ABSTRACT

LAEMTHONG, TUNYABOON. Engineering the Cell Surface of *Caldicellulosiruptor bescii* to Enhance Its Ability to Deconstruct Lignocellulosic Plant Biomass. (Under the direction of Dr. Robert Kelly).

Lignocellulose is considered to be a promising alternative feedstock for chemical and fuel production. Consolidated BioProcessing (CBP) is typically used for converting lignocellulose to chemicals and fuels. *Caldicellulosiruptor* species are extremely thermophilic, gram-positive fermentative anaerobes (T_{opt} 78^{o}C) and are considered to be promising candidates for CBP due to their ability to break down untreated plant biomass and metabolize pentose and hexose simultaneously. *Caldicellulosiruptor bescii*, one of the strongest cellulolytic species in the genus, has been systemically developed to become a better candidate among other species due to its established genetic manipulation tools. However, the strains developed so far have not yet attained industrial-level chemical production. Efficient plant biomass deconstruction ability by engineering the cell surface might be a key improvement of overall chemical and fuel production.

Caldicellulosiruptor species deploy secreted (Glucan Degradation Locus (GDL) enzymes and cell surface-associated multi-domain enzymes to deconstruct plant biomass (SLH Glycoside Hydrolases, SLH-GHs). Cell surface-associated enzymes often contain surface layer homology (SLH) domains, which attach enzymes to the bacterial cell surface. However, the *C. bescii* genome encodes only the smallest SLH glycoside hydrolase, which is Athe_0594. Representatives from all other classes of SLH-GHs (Calkro_0111, Calkro_0402, Calkro_0072, and Calkro_2036 from *Caldicellulosiruptor kronotskyensis*, Calhy_1629 and Calhy_2383 from *Caldicellulosiruptor hydrothermalis*) were produced recombinantly and biochemically characterized in vitro. Calkro_0111, Calkro_0072, and Calhy_2383 exhibited β-1,3-glucanase activity, Calkro_0402 was active on both β-1,3/1,4-glucan, and β-1,4-xylan, Calkro_2036 exhibited activity on both β-

1,3/1,4-glucan and β-1,4-glucan, and Calhy_1629 was active only on arabinan. Supplementing with individual enzymes or cocktails of SLH-GHs, increased in vitro sugar release from lignocellulosic plant biomass like sugar cane bagasse and poplar. Expression of non-native SLH-GHs in vivo improved total carbohydrate solubilization of sugar cane bagasse and poplar by up to 45% and 23%, respectively. Expression of Calkro_0402 in *C. bescii*, a xylanase/glucanase, improved xylose solubilization from poplar and sugar cane bagasse by over 70%. Even though *C. bescii* was a prolific cellulose plant biomass degrader, its secreted enzymes were naturally optimized. However, it could be further improved by supplementing non-native enzymes.

Caldicelllosiruptor sp. typically deploy other cell surface components to adhere to substrate as well. Tāpirins (Mr~70 kDa), non-catalytic proteins found nearby pili loci, are one of the cell surface structures found only in the genus *Caldicellulosiruptor*. Tāpirins were found binding specifically to cellulose, but were unable to bind to xylan, and were hypothetically involved in overall biomass hydrolysis. In vitro results indicated that tāpirins from less cellulolytic species, like Calhy_0908 from *Caldicellulosiruptor hydrothermalis* and Calkr_0826 from *Caldicellulosiruptor kritjanssonii* compared to Calkro_0844 from *Caldicellulosiruptor kronotskyensis*. Engineering the cell surface of *C. bescii* to become a better cellulose binder would increase chances of cellulolytic enzymes to approach the cellulose component in plant biomass. Three different tāpirins (Calhy_0908, Calkr_0826, and Calkro_0844) were overexpressed in *C. bescii* using P_{slp} promoter to test the relationship between substrate binding and substrate solubilization. *C. bescii* with overexpressed versions of Calhy_0908 and Calkr_0826 were closer in proximity of substrate compared to their parent strain and to the strain with Calkro_0844 expression. Live cell binding assays proved that the cellular binding fraction of Calkro_0826

expression strain was higher than the other strains. Moreover, poplar biomass solubilization of Calkr_0826 strain increased up to 15% compared with the parent strain.

Taken together, both catalytic and non-catalytic proteins of cellulolytic bacteria are important to further improve *C. bescii* to become an all-in-one microorganism to efficiently convert plant biomass into chemicals and fuels. This work will enable future efforts to engineer strains with non-native proteins that the wildtype strain lacks to align with feedstocks in order to further maximize industrial bio-based chemical and fuel production.

Engineering the Cell Surface of *Caldicellulosiruptor bescii* to Enhance Its Ability to Deconstruct Lignocellulosic Plant Biomass.

by
Tunyaboon Laemthong

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
2022

APPROVED BY:

_____	_____
Dr. Robert Kelly Committee Chair	Dr. Balaji Rao
_____	_____
Dr. Nathan Crook	Dr. Amy Grunden

BIOGRAPHY

Tunyaboon Laemthong was born and raised in Bangkok, Thailand. He attended Kasetsart University, Bangkok, Thailand for his undergraduate degree in chemical engineering. In college, he participated in undergraduate research projects, where his interest in biotechnology and microbiology started. During the last year of his undergraduate study, he had the opportunity to intern in one of the biggest petrochemical companies in Thailand, where he realized the importance of alternative energy, especially bio-based chemicals and fuels. These experiences inspired him to pursue graduate degree in Chemical Engineering. After, graduation and earning his Bachelor of Engineering in Chemical Engineering, he moved to the Missouri University of Science and Technology, where he received his Master's degree in Chemical Engineering. After graduation, he decided to a pursue PhD degree to expand his knowledge in biotechnology and chemical engineering. He began his PhD studies at North Carolina State University in Chemical and Biomolecular Engineering and spent the following years studying extreme thermophiles under the direction of Dr. Robert Kelly. After his completion of his PhD degree, he intends to pursue a career in academia doing research, which chemical engineering and biotechnology merge and pass on knowledge to the next generations.

ACKNOWLEDGMENTS

First of all, I would like to acknowledge my academic advisor, Dr. Robert Kelly, for all of his support, guidance, advice and leadership over the last few years. He is the most important person for my completing the PhD program since he gave me this great opportunity to join the Hyperthermophile Group and introduced me to the fascinating world of these microbes. I would like to thank all Kelly lab members and alumni, especially the *Caldi* team for their suggestions and help along the PhD journey. The undergraduate that has worked with me (Hailey O'Quinn) has been an important part of my graduate degree experience as well.

I am also grateful for the collaborations in the Chemical and Biomolecular Engineering Department. Also, I would like to thank our collaborators, Dr. Mike Adams's lab at the University of Georgia, Dr. Vladimir Lunin and Dr. Markus Alahunta at the National Renewable Energy Laboratory. I would also like to thank financial support from the Government of Thailand.

Most importantly, I have received so much support from my friends and family. I want to acknowledge my parents, Prasan and Kunnikar Laemthong, for their unconditional love and support. Without it, everything I have achieved would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES xiii
LIST OF FIGURES ix

CHAPTER 1: Bacterial and Archaeal Structural Surface 1
Introduction 2
S-layer structures and functions in different microorganisms 4
S-layer in gram-positive bacteria 4
S-layer in gram-negative bacteria 6
S-layer in archaea 7
S-Layer Homology (SLH) domains 9
Structure adjacent to S-layer 12
Pili 12
Pili structure and assembly (gram-positive vs. gram-negative) 12
Pili in gram-positive microbes 13
Pili in gram-negative microbes 13
Type IV pili 14
Archaeal pili 16
Flagella (bacterial and archaeal) 17
Secretome 18
Protein secretory machinery 18
Extracellular enzymes 20
References 26
CHAPTER 2: Enzymes from Extremely Thermophilic Bacteria and Archaea: Current Status and Future Prospects 47
Abstract 48
Introduction 49
Microbial diversity at high temperatures 50
Extremely thermophilic bacteria 51
Extremely thermophilic archaea 52
DNA polymerases 55
Proteases 57

