed β-(1,3)(1,4)-glucans, lichenan and barley glucan, as well as to cello-oligosaccharides, insoluble forms of cellulose, laminarin, and xylooligosaccharides. Carefully performed mutagenesis studies showed that the mixed β-(1,3)(1,4)-glucans interacted with only cleft B, laminarin bound preferentially to cleft A, xylooligosaccharides showed absolute specificity to cleft A, and cello-oligossacharides bound both cleft A and B (28). Other studies also showed that family 6 CBM is a very unusual binding domain containing two binding sites which exhibits different ligand specificities (20, 28, 43). Dvortsov et al. (22)

showed that CBM54 had a significant effect on the binding to lichenan and laminarin. CBM54 from LicA also bound to the insoluble polysaccharides, xylan, pustulan, chitin, chitosan, yeast cell wall glucan, and Avicel. Deletion mutants lacking the CBM54 had lower optimum temperatures and thermostability. LicA hydrolyzed β-1,3- and mixed β-1,3-1,4-linkages and exhibited the highest activity on lichenan. Lam16A, accounting for *Thermotoga maritima*'s laminarinase activity, is composed of one C-terminal CBM4, one GH16, and one N-terminal CBM4 (12). CBM4 domains were reported to be highly thermostable binding β-1,3- and β-1,3-1,4-glucans.

Calhy_0060-TM1, lacking SLH domains, exhibited highest activity on laminarin at 70°C and at pH 6. Calhy_0060-TM1 degraded laminarin, lichenan, and barley glucan, by an endo-acting mode, cleaving the β-1,3-glycosididc linkages randomly and releasing different sizes of oligosaccharides (see Figure 3.2). Calhy_0060, and its orthologs in *C. kronotskyensis*, *C. obsidiansis*, and *C. saccharolyticus*, contibute to the degradation of the hemicellulosic components of the grasses falling into neutral pH hot springs in terrestrial environments.

3.3. Calkro_0111

BLAST results to date indicate that Calkro_0111 is a unique enzyme in the *Caldicellulosiruptor* genus and beyond, containing one GH16, one GH55, one CBM54, two CBM32, and fibronectin domains. Calkro_0121 has the same conserved domains except the last CBM32 domain at the C-terminus. Calkro_0111 and Calkro_0121 likely formed by a gene duplication event. Calkro_0113 is also

homologous to Calkro_0111, containing one GH55 and three CBM32, occurring in the same gene neighborhood. However, there is no other gene with the same set of conserved domains annotated in the public databases. The region in the *C. kronotskyensis* genome containing Calkro_0111 to Calkro_0122, has genes encoding hemicellulases (Table 3.3). Calkro_0114 belongs to GH81, which are β-1,3-glucanases, Calkro_0118 and Calkro_0119 are ABC carbohydrate transporters, and Calkro_0120 belongs to GH2, which contains β-galactosidases and mannosidases. Calkro_0122 encodes for a GH88, which is annotated as a β-glucuronyl hydrolase. Calkro_0111, Calkro_0121, and Calkro_0113 are the only genes containing GH55 in the *C. kronotskyensis* 2002 genome. Other than Calkro_0111 and Calkro_0121 containing GH16 domains, and Calkro_0072 is the only GH16 also containing one CBM54, one CBM6, and five CBM4s.

Calkro_0111 is comprised of 2435 amino acids, with a predicted M_r of 264 kDa. Two truncation mutants (TMs) were produced: Calkro_0111-TM1 (95 kDa) was constructed by truncating Calkro_0111 gene at the 3306th base pair as this type of truncation separates the two catalytic domains. Calkro_0111-TM2 (144 kDa) contained the rest of the gene up to the C-terminal (see Figure 3.1). N-terminal S-layer homology motifs were not included in the Calkro_0111-TM1 construct.

Family 55 glycoside hydrolases (GH55) are described as exo/endo-β-1,3-glucanases on the CAZy website (27), and characterized as laminarin degraders (31). Although β-1,3-glucanases and β-1,3-1,4-glucanases have a similar protein

fold, β-1,3-glucanases hydrolyze only β-1,3-linkages, whereas β-1,3-1,4-glucanases cleave the β-1,4-linkage next to a β-1,3-linkage (24). β-1,3-glucanases are used in yeast extract production and in the conversion of algal biomass to fermentable sugars (24). In a previous study (34), Lam55, a GH55 from *Phanerochaete chrysosporium*, was cloned and overexpressed in *Pichia pastoris*. It was found to be active on laminarin (β-1,3/1,6-glucan) and on curdlan (β-1,3-glucan). No activity was observed on pustulan (β-1,6-glucan) and on lichenan (β-1,3-1,4-glucan).

Calkro_0111-TM1 contains the N-terminal region (CBM54+FN3+GH16+FLD+FN3) of Calkro_0111. Calkro_0111-TM1 was expressed and found mostly in the insoluble fraction of the cell lysate, i.e., the recombinant protein was misfolded forming inclusion bodies. It was re-solubilized under mildly denaturing conditions. The resolublized Calkro_0111-TM1, optimally active at pH7 and 68C, exhibited activity against laminarin, lichenan, and barley glucan, similar to the GH16-containing Calhy_0060. Calkro_0111-TM2 (FN3+FLD+GH55+CBM32+CBM32) was expressed and found in the soluble fraction of the cell lysate and purified using affinity chromatography. TLC image shows (see Figure 3.3) that it degraded laminarin into oligosaccharides. No activity was observed on lichenan, degradation of which requires β-1,3-1,4-glucanase activity to cleave the β-1,4-glycosidic linkage adjacent to a 3-*O*-substituted glucose residue (31). Optimum temperature and pH were found to be 75°C and pH 5 for Calkro_0111-TM2, respectively.