Thermostable amylases from extreme thermophiles 59
Improving thermal stability of α-amylases from extreme thermophiles 60
β-glucanases 61
Hemicellulases 63
Surface (S)-layer homology domain glycoside hydrolases 66
Xylose isomerases 66
Carbon dioxide processing enzymes 68
Carboxylases 68
Carbonic Anhydrases 70
Carbonic anhydrases and carboxylases in concert 71
Hydrogenases and hydrogen production 71
Genetic tools for extreme thermophiles 74
Metabolic engineering 77
Future directions 80
Acknowledgments 81
References 91
CHAPTER 3: Engineering *Caldicellulosiruptor bescii* with Surface Layer Homology Domain-Linked Glycoside Hydrolases Improves Plant Biomass Solubilization 132
Abstract 133
Importance 134
Introduction 135
Results and discussions 137
Recombinant production of *Caldicellulosiruptor* SLH-GHs 137
Biochemical characterization of SLH-GH enymzes 138
Substrate specificity of *Caldicellulosiruptor* SLH-GHs 138
Engineering *C. bescii*'s S-layer to improve solubilization of plant biomass 140
Lignocellulosic plant biomass solubilization by *C. bescii* secretrome supplemented with SLH-GH cocktails 141
Lignocellulosic plant biomass solubilization by *C. bescii* SLH-GH expression strains 142
Materials and methods 145

Bacterial strains, plasmid, and reagents 145
Genetic manipulation of *C. bescii* 146
Protein expression and purification 146
Polysaccharide substrate specificity 148
pH and temperature optima 148
Enzyme thermostability 148
SLH-GH specific activity 149
Confocal microscopy 149
Lignocellulosic plant biomass preparation 150
Lignocellulosic plant biomass hydrolysis with SLH-GH cocktails 150
Biomass solubilization 150
Quantitative saccharification 151
Acknowledgments 151
References 152
CHAPTER 4: Improving biomass solubilization ability of *Caldicellulosiruptor bescii* using unique non-catalytic carbohydrate binding protein (tāpirins) 177
Abstract 178
Introduction 179
Results and discussions 181
Possible binding site of tāpirins 181
Improving cellulosic substrate binding of *C. bescii* by overexpressing non-native tāpirins 182
Lignocellulosic plant biomass solubilization 183
Conclusions and future directions 184
Materials and methods 185
Bacterial strains, plasmids, and reagents 185
Site-directed mutation study 185
Protein expression and purification 186
In vitro tāpirin binding assay 187
Genetic manipulation of *C. bescii* 187
Florescence microscopy 188

Scanning Electron Microscopy (SEM) 188
Live-cell binding quantification 188
Biomass solubilization 189
Acknowledgments 190
References 191

LIST OF TABLES

Table 1.1 Comparison between bacterial and archaeal S-layer 3

Table 2.1 α-Amylolytic enzymes found in extreme thermophiles 82

Table 2.2 β-Glucanases found in extreme thermophiles 83

Table 2.3 Surface layer homology (SLH) domain glycoside hydrolases (GHs) encoded in *Caldicellulosiruptor* sp. genomes 84

Table 2.4 Autotrophic and assimilatory carboxylases 85

Table 2.5 Genetic systems for extreme thermophiles 86

Table 3.1 *Caldicellulosiruptor* sp. surface layer homology domain glycoside hydrolases 170

Table 3.2 Specific activity of SLH-GHs on selected substrates 171

Table 3.3 *In vitro* biomass hydrolysis after 24 hours of incubation 172

Table 3.4 Improvement in sugar cane bagasse carbohydrate solubilization of SLH-GH strains 173

Table 3.5 Improvement in poplar carbohydrate solubilization of SLH-GH strains 174

Table 3.6 Primers used in this study 175

Table 4.1 Protein residues that were replaced with cysteine residues to create double site-directed mutations 201

Table 4.2 Densitometry analysis showing changes in binding between original and site-directed mutated tāpirins from in vitro binding assay 201

Table 4.3 Live cell binding assay result of bound fraction from tāpirin expression strain after incubated with Avicel 201

Table 4.4 Post-fermentation analysis of poplar biomass solubilization and acetate production of tāpirin expression strains 202

Table 4.5 Primers used in this study 202

LIST OF FIGURES

Figure 1.1 Schematic diagram of bacterial and archaeal S-layer 4

Figure 1.2 Schematic diagram explaining proposed mechanism of CAZymes counting SLH domains found in plant biomass degrading bacteria 24

Figure 2.1 Phylogenic tree of 16S rDNA sequences from prokaryotic organisms..... 87

Figure 2.2 DNA polymerase proofreading 3'-5' exonuclease activity..... 88

Figure 2.3 Soluble hydrogenases and their applications for cell free hydrogen production 89

Figure 2.4 Mutant generation using uracil auxotrophy and 5-FOA counterselection 90

Figure 3.1 SDS-PAGE gel of purified recombinant SLH-GHs produced in *E. coli* or *C. bescii* host..... 162

Figure 3.2 *Caldicellulosiruptor* species and substrate specificity of their SLH-GHs 163

Figure 3.3 Thermostability analysis of SLH-GHs at different temperature 164

Figure 3.4 Fluorescence microscopy showing surface bound GHs via SLH domains in different SLH-GH strains revealing that SLH domains in *Caldicellulosiruptor* SLH-GHs function as S-layer binding domains to associate the rest of domains on the cell surface..... 165

Figure 3.5 Carbohydrate solubilization of different SLH-GH strains compared with their parent strain..... 166

Figure 3.6 His-tagged expressed protein detection using dot blot analysis confirming the expression of Calkro_0111 in *C. bescii*..... 167

Figure 3.7 pH optima of SLH-GHs 168

Figure 3.8 Temperature optima of SLH-GHs..... 169

Figure 4.1 *In vitro* binding assay comparing differences in substrate binding between original and site-directed mutational tāpirins..... 196

Figure 4.2 Fluorescence microscopy showing bound cells on substrate surface..... 197

Figure 4.3 Scanning Electron Microscopic (SEM) images showing deformation of post-fermentation substrate surface of Calkr_0826 strains compared with its parent strain (MACB1018)..... 198

Figure 4.4 Quantitative analysis of live cell binding to substrate (Avicel) 199

Figure 4.5 Poplar biomass solubilization of all strains after 7-day fermentation 200

CHAPTER 1

Bacterial and archaeal structural surface

Introduction

Cell envelopes are important structurally for prokaryotic microorganisms, and are essential for their viability and to protect them from external environmental stress (1, 2). Cell envelopes consist of several structural components. The Surface layer (S-layer) is common to archaea and bacteria. S-layers are paracrystalline arrays of identical proteins or glycoproteins with a thickness of 5 to 25 nm and a M_r ranging from 40 to 200 kDa. These proteins are located on the outside of the cell wall with an outer smooth surface and a wrinkled inner surface (2, 3). S-layers have several functions in microorganisms including serving as a protective coating for cells and as a molecular permeable sieve, and providing cell shape determination and cell adhesion (4). S-layers are commonly found in Gram-negative, Gram-positive, Gram-variable bacteria, as well as archaea. Even though these microorganisms have S-layers in common, their S-layer could associate with different cell envelope structures. Gram-positive bacteria have thick peptidoglycan layers causing their S-layer subunits to associate directly with peptidoglycan, meanwhile archaeal S-layer subunits are bound directly to pseudomurein due to small periplasmic space (5). However, in case of archaea that lack a stiff cell wall, the S-layer is the main cell wall component closely connected to the plasma membrane. In some archaea, biopolymers, like pseudomurein or methanochondroitin, are among the components on the cell envelope (6-8). In contrast, S-layers in gram-negative bacteria attach to outer cell wall components such as liposaccharides (9-11). For gram-positive bacteria, their unit cell structure is compact and exhibit filling-forms network with gaps and pores. The surface proteins associate more tightly with the outer membrane in contrast to gram-negative bacteria (12). Comparison between characteristics of bacterial and archaeal is shown in **Table 1.1 and Figure 1.1**.

Table 1.1 - Comparison between bacterial and archaeal S-layer

Characteristics	Bacterial S-layer	Archaeal S-layer
Thickness	~5-10 nm	Much thicker than 10 nm
Location	- Gram-positive: attach to peptidoglycan -Gram-negative: attach to lipopolysaccharide on the outer surface	Most archaeal S-layers attach directly to cell plasma membrane
Structure	Smooth outside surface and more wrinkled inner surface	Structure looks like mushroom with domains attached to the plasma membrane
Components	Proteins or glycoprotein subunits	Proteins or glycoprotein subunits
Lattice types	Oblique (p1), square (p2), or hexagonal (p6)	Mostly hexagonal (p6)

Pores are also an important part of the S-layer. In archaea, the predominant lattice type is the hexagonal shape, which provides a larger pore size compared with other types. A study on the S-layer structure in *Methanosarnica acetivorans* revealed that their S-layer pores consist of three different pore types depending on the amino acids positioning with slightly to highly negatively charge. These intense negatively charged pores with limited size function as charge and size barrier to protect the cells from troublesome or chaotropic enzymes (13).

S-layer structures and functions in different microorganisms

S-layer in gram-positive bacteria: S-layers have been studied in various microorganisms and their function varies. Not only have the S-layers structure been studied, but the proteins

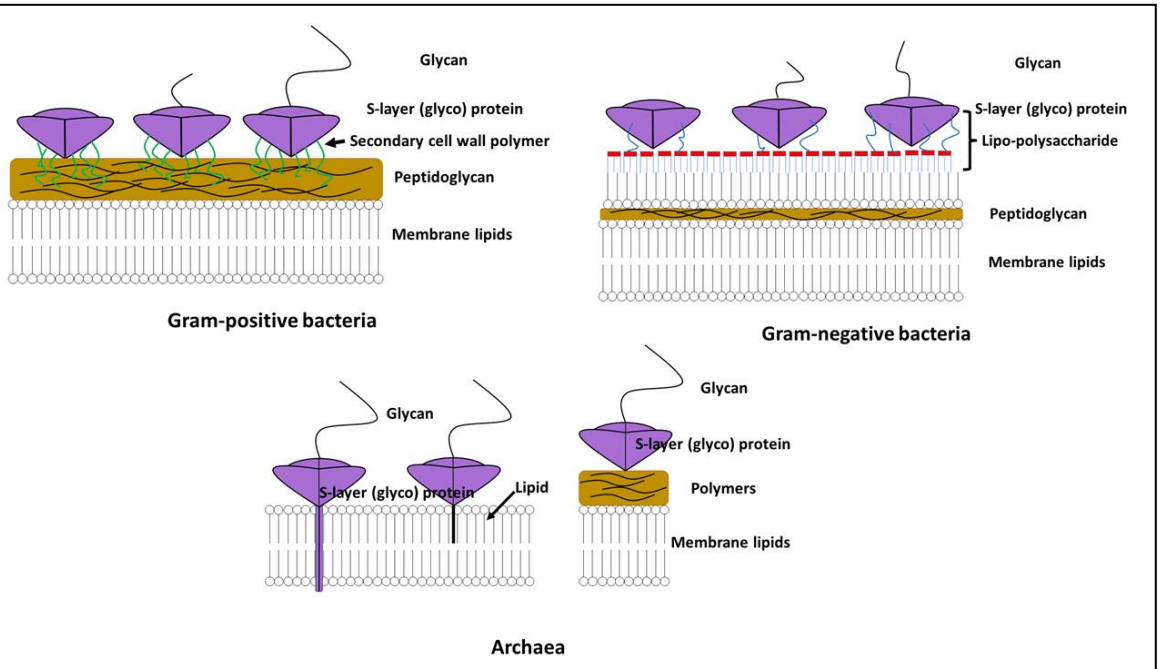


Figure 1.1 - Schematic diagram of bacterial and archaeal S-layer

expressed on S-layers have also been investigated. Within the genus *Bacillus*, *B. cereus* and *B. thuringiensis* cause *Bacillus* endophthalmitis, a severe intraocular infection leading to blindness once they infect the posterior of eyes. Its S-layer plays a key role in this pathogenesis due to special flagella, and pili on S-layer that facilitate their adhesion to human retinal cells. A mutant lacking S-layer proteins bound to retinal cells significantly less than the wild-type (14). Additionally, a small cytoplasmic protein, SlaQ, deposited on the cell membrane functions as a precursor of S-layer assembly in *B. anthracis*. This S-layer homology (SLH) protein is conserved among pathogenic bacilli, including *B. cereus*, *B. thuringiensis*, and *B. cytotoxius*. S-layer gene clusters are encoded in their genomes, indicating that the SlaQ protein supports S-layer creation (15). Sap and EA1, two secreted polypeptides, use SLH domains to bind to pyruvated secondary cell wall

polysaccharides adjacent to peptidoglycan. *B. anthracis* expresses these two SLPs at high levels to form a para-crystalline array on its surface. However, the major S-layer associated protein that promotes the host tissue adhesion leading to anthrax disease is BslA (16-18). These S-layer associated proteins in *B. anthracis* function as a protective coat and are involved in pathogenesis. S-layer proteins can also self-assemble as flat layers on various surfaces. The S-layer/allergen fusion protein (rSbpA/Bet) from *Lysinibacillus sphaericus* CCM 2177 with the signal peptide from *B. subtilis* α-amylase could be expressed in *B. subtilis*. The recombinant proteins from *B. subtilis* exhibited excellent recrystallization properties (19). This illustrates the biotechnological potential of using S-layer proteins for vaccine development.

The S-layer has been a target on the bacterial surface structure to find a cure for pathogenic bacteria causing major health problems worldwide, *Clostridium difficile*, for example. *C. difficile*'s S-layer is involved in sporulation of glycan-positive *C. difficile* strains, since the mutants with the defective S-layer or glycan-negative strains of *C. difficile* were consistently unable to form spores (20). SlpA apparently varies across the genus. *C. difficile* SlpA is composed of low-molecular mass subunits with high variability and high-molecular weight subunits. The low-molecular mass subunits are involved in anchoring the cell surface to host cells, more so than the higher-molecular mass subunits. *C. difficile* SlpA plays a significant role in pathogenesis (21).

The S-layer functions as a protective coat in *Lactobacillus* species; a major species in human intestines. S-layer of *L. acidophilus* inhibits the bacterial and viral infections by blocking the pathogen's interaction. The interaction between S-layer and viral molecules activates the signaling pathways to stop viral infection (22). S-layer proteins in *L. acidophilus* also help them adapt to stressful environments. S-layer proteins (SlpA and SlpX) expression level increased in the high-salt concentrations. The mutant lacking SlpA was more sensitive to osmotic pressure,

resulting from increased salt concentration (23). In addition, deletion of S-layer proteins, like IgdA, significantly changes the cell exterior resulting in loss of the cell's ability to adapt to stressful environments (24).

In the case of gram-positive bacteria living in extreme environments, like *Deinococcus radiodurans*. A complex-layered cell envelope supports the S-layer to become a radioresistant cell envelope. The hexagonal shaped S-layer consists of mainly single polypeptides. This S-layer is composed of elementary S-layer deinoxanthin binding subunits that bind to carotenoid deinoxanthin to become UV radiation resistance and increase the thermostability of cells. Interestingly, the S-layer complex is a porin-like structure with needle-like structures. This abnormal structure allows normal exchange of substances in and out of cells. Thus, the s-layer functions not only as a defensive barrier, but also as a selective permeable cell membrane (25).

Uncommon S-layer structures in *Viridibacillus arvi* are similar to those in *Lysinibaciilus* and *Bacillus* species. *V. arvi* (formerly *B. arvi*) was isolated from heavy metal contaminated environments. Slp1, Slp2, and Slp3 are its major the S-layer proteins. These S-layer proteins are assembled into a double layer with a square shape (P4 symmetry), as commonly found in *Bacillus* strains. However, the isoelectric points (pI) of mature protein were ~7.8-9.3 indicating a similar pI to proteins in *Lysinibacillus* strains. Moreover, *Lysinibacillus* and *Bacillus* S-layer proteins normally contain SLH domains to anchor to cell wall, but *V. arvi* S-layer proteins do not have SLH domains showing the uncommon feature of their S-layer proteins (26).

S-layer in gram-negative bacteria: Compared with gram-positive bacteria and archaea, only a few gram-negative bacterial S-layer have been studied. In *Caulobacter crescentus*, smooth lipopolysaccharides are essential for attaching S-layer to the cell surface. RsaA, a S-layer protein involved in secretion control, self-assembly and surface attachment, was unable to spontaneously

recrystallize without calcium addition (27). Protein-protein and protein-carbohydrate-lipid interaction occur on the S-layer of *Aeromonas hydrophila*. Truncated S-layer proteins could be exported from the periplasm to the cell surface; however, it could not be anchored to the cell surface or be assembled on the S-layer (28). In the case of fish pathogens, like *Aeromonas salmonicida*, the S-layer serves as a protective coat to thwart against fish defense mechanisms. Mutant S-layer proteins lacked the ability to self-assemble or anchoring to the cell surface containing the antigen (29).