3.4. Calhy_1629

Calhy_1629 is a novel enzyme among the genome sequenced *Caldicellulosiruptor* species; it is comprised of one GH43, one CBM54, and three SLH domains. Glycoside hydrolase family 43 enzymes exhibit β-xylosidase activity with an inverting mechanism, cleaving xylose residues from the non-reducing end of oligomeric xylosides (49). β-xylosidases are required for complete degradation of xylan to xylose (29). Therefore, Calhy_1629 likely plays an important role in the degradation of hemicelluloses to make cellulose accessible for enzymatic hydrolysis (49). For consolidated bioprocessing, Calhy_1629 could function in the final step in the breakdown of plant cell wall hemicellulose, producing pentose sugars for fermentation to ethanol directly (15, 50). Calhy_1629 was produced recombinantly without the SLH domains and found in the soluble fraction. It was active on xyloglucan, arabinoxylan, and on xylan, with the highest activity at pH 5 and at 75°C.

In general, β-xylosidases can contribute to the post-treatment of a xylanase-generated hydrolysates to produce xylose (49). Smaali et al. studied two β-xylosidases, XylBH43, a glycoside hydrolase family 43, and XylBH39, GH39, from *Bacillus halodurans* C-125. Together, these enzymes released 521 μg/ml of xylose from endoxylanase-generated wheat bran hydrolysates in 330 min (49). A previous study showed that a thermophilic Gram-positive soil bacterium, *Geobacillus stearothermophilus* T-6, also produced a GH43 enzyme, XynB3, intracellularly. Xyloligosaccharides produced by an extracellular endo-1,4-β-xylanase were

transported into the *G. stearothermophilus* cytoplasm via ABC transporters, where they were degraded to xylose by XynB3. The crystal structure of XynB3 revealed an N-terminal, five-bladed β-propeller catalytic domain, and a β-sandwich domain. XynB3 was shown to be an exoenzyme producing single xylose units (15). Das et al. used a glycoside hydrolase family 43 from *Clostridium thermocellum* for bioethanol production through simultaneous saccharification and fermentation (SFF) from pretreated biomass of mango, poplar, neem, and asoka (21). The enzyme was recombinantly produced and used to degrade phosphoric acid-acetone or ammonia fibre expansion (AFEX) pretreated biomass. The GH43 was active on rye arabinoxylan and oat-spelt xylan. The pentose sugars produced by GH43 from the biomass converted to bioethanol by *Candida shehatae*.

3.5. Calhy_2383

There are only four family 87 glycoside hydrolyses encoded by the genus *Caldicellulosiruptor*; three of which are in *C. hydrothermalis* (Calhy_0074, Calhy_2380, Calhy_2383), and one in *C. kronotskyensis* (Calkro_0091). Calhy_2383 is the only one containing SLH domains.

α-1,3-glucanases (mutanases) have been used for dental plaque removal. *Streptococcus mutans* produces mutan, an adhesive and insoluble α-1,3-linked glucan, from sucrose thereby forming dental plaque on the tooth surface (26). Another reported α-1,3-glucanase is from *Trichoderma harzianum*, which was found to secrete this enzyme when grown on polysaccharides, fungal cell walls, and

autoclaved mycelium. The enzyme showed lytic and antifungal activity against fungal plant pathogens (2). Calhy_2383 and other family 87 glycoside hydrolases from *C. hydrothermalis* and *C. kronotskyensis* are thought to play roles in the degradation of fungal component of lichenan. However, it is not clear as to what role they would play in the processing of biomass for biofuel production, as plant cell walls do not appear to contain α-1,3-glucans.

3.6. Structure and localization of Csac_0678, a bi-functional endoglucanase/xylanase

Csac_0678 is composed of one GH5 domain, one family 28 carbohydrate binding module (CBM28), and three SLH domains (see Table 3.2). Orthologs of Csac_0678 are found in all *Caldicellulosiruptor* spp., indicating that it might be critically important for utilization of the carbohydrate component of plant cell walls. Previous work with Csac_0678 and its deletion mutants showed that it exhibited both xylanase and endoglucase activities, and degraded barley glucan, xylan, Avicel, switchgrass, and filter paper (42). Deletion of the SLH domains did not affect the activity, but removal of CBM28 rendered Csac_0678 inactive on crystalline cellulose. Western blot analysis of the S-layer extracted from *C. saccharolyticus* cells proved that Csac_0678 was associated with the S-layer.

The crystal structure of the GH5 catalytic domain of Csac_0678 was solved to 2.0 Å resolution. Csac_0678 GH5 has a typical TIM-barrel protein fold (Figure 3.4), consisting of eight alpha helices and eight beta sheets. This protein fold ubiquitously

occurs in nature. Most of the proteins containing a TIM-barrel protein fold are enzymes belonging to five to six different enzyme classes (41).

To further investigate the localization of Csac_0678 on the S-layer, polyclonal antibodies directed against Csac_0678 were used to view the enzyme through Structured Illumination Microscopy (Alexa Fluor 488 labeled secondary antibodies were used) (see Figure 3.5). *C. saccharolyticus* was stained with SYTO61 (red). The SIM imaging supports the notion that he enzyme is cell-surface associated, presumably with the S-layer.

Csac_0678 and its orthologs have bi-functional activity towards cellulose and xylan. *Caldicellulosiruptor* species secrete Csac_0678, where it anchors to the S-layer fraction. Csac_0678 molecules attach to the biomass substrates by means of their family 28 CBM while the TIM-barrel protein fold catalytic domain degrades the (hemi)cellulosic substrates.

4. Summary

This study considered cell surface attached, multi-domain cellulolytic and hemicellulolytic enzymes from the genus *Caldicellulosiruptor* (Table 3.4). Although this genus does not produce cellulosomes, it uses a similar strategy to address the recalcitrant nature of lignocellulosic biomass. *Caldicellulosiruptor* species produces a set of SLH-containing glycoside hydrolases which relate to the dockerin-associated glycoside hydrolases from cellulosome-producing organisms (3). The difference is that the SLH containing glycoside hydrolases are not attached to a scaffoldin, as in

the case with the cellulosome, and function as individual enzymes while still attached to the cell wall and to the polysaccharide. In this regard, there is much to learn from *Caldicellulosiruptor* species to understand alternative strategies for biomass degradation.

Figures 3.6 and 3.7 summarize how carbohydrate active (CAZy) enzymes from *Caldicellulosiruptor* species function to degrade and process carbohydrates. Shown here are the cases of *C. hydrothermalis* (Figure 3.6) and *C. saccharolyticus* (Figure 3.7). Some of these GHs are intracellular, degrading smaller oligosaccharides transported into the cell; for example, Csac_1080, a mannanase, and Csac_1561, a α-N-arabinofuranosidase, are such examples of glycoside hydrolases with no signal peptides. The GHs with SLH domains are shown as attached to the cell wall. After they are produced in the cell, they are transported out of the cell and anchor to the S-layer. Their localization on the cell likely contributes to the overall efficiency in degrading biomass. Free, extracellular enzymes, such as CelA (Csac_1076) and CelB (Csac_1078), can degrade crystalline cellulose, to allow for the transport of oligoglucans into the cell prior to degradation to fermentable sugars (6).

It is interesting that all of the SLH containing enzymes (Table 3.4) from *Caldicellulosiruptor* species are hemicellulolytic, with an exception of Csac_0678 which is both cellulolytic and hemicellulolytic. These enzymes target the degradation of hemicellulose, which constitutes 20-30% of the plant cell wall. Apart from fermenting pentose sugars produced by hemicellulolytic degradation for their

metabolism, presumably, they uncover the hemicellulose component to make cellulose accessible for cellulase attack, as was proposed to be part of an “onion peeling” mechanism (Kataeva, I. et al., unpublished).

Degradation of plant biomass requires the synergistic action of many GHs, either directed by the cell or in formulations of individual enzymes used for biotechnological applications. A previous work showed that there was synergism between cellulases from different species: *T. fusca* and *T. reesei* (32). We proved that the combination of Calkr_2245 (xylanase) and Calhy_1629 (β-xylosidase) had an increased activity compared to the sum of activities of the individual enzymes, indicating that there is some kind of synergism between the two enzymes. While no evidence of high levels of synergism were noted here for the Caldicellulosiruptor species SLH-domain GHs, the lack of other enzymes, not associated with the S-layer, in the cocktails created likely explains this result (unpublished results, Ozdemir and Kelly). Further work is needed to develop an overall explanation of the 50 or so GHs in the Caldicellulosiruptor species work together to produce fermentable sugars from plant cell walls. This kind of insight will be very useful in developing synthetic enzyme mixtures for biofuels applications.

REFERENCES

1. Adav, S. S., A. Ravindran, and S. K. Sze. Quantitative proteomic analysis of lignocellulolytic enzymes by *Phanerochaete chrysosporium* on different lignocellulosic biomass. J Proteomics 75:1493-504.
2. Ait-Lahsen, H., A. Soler, M. Rey, J. de La Cruz, E. Monte, and A. Llobell. 2001. An antifungal exo-alpha-1,3-glucanase (AGN13.1) from the biocontrol fungus *Trichoderma harzianum*. Appl Environ Microbiol 67:5833-9.
3. Bayer, E. A., Shoham, Y., Lamed, R. 2008. Cellulosome-enhanced conversion of biomass: on the road to bioethanol, p. 75-96. *In* J. D. Wall, Harwood, C.S., Demain, A. (ed.), Bioenergy. ASM, Washington.
4. Berger, E., D. Zhang, V. V. Zverlov, and W. H. Schwarz. 2007. Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. FEMS Microbiol Lett 268:194-201.
5. Black, G. W., J. E. Rixon, J. H. Clarke, G. P. Hazlewood, L. M. Ferreira, D. N. Bolam, and H. J. Gilbert. 1997. Cellulose binding domains and linker sequences potentiate the activity of hemicellulases against complex substrates. J Biotechnol 57:59-69.

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

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

8. Blumer-Schuette, S. E., I. Ozdemir, D. Mistry, S. Lucas, A. Lapidus, J. F. Cheng, L. A. Goodwin, S. Pitluck, M. L. Land, L. J. Hauser, T. Woyke, N. Mikhailova, A. Pati, N. C. Kyrpides, N. Ivanova, J. C. Detter, K. Walston-Davenport, S. Han, M. W. Adams, and R. M. Kelly. 2011. Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus*. J Bacteriol 193:1483-4.

9. Bond, K., F. Stutzenberger. 1989. A note on the localization of cellulosome formation in *Thermomonospora curvata*. J. Appl. Bacteriol. 67:605-609.

10. Boraston, A. B., D. N. Bolam, H. J. Gilbert, and G. J. Davies. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382:769-781.

11. Boraston, A. B., E. Kwan, P. Chiu, R. A. Warren, and D. G. Kilburn. 2003. Recognition and hydrolysis of noncrystalline cellulose. J Biol Chem 278:6120-7.

12. Boraston, A. B., R. A. Warren, and D. G. Kilburn. 2001. beta-1,3-Glucan binding by a thermostable carbohydrate-binding module from *Thermotoga maritima*. Biochemistry 40:14679-85.

13. Bouton, J. H. 2007. Molecular breeding of switchgrass for use as a biofuel crop. Curr Opin Genet Dev 17:553-8.

14. Bredholt, S., J. Sonne-Hansen, P. Nielsen, I. M. Mathrani, and B. K. Ahring. 1999. *Caldicellulosiruptor kristjanssonii* sp. nov., a cellulolytic, extremely thermophilic, anaerobic bacterium. Int J Syst Bacteriol 49 Pt 3:991-6.

15. Brux, C., A. Ben-David, D. Shallom-Shezifi, M. Leon, K. Niefind, G. Shoham, Y. Shoham, and D. Schomburg. 2006. The structure of an inverting GH43 beta-xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. J Mol Biol 359:97-109.

16. Bugg, T. D., M. Ahmad, E. M. Hardiman, and R. Rahmanpour. Pathways for degradation of lignin in bacteria and fungi. Nat Prod Rep 28:1883-96.

17. Burton, R. A., M. J. Gidley, and G. B. Fincher. Heterogeneity in the chemistry, structure and function of plant cell walls. Nat Chem Biol 6:724-32.

18. Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37:D233-238.

19. Centeno, M. S., A. Goyal, J. A. Prates, L. M. Ferreira, H. J. Gilbert, and C. M. Fontes. 2006. Novel modular enzymes encoded by a cellulase gene cluster in *Cellvibrio mixtus*. FEMS Microbiol Lett 265:26-34.

20. Czjzek, M., D. N. Bolam, A. Mosbah, J. Allouch, C. M. Fontes, L. M. Ferreira, O. Bornet, V. Zamboni, H. Darbon, N. L. Smith, G. W. Black, B. Henrissat, and H. J. Gilbert. 2001. The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. J Biol Chem 276:48580-7.

21. Das, S. P., R. Ravindran, S. Ahmed, D. Das, D. Goyal, C. M. Fontes, and A. Goyal. 2012. Bioethanol Production Involving Recombinant *C. thermocellum* Hydrolytic Hemicellulase and Fermentative Microbes. Appl Biochem Biotechnol (Epub ahead of print).

22. Dvortsov, I. A., N. A. Lunina, L. A. Chekanovskaya, W. H. Schwarz, V. V. Zverlov, and G. A. Velikodvorskaya. 2009. Carbohydrate-binding properties of a separately folding protein module from beta-1,3-glucanase Lic16A of *Clostridium thermocellum*. Microbiology 155:2442-9.

23. Elkins, J. G., A. Lochner, S. D. Hamilton-Brehm, K. W. Davenport, M. Podar, S. D. Brown, M. L. Land, L. J. Hauser, D. M. Klingeman, B. Raman, L. A.

Goodwin, R. Tapia, L. J. Meincke, J. C. Detter, D. C. Bruce, C. S. Han, A. V. Palumbo, R. W. Cottingham, M. Keller, and D. E. Graham. 2010. Complete genome sequence of the cellulolytic thermophile *Caldicellulosiruptor obsidiansis* OB47T. J Bacteriol 192:6099-100.

24. Fuchs, K. P., V. V. Zverlov, G. A. Velikodvorskaya, F. Lottspeich, and W. H. Schwarz. 2003. Lic16A of *Clostridium thermocellum*, a non-cellulosomal, highly complex endo-beta-1,3-glucanase bound to the outer cell surface. Microbiology 149:1021-31.

25. Geslin, C., M. Le Romancer, G. Erauso, M. Gaillard, G. Perrot, and D. Prieur. 2003. PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "*Pyrococcus abyssi*". J. Bacteriol. 185:3888-3894.

26. Hakamada, Y., N. Sumitomo, A. Ogawa, T. Kawano, K. Saeki, K. Ozaki, S. Ito, and T. Kobayashi. 2008. Nucleotide and deduced amino acid sequences of mutanase-like genes from *Paenibacillus* isolates: proposal of a new family of glycoside hydrolases. Biochimie 90:525-33.

27. Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon, and G. Davies. 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.

28. Henshaw, J. L., D. N. Bolam, V. M. Pires, M. Czjzek, B. Henrissat, L. M. Ferreira, C. M. Fontes, and H. J. Gilbert. 2004. The family 6 carbohydrate

binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. J Biol Chem 279:21552-9.

29. Herrmann, M. C., M. Vrsanska, M. Jurickova, J. Hirsch, P. Biely, and C. P. Kubicek. 1997. The beta-D-xylosidase of *Trichoderma reesei* is a multifunctional beta-D-xylan xylohydrolase. Biochem J 321 (Pt 2):375-81.

30. Hori, C., K. Igarashi, A. Katayama, and M. Samejima. 2011. Effects of xylan and starch on secretome of the basidiomycete *Phanerochaete chrysosporium* grown on cellulose. FEMS Microbiol Lett 321:14-23.

31. Hua, C., H. Yi, and L. Jiao. 2011. Cloning and expression of the endo-1,3(4)-beta-glucanase gene from *Paecilomyces* sp. FLH30 and characterization of the recombinant enzyme. Biosci Biotechnol Biochem 75:1807-12.

32. Irwin, D. C., M. Spezio, L. P. Walker, and D. B. Wilson. 1993. Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. Biotechnol Bioeng 42:1002-13.

33. Kataeva, I. A., S. J. Yang, P. Dam, F. L. Poole, 2nd, Y. Yin, F. Zhou, W. C. Chou, Y. Xu, L. Goodwin, D. R. Sims, J. C. Detter, L. J. Hauser, J. Westpheling, and M. W. Adams. 2009. Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. J Bacteriol 191:3760-1.

34. Kawai, R., K. Igarashi, and M. Samejima. 2006. Gene cloning and heterologous expression of glycoside hydrolase family 55 beta-1,3-glucanase

from the basidiomycete *Phanerochaete chrysosporium*. Biotechnol Lett 28:365-71.

35. Kleine, J., and W. Liebl. 2006. Comparative characterization of deletion derivatives of the modular xylanase XynA of *Thermotoga maritima*. Extremophiles 10:373-81.

36. Li, X. L., S. Spanikova, R. P. de Vries, and P. Biely. 2007. Identification of genes encoding microbial glucuronoyl esterases. FEBS Lett 581:4029-35.

37. Lynd, L. R., P. J. Weimer, W. H. van Zyl, and I. S. Pretorius. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506-77.

38. Meissner, K., D. Wassenberg, and W. Liebl. 2000. The thermostabilizing domain of the modular xylanase XynA of *Thermotoga maritima* represents a novel type of binding domain with affinity for soluble xylan and mixed-linkage beta-1,3/beta-1, 4-glucan. Mol Microbiol 36:898-912.

39. Miller, G. L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. 31:426-428.

40. Millward-Sadler, S. J., D. M. Poole, B. Henrissat, G. P. Hazlewood, J. H. Clarke, and H. J. Gilbert. 1994. Evidence for a general role for high-affinity non-catalytic cellulose binding domains in microbial plant cell wall hydrolases. Mol Microbiol 11:375-82.