S-layer in archaea: The outermost cell envelope structure in archaea is an S-layer. Archaea often live in extreme environments, characterized by high pressure, high temperature, high acidity, and have archaeal S-layers (30). The most well-characterized archaeal S-layer exhibits a pseudo-crystalline protein layer that covers the cell membrane. Since archaea lack peptidoglycan in the cell envelope, archaeal S-layers play a similar role to bacterial peptidoglycan to maintain cell stability and shape. Archaeal S-layers have antifouling properties, which help reduce non-specific binding of macromolecules thereby maximizing nutrient uptake. They also reduce flow resistance thereby facilitating archaeal flagella-driven motility since their properties are similar to nano shark-skin effect (31, 32). The archaeal S-layer encloses the cytoplasmic membrane and is decorated with an outermost layer of heteropolysaccharide, depending on the cell culture environment. The archaeal S-layer is stable against proteases, detergents, and survives high temperatures (33).

In hyperthermophilic archaeal genera like *Sulfolobus* and *Saccharolobus*, the typical archaeal S-layer protects cells from stressful environments and maintains cell shape (34). Endogenous CRISPR Type III was used to knockdown the *slab* gene, which acts as membrane-anchoring S-layer subunit for *Saccharolobus solfataricus* (35). Silencing the *slab* gene resulted in

loss of cell wall integrity and deformed cell shape because the S-layer lattices lost their binding affinity. Moreover, the defective S-layer may have affected viral infection (36). Additionally, *S. solfataricus* S-layer glycoprotein is the most complex archaeal glycan yet identified due to a monosaccharide added to glycan. This form of S-layer glycan is different from other glycans found in archaea (37). In order to adapt themselves to high temperature and acidic environments, the S-layer of *S. acidocaldarius* is highly glycosylated. The high glycosylation ratio *S. acidocaldarius* S-layer protein SlaA is higher than other archaeal S-layer proteins. Additionally, the glycans found on the S-layer act as the anchor for specific host or cells (38). The S-layer of *S. acidocaldarius* forms a uranyl carboxylate complex in moderate acidic pH environments (pH 4.5-6) by stimulating deprotonated carboxylic groups of specific protein residues, such as glutamic and aspartic acids (33). Mineral formation occurs passively on the S-layer within both the S-layer pore space and the quasi-periplasmic space after the metal ions interact indirectly with negatively charged molecules on the surface. The budding mechanism, called Endosomal Sorting Complex Required for Transport III, is used during the metal encrustation on S-layer. This phenomenon is a driving force for extracellular membrane production in order to eliminate damaged S-layer portions (39, 40).

Sulfolobus S-layers are comprised mainly of SlaA and SlaB proteins, which are conserved in *Sulfolobus* species. SlaB is similar to other archaeal S-layer proteins, such as found in *Staphylothermus marinus*, but no SlaA homologs are found in other archaea (41). SlaA serves as an S-layer assembly driving factor, in that its shapes determine the lattice dimension and topology. Meanwhile, SlaB functions as a membrane anchor and a cell periplasmic space ruler. The high porosity of S-layers increases its flexibility, allowing archaea to morph and divide. Additionally, the size and shape of pores have evolved to assist the settlement of cellular filament structures, like archaeal flagella and adhesive pili (42). In *Saccharolobus islandicus*, SlaA and SlaB are the

major glycosylated proteins found in the S-layer. Deletion of the *slaA* gene in the chromosome resulted in an increase in both cell size (with higher sensitivity to osmotic pressure) and the chance of cellular aggregation. In addition, mutants lacking the *slaA* gene vary in numbers of chromosome copies because cells were likely to clump, thus resulting in cell division defects. Meanwhile, the mutant lacking the *slaB* gene formed a partial SlaA (43). This suggested that both SlaA and SlaB play a key role in S-layer formation.

In *Haloferax volcanii*, the S-layer serves as the main component of the cell wall, providing structural rigidity and protection from the extreme environments. The recent model of S-layer glycoproteins indicated that the C-terminal transmembrane domain of the S-layer glycoproteins of *H. volcanii* anchors to the cell membrane. Meanwhile, the rest of *H. volcanii* S-layer glycoproteins were lipid-modified and associated with cell membrane once the S-layer gets hydrolyzed (44). Furthermore, the driving force of S-layer structural changes with salinity. Differences in salt concentration impact the protein N-glycosylation process, resulting in changes in both glycan and glycosylation sites. These changes in *H. volcanii* help it survive in the hypersaline environment (45). This indicates that the external environment affects S-layer structural changes. In hyperthermophilic archaea, such as *Pyrobaculum organotrophum*, S-layer subunits integrate into a periplasmic space, formed between the plasma membrane and the S-layer, which is used as a storage for enzymes and proteins involved in nutrient digestion and uptake (46).

S-Layer Homology (SLH) domains

Studies have been focused on characterization of the S-layer homology (SLH) domain containing proteins. Some S-layer proteins covalently anchor to the cell wall, whereas others bind to the cell wall via the interaction between the secondary cell wall polymers and S-layer homology

(SLH) domains. SLH domains in S-layer proteins are responsible for cell wall targeting. SLH domains have been identified in several extracellular proteins, including enzymes and S-layer proteins. SLH domains are found either at the N- or C- terminus of proteins. Recombinant S-proteins lacking SLH domains do not attach to the cell wall (47, 48). Most S-layer proteins bind to the secondary cell wall polymers, rather than directly to peptidoglycan. For example, SLH domains in the xylanases, Xyn10A, from *Acetivibrio thermocellus* and Xyn10B from *Thermoclostridium stercorarium* associate with the secondary cell wall polymer. Meanwhile *T. stercoranium* SdbA, whose SLH domains are homologous to Xyn10A and Xyn10B, binds to peptidoglycan (49). SpaA, an S-layer protein in *Paenibacillus alvei*, contains the SLH domains close to the end of the N-terminus. The SLH domains from SpaA and SlhA have specific protein residues conserved, which are also conserved in the S-layer protein found in *Bacillus anthracis*. These conserved protein residues have also been found in grooves of proteins containing single cell wall polymer-binding sites. This provides a structural basis of cell wall anchoring by SLH domains in *P. alvei* (50). Likewise, the inter-prong grooves of SLH domains found in thr protein of *B. anthracis* contribute to the secondary cell wall polymer association (17). SpaA contains SLH domains, containing the TRAE motif, a cell wall binding amino acid motif, as well as TRAQ and TVEE. These motifs are necessary for cell wall binding of SLH domains because SLH domains were sufficiently driven by two motifs. SLH domains have dual functions as secondary cell wall polymers and for peptidoglycan recognition. Foreign proteins containing SpaA SLH domains could be expressed on the surface of *P. alvei* (51). Interestingly, another SLH containing cell surface protein, SlaA, plays an important in bacterial swarming. Deletion of *slaA* caused the bacterial adhesion, colony morphology, and a decrease in biofilm formation. The innermost SLH

domain could potentially anchor the SlhA protein to the peptidoglycan layer, whereas the other domain could serve as an anchor to the single cell wall polymer (52).

The SLH motifs are also observed and studied in other microorganisms. For example, the S-layer protein, SbpA of *B. sphaericus,* utilizes non-covalent interactions to associate to the peptidoglycan-containing layer. Previously, it was believed that the three SLH motifs in the SLH domain were responsible for anchoring the subunits to the cell wall using the secondary cell wall polymer. This study, in contrast, proposed that these SLH motifs were not sufficient to perform the full function of lattice formation and self-assembly process. The SLH-like motif was necessarily required to accomplish the process (53). In the same genus, a pathogen like *B. anthracis* possesses S-layer proteins. Pathogenic bacteria require iron from heme and hemoglobin from mammalian host to survive. Under the iron starvation, *B. anthracis* utilizes BslK, a noncovalently associated S-layer protein containing SLH domains on the external cell surface to scavenge heme via protein-protein interactions and rapidly transfer the heme into an iron for survival (54).

SLH domains are also highly conserved in moderate thermophiles like *A. thermocellus*. For example, AncA, Slp1, and Slp2 are S-layer proteins containing three SLH domains and have binding specificity to different cell envelope structures. Previously, some studies suggested that SLH domains from Gram-positive bacteria, such as *Geobacillus stearothermophilus* and *B. sphaericus,* bind to secondary cell wall polymers rather than to peptidoglycan. However, this study proved that *A. thermocellus* Slp1-SLH and Slp2-SLH bound specifically to peptidoglycan, but not to the secondary cell-wall polymer, whereas AncA-SLH attached to only secondary cell wall polymer (55).