41. Nagano, N., C. A. Orengo, and J. M. Thornton. 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol 321:741-65.

42. Ozdemir, I., S. E. Blumer-Schuette, and R. M. Kelly. 2012. S-layer homology domain proteins Csac_0678 and Csac_2722 are implicated in plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Appl Environ Microbiol 78:768-77.

43. Pires, V. M., J. L. Henshaw, J. A. Prates, D. N. Bolam, L. M. Ferreira, C. M. Fontes, B. Henrissat, A. Planas, H. J. Gilbert, and M. Czjzek. 2004. The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. J Biol Chem 279:21560-8.

44. Planas, A. 2000. Bacterial 1,3-1,4-beta-glucanases: structure, function and protein engineering. Biochim Biophys Acta 1543:361-382.

45. Pokkuluri, P. R., N. E. Duke, S. J. Wood, M. A. Cotta, X. L. Li, P. Biely, and M. Schiffer. 2011. Structure of the catalytic domain of glucuronoyl esterase Cip2 from *Hypocrea jecorina*. Proteins 79:2588-92.

46. Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. Molecular Cloning Laboratory Manual, 2nd ed, vol. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

47. Scheller, H. V., and P. Ulvskov. 2010. Hemicelluloses. Annu Rev Plant Biol 61:263-89.

48. Simmons, B. A., D. Loque, and H. W. Blanch. 2008. Next-generation biomass feedstocks for biofuel production. Genome Biol 9:242.

49. Smaali, I., C. Remond, and M. J. O'Donohue. 2006. Expression in Escherichia coli and characterization of beta-xylosidases GH39 and GH-43 from *Bacillus halodurans* C-125. Appl Microbiol Biotechnol 73:582-90.

50. Stjohn, F. J., J. D. Rice, and J. F. Preston. 2006. Paenibacillus sp. strain JDR-2 and XynA1: a novel system for methylglucuronoxylan utilization. Appl Environ Microbiol 72:1496-506.

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

52. Vanfossen, A. L., I. Ozdemir, S. L. Zelin, and R. M. Kelly. 2011. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Biotechnol. Bioeng. 108:1559-1569.

53. Vanfossen, A. L., M. R. Verhaart, S. M. Kengen, and R. M. Kelly. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl. Environ. Microbiol. 75:7718-7724.

54. Verjans, P., E. Dornez, M. Segers, S. Van Campenhout, K. Bernaerts, T. Belien, J. A. Delcour, and C. M. Courtin. 2009. Truncated derivatives of a multidomain thermophilic glycosyl hydrolase family 10 xylanase from *Thermotoga maritima* reveal structure related activity profiles and substrate hydrolysis patterns. J Biotechnol 145:160-7.

55. Watanabe, H., and G. Tokuda. 2001. Animal cellulases. Cell Mol Life Sci 58:1167-78.

56. Winterhalter, C., and W. Liebl. 1995. Two Extremely Thermostable Xylanases of the Hyperthermophilic Bacterium *Thermotoga maritima* MSB8. Appl Environ Microbiol 61:1810-5.

57. Wulff-Strobel, C. R., and D. B. Wilson. 1995. Cloning, sequencing, and characterization of a membrane-associated *Prevotella ruminicola* B(1)4 beta-glucosidase with cellodextrinase and cyanoglycosidase activities. J Bacteriol 177:5884-90.

58. Xiao, Z., R. Storms, and A. Tsang. 2005. Microplate-based carboxymethylcellulose assay for endoglucanase activity. Anal. Biochem. 342:176-178.

59. Xiao, Z., R. Storms, and A. Tsang. 2004. Microplate-based filter paper assay to measure total cellulase activity. Biotechnol. Bioeng. 88:832-837.

60. Xu, Q., W. Gao, S. Y. Ding, R. Kenig, Y. Shoham, E. A. Bayer, and R. Lamed. 2003. The cellulosome system of *Acetivibrio cellulolyticu*s includes a novel type of adaptor protein and a cell surface anchoring protein. J Bacteriol 185:4548-57.

TABLES

Table 3.1. Primers for the PCR reactions to clone Calkr_2245-TM1, Calhy_0060, Calkro_0111-TM1&2, and Calhy_1629

Gene Name	F/R	Sequence (5→3)
Calkr_2245-TM1	F	GAG GAG AAG CCC GGT TAC GCC ATG GTT AAA AT
	R	GAC GAC GAC AAG ATG GCT TCA TCA GGG GTT
Calhy_0060	F	GAC GAC GAC AAG ATG AGT TCA AAT GGC AAG ATA
	R	GAG GAG AAG CCC GGTTACTGGTCA CTC ACA AC
Calkro_0111-TM1	F	GAC GAC GAC AAG ATG CAG AAA ATA AAG GGG AAT
	R	GAG GAG AAG CCC GGTTAT ATA GGT GAT TCA TTA AC
Calkro_0111-TM2	F	GAC GAC GAC AAG ATG ACA GAT ACA GGTTTG AG
	R	GAG GAG AAG CCC GGTTACTTG CTC CCATAC AC
Calhy_1629	F	GAC GAC GAC AAG ATG TCT ACA AAT CCA ATA GTG
	R	GAG GAG AAG CCC GGTTAT TTT GTT TGA AAC AAT T

Table 3.2. Representative SLH-domain proteins and their orthologs in *Caldicellulosiruptor* genomes that contain CBM and GH domains

DOMAINS	ORF	# AA
	**Csac_0678** (Calkro_2036, Calla_0352, Calhy_2064, Calkr_2007, Athe_0594, COB47)	755
	**Calkr_2245** (Calow_1924, Calkro_0402, Calla_0206)	2159
	**Calhy_0060** (Calkro_0072, COB47_0076, Csac_2548-9)	1732
	**Calkro_0111** (Calkro_0121)	2435
	**Calhy_1629**	1440
	**Calhy_2383**	2007

Table 3.3 Calkro_0111-0122 loci containing genes encoding hemicellulolytic enzymes from *C. kronotskyensis*

Locus tag	Protein name
Calkro_0111	SLH-SLH-SLH-CBM54-FN3-GH16-FLD-FN3-FN3-FLD-GH55-CBM32-CBM32
Calkro_0112	LuxR family transcriptional regulator
Calkro_0113	GH55-CBM32-CBM32-CBM32
Calkro_0114	GH81-CBM32-CBM32-CBM32-CBM6
Calkro_0115	GH30
Calkro_0116	AraC family transcriptional regulator
Calkro_0117	Extracellular solute-binding protein
Calkro_0118	Transporter
Calkro_0119	Transporter
Calkro_0120	GH2
Calkro_0121	SLH-SLH-SLH-CBM54-FN3-GH16-FLD-FN3-FN3-FLD-GH55-CBM32
Calkro_0122	GH88

GH55 and GH81:β-1,3-glucanase, GH30:β-glucosidase/xylosidase, GH2: β-galactosidase/mannosidase, GH88: β-glucuronyl hydrolase, GH16: β-1,3-1,4-glucanase

Table 3.4. Summary of the characterization of cellulolytic and hemicellulolytic enzymes containing CBMs and SLH domains from the genus *Caldicellulosiruptor*

ORF#	GH Family	CBM Family (Type)	Opt T (°C)	Opt pH	Substrates
Csac_0678	5	28(B)	75	5	cellulose, CMC, xylan, barley glucan, lichenan, switchgrass, filter paper
Calkr_2245	10	22(B) 9(C)	70 (TM1)	7 (TM1)	xylan arabinoxylan
Calhy_0060	16	4(B) 6(B) 54*	70	6	laminarin, lichenan, barley glucan
Calkro_0111	16 (TM1), 55 (TM2)	54* 3(C)	68 (TM1) 75 (TM2)	7 (TM1) 5.5 (TM2)	laminarin, lichenan, barley glucan (TM1), laminarin (TM2)
Calhy_1629	43	54*	75	5	xyloglucan, arabinoxylan, xylan

* CBM54 carbohydrate binding domains have been discovered recently and their biochemical properties are yet unknown.

FIGURES

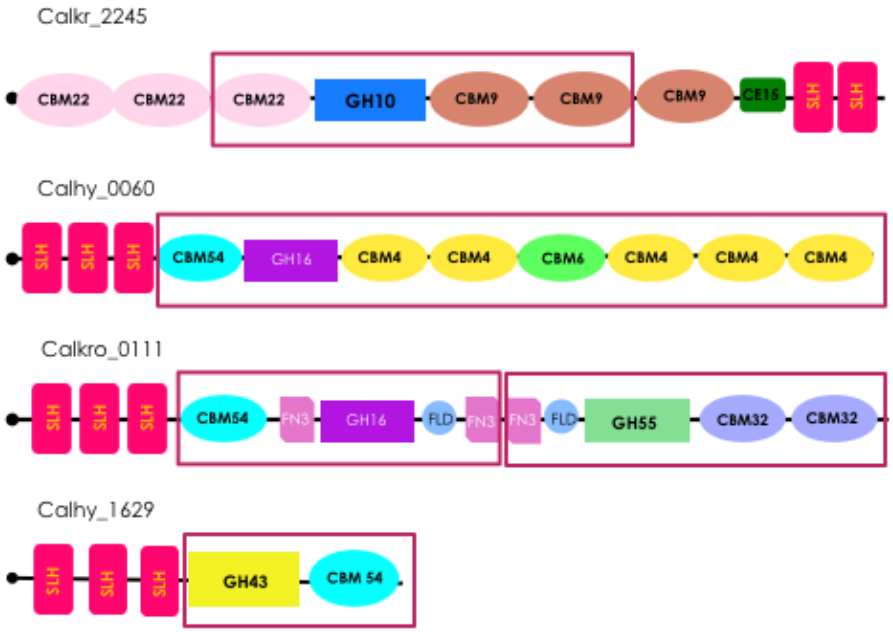

Figure 3.1. Truncation mutants (TMs) of *Caldicellulosiruptor* SLH-domain GHs. The following numbers indicate the first and the last amino acids included in the TMs. Calkr_2245-TM1:327-868 aa, Calhy_0060:423-1732 aa, Calkro_0111-TM1:212-1053 aa, Calkro_0111-TM2:1053-2435, and Calhy_1629:212-1440 aa.

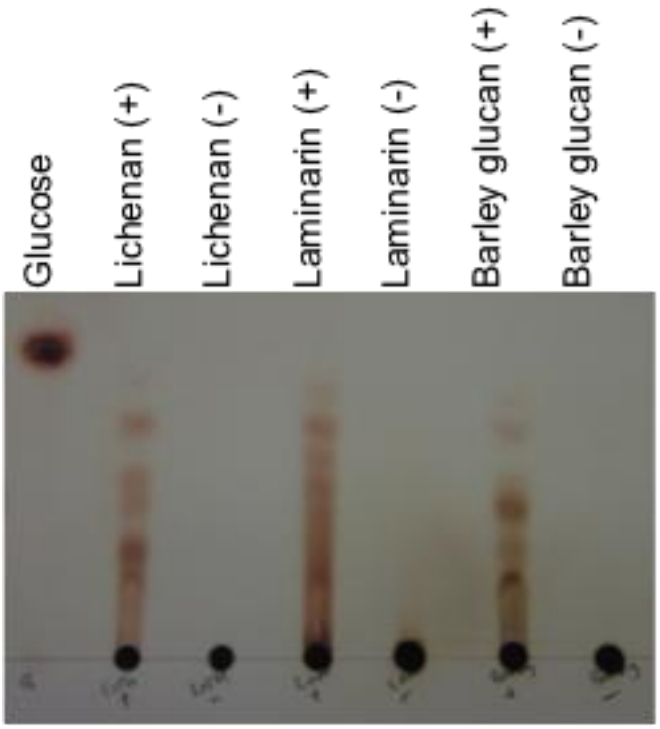

Figure 3.2. Thin layer chromatography (TLC) image showing the hydrolysis products of Calhy_0060. It degraded lichenan, laminarin, and barley glucan into their oligosaccharides.