Structures adjacent to S-layer

Most microorganisms possess either pili or fimbriae, which are long filamentous structures extending from the cell surface, for various proposes. Pili are protein polymers that are involved in several processes, including biofilm formation, colonization, mobility, host cell and substrate adhesion, as well as pathogenicity (56).

Pili

Pili structure and assembly (gram-positive vs. gram-negative)

Pili are characterized into three different types: chaperone-usher pathway pili, mating pili or sex pili, and type IV pili (T4P). Chaperone-usher (CU) pathway pili are highly conserved in the secretion-assembly system among the Gram-negative proteobacteria, which are the major pathogenic bacteria that can cause diseases in humans and animals. Pili assembled by this pathway are mainly virulence factors of pathogenic bacteria. Pathogenic bacteria lacking this pathway lose their pathogenicity (57). The CU pathway consists of a periplasmic chaperone and an outer membrane usher. The periplasmic usher functions as a stabilizer for CU fimbriae, which are linear non-covalent multi-subunit polymers. Then, they translocate these polymers to the outer membrane usher for polymerization, which is later transported to the cell surface. Mating pili are deployed for DNA transfer between microorganisms. Sex pili subunits are formed from hydrophobic subunits existing in the inner membrane of bacteria prior to assembly. It was assumed that these sex pili depolymerize after the cell attaches to another cell, thereby resulting in bringing two cells closer to transfer DNA through pores. Similarly, T4P subunits used for pilus assembly are found in the inner membrane as a source of inner membrane proteins prior to assembly and after depolymerization. However, only largely conserved N-terminal α-helices with high

hydrophobicity in each subunit are inserted into the bacterial membrane, while most residues of sex pili subunit are inserted into the membrane (58).

Pili in gram-positive microbes:

General pili assembly in Gram-positive bacteria has been tentatively described, based on *Corynebacterium diphtheriae*, starting from multiple identical units of a backbone pilin that form the pilin shaft along with at least one or two minor pilins (59). Normally, there is one minor pilin at the tip, and the other one is at the shaft base. Once the pilin subunits move away from the cell membrane, they are connected as beads on a string using a sortase that is specific for pilin. The housekeeping sortase joins C-terminal subunits on a lysine residue introducing the covalent amide bond (60). Then, a general sortase covalently joins the pilus base to specific amino acid groups on cell wall to anchor the pilin assembly (61-63).

Pili in gram-negative microbes:

Pili of Gram-negative bacteria carry out diverse functions and are involved in pathogenicity. Pili assembly uses multiprotein complexes and occur on either only the outer membrane or both inner and outer membrane (57). To date, there are five known types of pili in Gram-negative bacteria: chaperone-usher (CU) pili, type IV pili (T4P), conjugative type IV secretion pili, curli fibers, and type V pili, which was recently discovered (64). The CU pathway is widespread across the Gram-negative bacterial pili assembly (65). For the CU pathway, a chaperone in the periplasmic space starts binding the pilin subunits, then the integral outer membrane channel protein, named usher, turns bound subunits into pilus fiber at the outer membrane (66). T4P are filaments composed of approximately 15-20 pilus subunits emerging from

the bacterial surface. Pilin subunit synthesis occurs in the cytosol and Sec machinery then transports the subunits across the inner membrane. They are anchored in the membrane by the N-terminal of commonly shared protein residues found in T4P. The prepilin peptidase removes the N-terminal leader sequence adjacent to the cytoplasmic side, followed by adding the methyl group on the N-terminal residue. The oxidoreductase enzyme introduces disulfide bonds to the anchored subunits, leading to globular domain folding in the periplasm (67). Gram-negative bacteria produce extracellular proteinaceous fibers, called Curli fibers that are involved in biofilm formation and sedimentation. Curli specific genes (*csg*) are primary gene regulating Curli fibers found in Proteobacteria and Bateroidetes (68, 69). Dedicated secretion-assembly, or the nucleation-precipitation pathway, mainly used for Curli fibers production, is a Type VIII secretion system (70). Last, type V pili have been also found in Gram-negative bacteria. The structure and the assembly mechanism of these pili have remained largely unknown; however, these pili have key roles in virulence of pathogenic bacteria (57). The domain structure of type V pili is clearly different from CU pili. The major difference is that the pilin domain of CU pili is on the N-terminus, whereas pilin domain of type V pili is on the C-terminus. Moreover, CU pili required a chaperone protein for assembly, but type V pili tend to have self-assembly ability with help from arginine-related proteases (71).

Type IV pili

Type IV pili (T4P) are not only commonly found in Gram-negative bacteria, but are also found in some Gram-positive bacteria. For Gram-negatives, T4P have several functions such as, locomotion, adherence to host cells, pathogenicity, DNA uptake, and protein secretion. T4P are found across the gram-negative bacteria (72). However, T4P are widespread in Gram-positive

bacteria. Since the cell wall structure of bacteria is different, there are some differences and similarities between Gram-positive and Gram-negative T4P (73, 74). It was noted that the amount of cysteine residues in Gram-positive T4P found in *Clostridium difficile, Clostridium perfringers, Dichelobacter nodosus* was significantly less than Gram-negative T4P (75,76). However, this was assumed to be due to the anaerobic conditions under which these organisms thrive, which precludes proper formation of disulfide linkages (76). However, studies of T4P in aerobic gram-positive bacteria also found issues with disulfide bond formation (77). This suggests that the lack of cysteine residues is commonly seen in T4P among Gram-positive bacteria (78). The interaction between the conserved hydrophobic N-terminal α-helix of each pilin initiates the assembly of T4P pilins into a pilus filament. Moreover, the pilin assembly requires Glutamate at position 5, which is highly conserved on each pilin subunit, to create a salt bridge of connecting subunit via the N-terminal amine (79). T4P are important for DNA uptake in mesophilic and thermophilic bacteria since their common function was to transport macromolecules, like protein or DNA, into the cytoplasm. Both pili assembly and disassembly are involved in DNA translocation for mesophiles. In extreme thermophilic microorganisms, like *Thermus thermophilus*, DNA uptake machinery is linked to pili biogenesis, indicating that there are some structural similarities between mesophilic and extremely thermophilic bacteria (80).

In *C. difficile*, genes required for T4P assembly and function have been reported recently. PilJ has unique structural features, namely a dual-pilin fold with a binding affinity to zinc ion (81). In *C. perfringers*, pilT, which has been frequently found in Gram-negative bacteria, is involved in extended pili retraction was also found in *C. perfringers* (76). Their pilT gene has a highly conserved motif compared with other species. Moreover, pilT mutant strains indicate the lack of pili on their surface. PilT might have roles in pili retraction or sensing mechanism to prevent them

from non-functional pili assembly (76). T4P has been recently discovered in Gram-positive cellulolytic microorganisms, like *Caldicellusiruptor bescii* (82). During plant biomass deconstruction, there were some uncharacterized hypothetical proteins found, including T4P in cellulose-binding fractions via proteomic screening (83). Based on the *in vivo* and *in vitro* data, T4P found in the genome of *C. bescii* was shown to play a role in binding with insoluble polysaccharides found in plant biomass (82). The T4P cellular machine responsible for pili assembly and retraction has been studied thoroughly in *Myxococcus xanthus.* The T4P machine of *M. xanthus* is composed of an outer membrane spanning pore with the ring below the outer membrane. Then, two visible rings are located in the periplasm, and another ring in the cytoplasm that surrounds the disc-like structure. There is a long stem arising from the inner membrane passing through the periplasmic rings and the outer membrane pore (84)**.**

Archaeal pili

T4P pili are not only found in bacteria, but are also present in archaea (85). Some hypotheses indicate that T4P can be found across prokaryotes for the same reason, which is prokaryotic adhesion. Archaeal flagella have been deployed mostly for their motility. The major components for archaeal flagellum synthesis share structural similarity to T4P, except for the adhesion function (32). Archaeal pili are different from archaeal flagella and bacterial T4P due to the significant differences in quaternary structure of pili, even though their subunits look similar (86). Archaeal flagella have N-terminal domain homologs of the N-terminal domain of bacterial T4P (87). The outer globular domains of bacterial T4P do not connect to each other, while the outer globular domains of archaeal flagella do. This indicates that bacterial T4P do not need any assembly for motility, but archaeal flagella need the contacted domains for some mechanical

properties of the filament leading flagella to flex (88). This shows how proteins adapt for uses for different purposes through evolution (89). Genes required for archaeal flagella vary, depending on the species (88). Based on available data, *flaHIJ* are found across archaeal flagella. The conserved genes in flagella biosynthesis are similar to bacterial T4P genes. Studies indicated that *flaHIJ* genes are vital for the assembly (90,91). Moreover, *flaK* is identified as a pre-flagellin peptidase. Archaeal flagella biosynthesis begins once the flagellin subunits are synthesized as preproteins along with the signal peptides. The signal peptides are removed once the preproteins become mature proteins leading to the flagella filaments (92).

Flagella (bacterial and archaeal)

Flagella structure and assembly have been extensively studied in both *E. coli* and *Salmonella* (93,94). Flagella filaments are formed by a single flagellin subunit. However, some bacterial species require multiple flagellin subunits for flagella synthesis (95). Flagella phenotypes are divided into three distinct types; *fla*, *che*, and *mot*, based on the motility defect of the mutant. The architectural domains of *fla* types are composed of the basal body forming a platform in the plasma membrane and has the C-ring, a gear-like structure connected to the cytoplasmic side. The rod arises from the basal body and moves to the cell envelope vertically, whereas the hook curved from the basal body functions as a joint. Lastly, the filament, the largest piece of the structure, forms a long helical polymer working as propeller. The che type refers to the mutant lacking colonizing ability on agar plates. The molecule-specific chemoreceptors play a role in changes of sensory adaptation and a shared signal transduction system interacting with the flagella to turn clockwise or counterclockwise (96).

The assembly of mot type or the motor is done by connecting the stator complexes around the MS ring in peripheral membrane. For *E. coli*, two MotA and four MotB copies are used to form the stator complexes. Once the flagellar regulon is activated these proteins are expressed. Changes in MotB are stimulated by the complex formation to attach to rigid peptidoglycan layer (97).

Secretome

Protein secretory machinery

Bacteria normally synthesize proteins intracellularly, some of which are exported extracellularly into the environment via two major bacterial protein export secretory pathways, Sec and Tat. Apart from these two commonly known pathways, there are also other secretion systems found mostly in gram-negative bacteria, such as Type I Secretion system (T1SS), T2SS, T3SS, T4SS, T5SS, and T6SS (98). However, this section will cover only Sec and Tat pathways, which are commonly found in most microorganisms. The most important machinery to export unfolded proteins out of the cytosol is Sec, or the ubiquitous general secretory pathway (99). For most proteins, the Sec pathway, consisting of ubiquitous, essential and universal export machinery, is usually deployed to translocate them through the plasma membrane or integrate them into the plasma membrane. The main component of the Sec pathway is the transmembrane SecYEG channel translocating proteins across or into plasma membrane (100). The other component that is essential to the pathway is SecA, which functions as an ATPase motor of post-translational export including the proof-reading function. SecA is located at the ribosome and works along with SecYEG as the translocase holoenzyme. SecA is apparently a multi-functional protein, which can locate to either the cytoplasmic space or the ribosome. The two-helix finger and a clamp work together during the ATPase cycle in order to push the precursor proteins through SecYEG channel,

independent of the sequence of the polypeptide (101, 102). Another function is to bind to the cytoplasmic or membrane associated dimer, acting as a preprotein receptor for secreted proteins (101). During the post-translational stage, the precursor proteins with less hydrophobic signal peptides are completely synthesized and translocated from the ribosome. At this stage, the precursor proteins interact with posttranslationally interacting proteins, like the chaperones GroEL-GroES, DnaK-DnaJ-GrpE, the specific targeting chaperone SecB, and the soluble form of SecA. These interact with a precursor protein to ensure that they are not aggregated and are exported in a competent and unfolded structure. Afterwards, the translocated precursor proteins are transferred to the translocase, then translocated to the SecYEG pore with the use of ATPase from SecA. The SecYEG complex consists of SecD and SecF, which apply a force to translocate proteins through the SecYEG pore (103, 104).

Another important protein secretion pathway is Twin-Arginine Translocation (Tat) pathway which works parallelly to the Sec pathway (105). As mentioned above, the Tat pathway is mostly for folded protein translocation across the bacterial membranes. The Tat pathway is known for its capability to handle completely folded cargo proteins. Distinguishing proteins to be transported via Sec on Tat pathways depends on the signal peptides presenting on the N-terminal of proteins. Proteins that would be translocated via the Tat pathway normally conserve twin-arginine (RR) motifs in their signal peptides (99). To date, the Tat pathway is known to be involved in many vital processes in microorganisms, such as energy metabolism, cell envelope synthesis, cell division, motility and even pathogenicity (106, 107). The Tat pathway has been extensively studied in *E. coli* (108-110). Based on the *E. coli* system, the Tat pathway consists mainly of three major membrane-integral components, TatA, TatB, TatC, and minor components like TatE. TatA, whose size is about ~9 kDa, is the main component found in the Tat pathway. It is likely

responsible for forming the translocase channel. The structure of TatA, TatB and TatE are found similar to each other with a different short N-terminal domain exposed to the periplasm. Subsequently, TatA is sometimes replaced by TatE. However, the size of TatE (~7 kDa) is substantially smaller than TatA. TatE interacts with Tat signal peptide and can prevent TorA from premature cleavage. TatB (~18 kDa) binds specifically to the Tat signal peptides and the mature proteins (111). Even though there are some structural similarities between TatB and TatA, TatB contains longer amphipathic helix and a longer unstructured C-terminal region. It is possible that TatB originated from TatA. TatC, the largest among other components, is highly-conserved across the Tat pathway. It helps in cargo binding. TatC structure is different to other components because it contains six transmembrane helices and an N- and C- topology (112, 113). The mechanism of Tat pathway initiates once the cargo signal peptide interacts with the docking complex consisting of TatB and TatC. TatC then inserts the signal peptide into the membrane, leading to the oligomerization of TatA subunits. This step is assumed to create the transient translocation pore surrounding the substrate. Some species may use TatE as a substitution for TatA, due to their analogous structure (112, 113).

Extracellular enzymes

Microorganisms secrete enzymes and those secreted enzymes function outside the cell, and are referred to as exoenzymes or extracellular enzymes. Heterotrophic bacteria normally secrete enzymes to interact with the environment. Some utilize the secreted enzyme to deconstruct organic macromolecules for their survival, such as breaking down nutrients in the environment, as seen with cellulolytic bacteria (114). In nature, living microorganisms require nutrients, which are found in the form of complex chemical macromolecules consisting of variety of polymers (115).

These polymers are not small enough to transport across the microbial membrane; thus, they need enzymes to deconstruct polymers and transport usable molecules into cells. This mechanism is deployed among microorganisms found in marine, soil, extreme environments, etc. (116-120). Some pathogenic bacteria secrete extracellular enzymes as virulence factors to break down host defensive cells to invade the host and cause diseases (121).

Extracellular hydrolytic enzymes are found on the outside of the cell surface, which can be categorized into two types: the cell surface-associated enzyme and free enzymes. The surface-associated enzymes are attached to some part of the outer surface structure of cells, while the free enzymes are released into the environment (122, 123). Presumably, the enzyme deployment strategy is often related to the different lifestyle of microorganisms. The surface- associated enzymes are likely useful to the microorganisms that live in low nutrient environments where free enzymes are not the most cost-effective strategy to recruit dilute nutrients. This is supported by a mathematical model of enzyme strategies in free-living heterotrophic bacteria in the oceans (124).

To date, extracellular enzymes have been widely used in industrial processes. The biotechnological uses of bacteria include production of valuable products contained in feedstocks found abundantly in nature, such as plant biomass (125). Bioprocessing can convert various substrates into intermediate or final products by using extracellular enzymes from different microorganisms. Enzymes or biocatalysts are often deployed in the industries due to their unique properties (126). As industries evolve to be environmentally friendly, toxic or hazardous chemicals used in some industrial processes are being replaced with enzymes. This means biotechnological processes have come to play in the conventional processes (127). Enzymes are not only making the processes become more environmentally friendly, they also have high specificity, selectivity to substrates, and decrease energy consumption compared with other approaches (128).