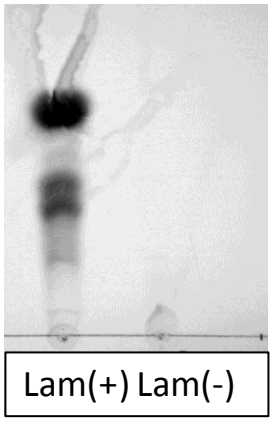

Figure 3.3. Thin layer chromatography (TLC image) of Calkro_0111-TM2 reaction with 1 % (w/v) laminarin. *Lane on the left (Lam(+)) is the reaction and the one on the right is the control with no enzyme (Lam(-)).

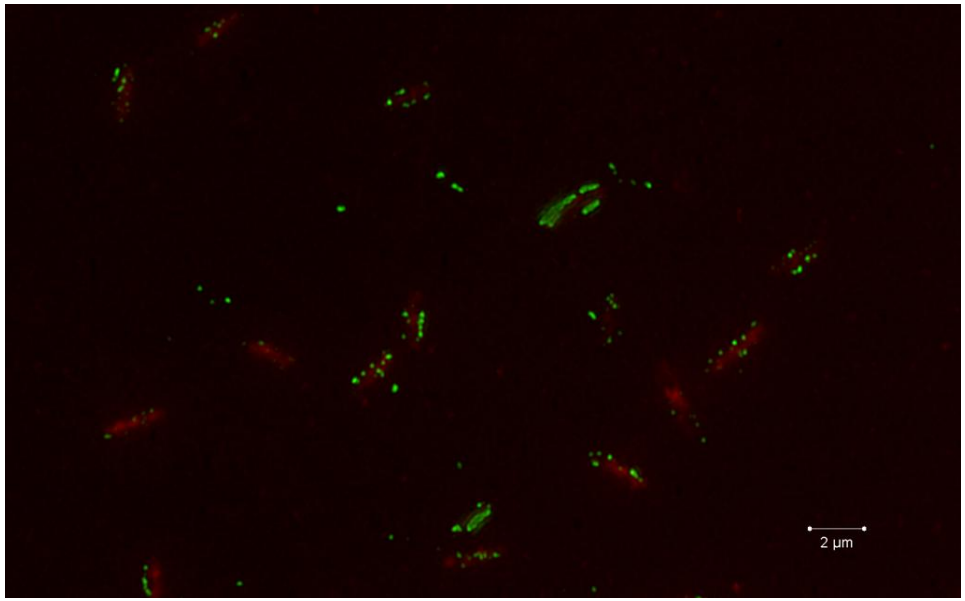

Figure 3.4. Structured illumination (SIM) micrograph of *C. saccharolyticus* cells labeled with SYTO61 (red) and incubated with the primary anti-Csac_0678 and Alexa Fluor 488 (green) secondary antibodies.

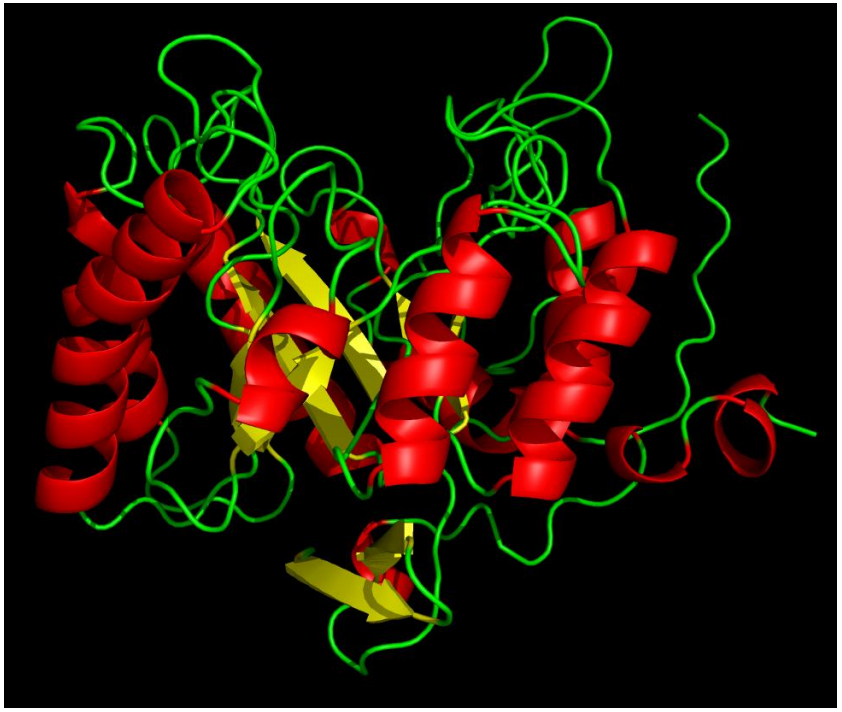

Figure 3.5. Crystal structure of Csac_0678 GH5 domain containing a TIM-barrel protein fold