Cellulases (endo-β-1,4-glucanase) and xylanases (β-1,3-glucanase and endo-β-1,4-mannase) are widely used for saccharification in industry in the last decade. Several industries deploy enzyme-based plant biomass conversion as one of their major processes to overcome the plant biomass recalcitrance (129). Examples are ethanol production, food processing, animal food processing, textile industry (130, 131). Cellulases and xylanases are normally obtained from either bacteria or fungi (132, 133). In the bioethanol production, lignocellulosic plant biomass is often used as low-cost, abundant feedstocks. Plant biomass is mainly composed of cellulose, hemicellulose, and lignin. The major substrates used in the industry are cellulose and hemicellulose (134). Chemical and physical pretreatments are often done prior to the fermentation by yeast to convert released simple sugars into ethanol. There are several conventional established pretreatments, such as steam explosion, acid or alkaline treatment, or organic solvents (135). However, these steps are not considered environmentally friendly and cost-effective processes. Meanwhile, replacing these steps with cellulolytic enzymes would potentially improve the processes. Cellulose and hemicellulose are obtained after the pretreatment and then converted into fermentable sugars by cellulases and hemicellulases (136). However, other hydrolytic enzymes are also used for plant biomass depending on the types of the feedstocks. These enzymes include amylases, esterases, phytases, and β-glucanases (137, 138).

In the food industries, microbial extracellular enzymes are also deployed in several applications. Microbial enzymes have been used in the processes due to their stability compared with enzymes from plants and animals. One obvious benefit that microbial enzymes has is that they can be produced via fermentation, which requires much less space and time (139). Food enzymes found in the food industry include cellulases, proteases, lipases, amylases, and pectinase.

These enzymes are mainly deployed in food production, such as starch processing, wine fermentation, cheese manufacturing, etc. (140).

As mentioned above, secreted enzymes are currently deployed in a variety of industries. However, there are also other enzymes that are anchored to the cell surface. The enzymes are bound to cell surface by some surface-associated proteins or some cell surface structure mentioned earlier in this chapter.

Some Carbohydrate-Active enZymes (CAZymes) containing SLH domains have been characterized in several plant biomass degrading microorganisms (141). *Caldanaerobius polysaccharolyticus*, a thermophilic anaerobe whose genome encodes hydrolytic enzymes that have been used for xylan hydrolysis. In particular, Xyn10A (160-170 kDa), a multimodular endoxylanase, containing two family 22 CBMs, GH10 endoxylanase, two family 9 CBMs, and followed by Surface Layer Homology (SLH) domains at the C-terminus has been applied. Xyn10A was predicted to anchor to the cell surface via SLH domains and utilize the GH10 domain to hydrolyze the xylan found in plant biomass, as described in **Figure 1.2** (142). In *Bacilllus* sp., AapT containing GH and SLH is α-amylase-pullulanase. The enzyme can release maltose, maltotriose, and maltotetraose from raw potato starch and amylose (143). There are various CAZymes containing SLH domains that have been characterized in xylan-degrading bacteria like *Paenibacillus* sp. LamA (44 kDa) containing a signal peptide, SLH domains, GH16, three CBM4 domains, and other domains was active towards insoluble complex substrates like laminarin. This enzyme released mainly laminaritriose from laminarin substrates. Moreover, the catalytic domain on the enzyme could specifically hydrolyze β-1,3- and β-1,3-1,4-glucans (144).

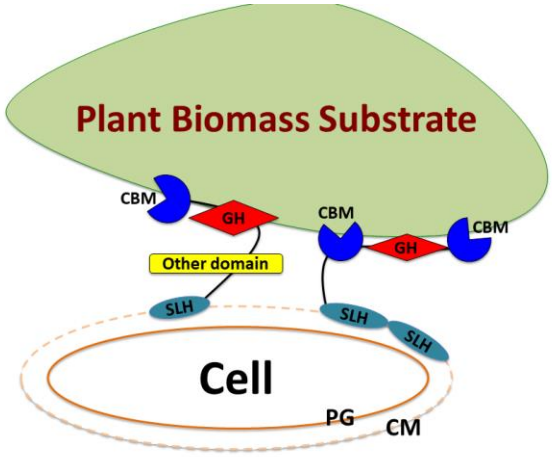


Figure 1.2 - Schematic diagram explaining proposed mechanism of CAZymes containing SLH domains found in plant biomass degrading bacteria. PG: peptidoglycan, CM: cell membrane, CBM: Carbohydrate binding module, GH: glucoside hydrolase domain, SLH: Surface Layer Homology

Furthermore, a surface associated enzyme from *Paenibacillus* sp. strain W-61, Xyn5, is a cell surface -anchored modular xylanase. Xyn5 consists of two family 22 CBM, GH10, CBM9, and three repeating SLH domains on the C-terminus. Xyn5 was disseminated in the cell envelope between peptidoglycan and the S-layers (145). Xyn5 is localized on the cell surface via the SLH domains, meanwhile CBM9 modules combine the cellulose microfibrils of the plant cell wall (146). *Paenibacillus sp*. strain FPU-7 produces a surface-associated chitinase, ChiW (150 kDa). *Paenibacillus sp.* strain FPU-7 ChiW consists of repeating SLH domains after the signal peptide, followed by catalytic GH18 domains. ChiW could be successfully produced recombinantly in *E. coli* after signal peptide and SLH domains deletion147. The activity of mutant chitinase that remained was sufficiently high to degrade the insoluble chitin and still maintained its thermostability (129).

There are also other thermostable CAZymes containing SLH domains from *Acetivibrio* sp. that have been characterized (148). *A. thermocellus* XynX xylanase (116 kDa) consists of N-

terminal thermostabilizing homologous, GH, CBM, and three repeating SLH domains (149). The size of non-cellulosomal fraction, where the SLH domain was deleted, was much less than the cellulosomal fraction which was the full-length of XynX. This indicated that the SLH domains are relevant to the cell envelope interaction of the XynX (150). Similarly, deletion of SLH domains at N-terminus of the cellulosomal *A. thermocellus* Lic16A (148 kDa) did not remove the intermolecular interaction between the surface attached enzymes containing SLH domains at the C-terminus. SLH domains at the N-terminus interact with each other and to form an interaction between polysaccharide-hydrolyzing enzyme network on the cell surface (151). A multi-domain cell surface-associated xylanase was also discovered in other species like *C. josui* Xyn10A (115 kDa), *T. stercorarium* XynC (115 kDa) and *T. stercorarium* Xyn10FigureF῀B (114 kDa). *C. josui* Xyn10A and *T. stercorarium* XynC are composed of a signal peptide and other six domains, namely two thermostabilizing domains, GH10, CBM9, and two SLH domains. These two enzymes are cell-associated rather than being targeted to the secretome. *T. stercorarium* Xyn10B was easily cleaved from the cell surface using hydrofluoric acid treatment. It was assumed to be secreted and anchored to the cell surface via SLH domains (152-154). To this point, the repeated SLH domains have been found frequently in thermophiles, either at the C- or N-terminus of the cell surface associated extracellular enzymes. For example, XynA xylanase (155-157), AmyB (pullulanase) (158), β-1,4-xylanases from *Thermoanerobacter* sp. (159), and Calkro_0402 xylanase from *Caldicellulosiruptior* sp. (160) have been characterized. It is not known why thermophiles have the SLH-containing CAZymes encoded in their genomes. This was possibly an adaptation to unusual environments, such as hot springs, where free extracellular enzymes would diffuse away from these microorganisms without benefiting them. The extracellular enzymes are found to be useful for various industrial processes. However, some processes still require highly thermostable

enzymes that can be derived from extremely thermophilic bacteria or archaea. Therefore, the next chapter is specific to enzymes from extremely thermophilic microbes that are currently used for industrial applications.

REFERENCES

1. Beveridge TJ, Graham LL. 1991. Surface layers of bacteria. Microbiol Rev 55:684-705.
2. Sa M, Sleytr UWEB. 2000. MINIREVIEW S-Layer Proteins. 182:859-868.
3. Sleytr UB, Schuster B, Egelseer EM, Pum D. 2014. S-layers: Principles and applications. FEMS Microbiol Rev 38:823-864.
4. Engelhardt H, Peters J. 1998. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer–cell wall interactions. J Struct Biol 124:276-302.
5. Küpcü S, Sára M, Sleytr UB. 1995. Liposomes coated with crystalline bacterial cell surface protein (S-layer) as immobilization structures for macromolecules. Biochim Biophys Acta Biomembr 1235:263-269.
6. Kandler O. 1993. Cell wall biochemistry and three-domain concept of life. Syst Appl Microbiol 16:501-509.
7. König H. 1988. Archaeobacterial cell envelopes. Can J Microbiol 34:395-406.
8. Kreisl P, Kandler O. 1986. Chemical structure of the cell wall polymer of methanosarcina. Syst Appl Microbiol 7:293-299.
9. Domínguez-Medina CC, Pérez-Toledo M, Schager AE, Marshall JL, Cook CN, Bobat S, Hwang H, Chun BJ, Logan E, Bryant JA, Channell WM, Morris FC, Jossi SE, Alshayea A, Rossiter AE, Barrow PA, Horsnell WG, MacLennan CA,

Henderson IR, Lakey JH, Gumbart JC, López-Macías C, Bavro VN, Cunningham AF. 2020. Outer membrane protein size and LPS O-antigen define protective antibody targeting to the Salmonella surface. Nat Commun 11.