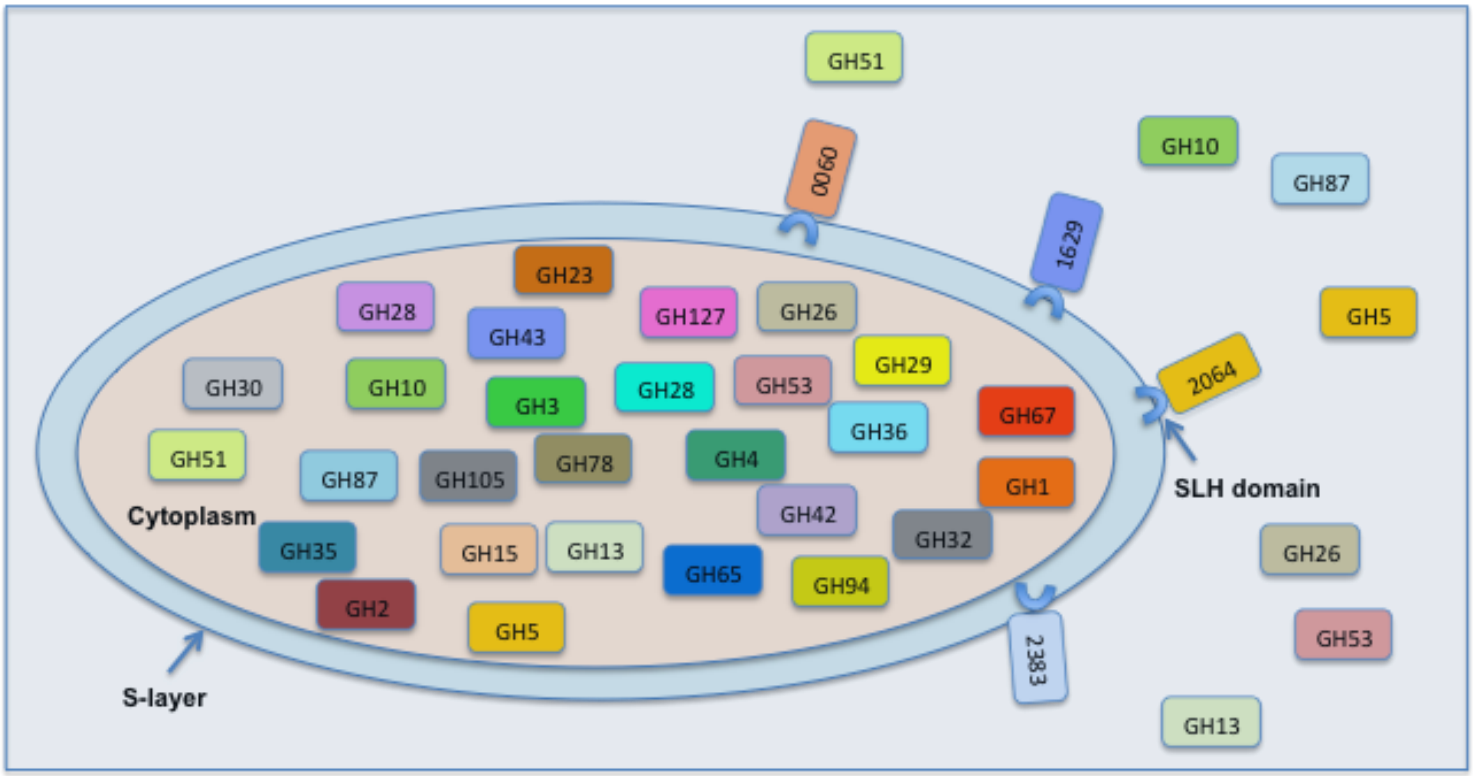

Figure 3.6. Localization of the putative 62 glycoside hydrolases (GHs) (representing 29 families) from *C. hydrothermalis*. Nine GHs (representing 7 GH families) are secreted and 4 (Calhy_0060-GH16, Calhy_1629-GH43, Calhy_2064-GH5, and Calhy_2383: GH87) are associated with the S-layer. GHs and CAZy family they belong to are represented as color-coded boxes.

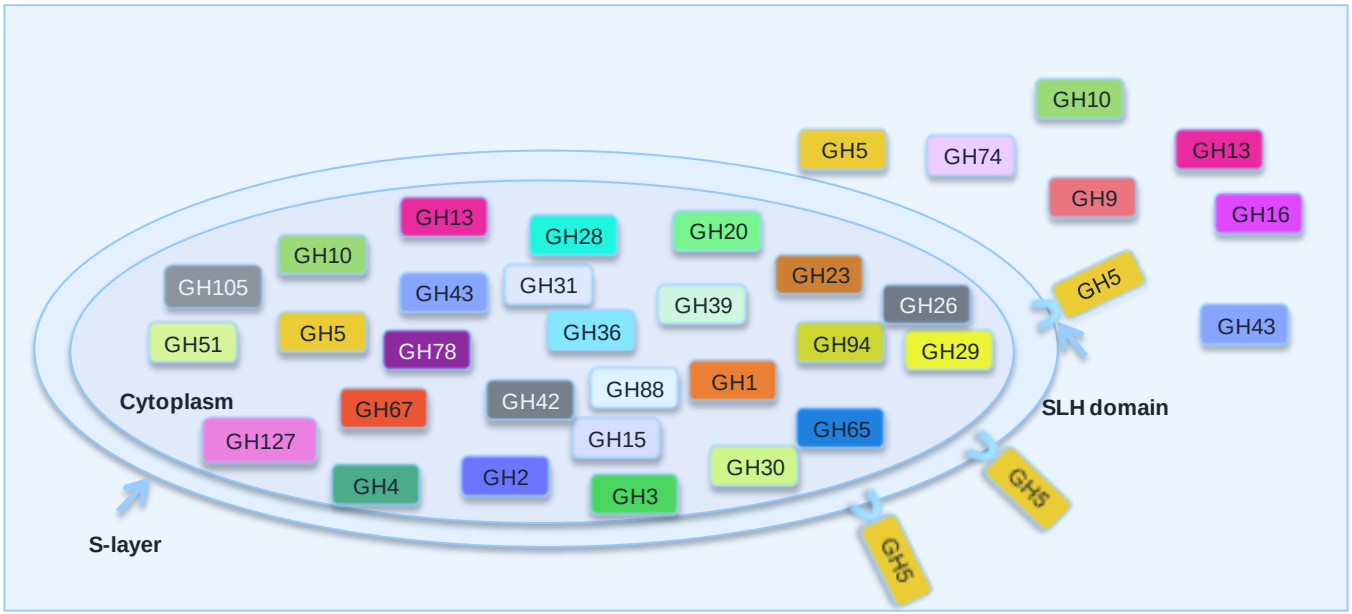

Figure 3.7. Localization of the putative 59 glycoside hydrolases (GHs) (representing 31 families) from *C.sacchrolyticus*. Thirteen GHs (representing 7 GH families) are secreted and one (Csac_0678) is associated with the S-layer. GHs and CAZy family they belong to, are represented as color-coded boxes.