10. Engelhardt H, biology JPJos, undefined. 1998. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer–cell wall interactions. Elsevier.

11. von Kügelgen A, Tang H, Hardy GG, Kureisaite-Ciziene D, Brun YV, Stansfeld PJ, Robinson CV, Bharat TAM. 2020. In situ structure of an intact lipopolysaccharide-bound bacterial surface layer. Cell 180:348-358.e15.

12. Engelhardt H, Gerbl-Rieger S, Krezmar D, Schneider-Voss S, Engel A, Baumeister W. 1990. Structural properties of the outer membrane and the regular surface protein of *Comamonas acidovorans*. J Struct Biol 105:92-102.

13. Arbing MA, Chan S, Shin A, Phan T, Ahn CJ, Rohlin L, Gunsalus RP. 2012. Structure of the surface layer of the methanogenic archaean *Methanosarcina acetivorans*. Proc Nat Acad Sci U S A 109:11812-11817.

14. Mursalin MH, Coburn PS, Livingston E, Miller FC, Astley R, Flores-Mireles AL, Callegan MC. 2020. *Bacillus* S-layer-mediated innate interactions during endophthalmitis. Front Immunol 11:215-215.

15. Nguyen-Mau SM, Oh SY, Schneewind DI, Missiakas D, Schneewind O. 2015. *Bacillus anthracis* SlaQ promotes S-layer protein assembly. J Bacteriol 197:3216-3227.

16. Kern J, Schneewind O. 2010. BslA, the S-layer adhesin of *B. anthracis*, is a virulence factor for anthrax pathogenesis. Mol Microbiol 75:324-332.

17. Kern J, Wilton R, Zhang R, Binkowski TA, Joachimiak A, Schneewind O. 2011. Structure of surface layer homology (SLH) domains from *Bacillus anthracis* surface array protein. J Biol Chem 286:26042-26049.

18. Sychantha D, Chapman RN, Bamford NC, Boons GJ, Howell PL, Clarke AJ. 2018. Molecular basis for the attachment of S-layer proteins to the cell wall of *Bacillus anthracis*. Biochemistry 57:1949-1953.

19. Ilk N, Schumi CT, Bohle B, Egelseer EM, Sleytr UB. 2011. Expression of an endotoxin-free S-layer/allergen fusion protein in gram-positive *Bacillus subtilis* 1012 for the potential application as vaccines for immunotherapy of atopic allergy. Microb Cell Fact 10:6-6.

20. Richards E, Bouché L, Panico M, Arbeloa A, Vinogradov E, Morris H, Wren B, Logan SM, Dell A, Fairweather NF. 2018. The S-layer protein of a *Clostridium difficile* SLCT-11 strain displays a complex glycan required for normal cell growth and morphology. J Biol Chem 293:18123-18137.

21. Merrigan MM, Venugopal A, Roxas JL, Anwar F, Mallozzi MJ, Roxas BAP, Gerding DN, Viswanathan VK, Vedantam G. 2013. Surface-layer protein A (SlpA) is a major contributor to host-cell adherence of *Clostridium difficile*. PLoS ONE 8.

22. Acosta MP, Geoghegan EM, Lepenies B, Ruzal S, Kielian M, Martinez MG. 2019. Surface (S) layer proteins of *Lactobacillus acidophilus* block virus infection via DC-SIGN interaction. Front Microbiol 10.

23. Palomino MM, Waehner PM, Fina Martin J, Ojeda P, Malone L, Sánchez Rivas C, Prado Acosta M, Allievi MC, Ruzal SM. 2016. Influence of osmotic stress on the

profile and gene expression of surface layer proteins in *Lactobacillus acidophilus* ATCC 4356. Appl Microbiol Biotechnol 100:8475-8484.

24. Klotz C, Goh YJ, O'Flaherty S, Johnson B, Barrangou R. 2020. Deletion of S-layer associated Ig-like domain protein disrupts the *Lactobacillus acidophilus* cell surface. Front Microbiol 11.

25. Farci D, Aksoyoglu MA, Farci SF, Bafna JA, Bodrenko I, Ceccarelli M, Kirkpatrick J, Winterhalter M, Kereïche S, Piano D. 2020. Structural insights into the main S-layer unit of *Deinococcus radiodurans* reveal a massive protein complex with porin-like features. J Biol Chem 295:4224-4236.

26. Suhr M, Lederer FL, Günther TJ, Raff J, Pollmann K. 2016. Characterization of three different unusual s-layer proteins from *Viridibacillus arvi* jg-b58 that exhibits two super-imposed s-layer proteins. PLoS ONE 11:e0156785-e0156785.

27. Nomellini JF, Kupcu S, Sleytr UB, Smit J. 1997. Factors controlling in vitro recrystallization of the *Caulobacter crescentus* paracrystalline S-layer. J Bacteriol 179:6349-6354.

28. Doig P, McCubbin WD, Kay CM, Trust TJ. 1993. Distribution of surface-exposed and non-accessible amino acid sequences among the two major structural domains of the S-layer protein of *Aeromonas salmonicida*. J Mol Biol 233:753-765.

29. Thomas S, Austin JW, McCubbin WD, Kay CM, Trust TJ. 1992. Roles of structural domains in the morphology and surface anchoring of the tetragonal paracrystalline array of *Aeromonas hydrophila*. Biochemical characterization of the major structural domain. J Mol Biol 228:652-661.

30. Klingl A. 2014. S-layer and cytoplasmic membrane–exceptions from the typical archaeal cell wall with a focus on double membranes. Front Microbiol 5:624.

31. Li PN, Herrmann J, Tolar BB, Poitevin F, Ramdasi R, Bargar JR, Stahl DA, Jensen GJ, Francis CA, Wakatsuki S, van den Bedem H. 2018. Nutrient transport suggests an evolutionary basis for charged archaeal surface layer proteins. ISME J 12:2389-2402.

32. Pohlschroder M, Pfeiffer F, Schulze S, Farid M, Halim A. 2018. Archaeal cell surface biogenesis. FEMS Microbiol Rev 027:694-717.

33. Reitz T, Rossberg A, Barkleit A, Steudtner R, Selenska-Pobell S, Merroun ML. 2015. Spectroscopic study on uranyl carboxylate complexes formed at the surface layer of *Sulfolobus acidocaldarius*. Dalton Trans 44:2684-2692.

34. Jing H, Takagi J, Liu J-h, Lindgren S, Zhang R-g, Joachimiak A, Wang J-h, Springer TA. 2002. Archaeal surface layer proteins contain β propeller, PKD, and β helix domains and are related to metazoan cell surface proteins. Structure 10:1453-1464.

35. Zink IA, Pfeifer K, Wimmer E, Sleytr UB, Schuster B, Schleper C. 2019. CRISPR-mediated gene silencing reveals involvement of the archaeal S-layer in cell division and virus infection. Nat Commun 10:1-14.

36. Zink IA, Pfeifer K, Wimmer E, Sleytr UB, Schuster B, Schleper C. 2019. CRISPR-mediated gene silencing reveals involvement of the archaeal S-layer in cell division and virus infection. Nat Commun 10.

37. Palmieri G, Balestrieri M, Peter-Katalinić J, Pohlentz G, Rossi M, Fiume I, Pocsfalvi G. 2013. Surface-exposed glycoproteins of hyperthermophilic *Sulfolobus*

solfataricus P2 show a common N- glycosylation profile. J Proteome Res 12:2779-2790.

38. van Wolferen M, Shajahan A, Heinrich K, Brenzinger S, Black IM, Wagner A, Briegel A, Azadi P, Albers SV. 2020. Species-specific recognition of Sulfolobales mediated by uv-inducible pili and s-layer glycosylation patterns. mBio 11.

39. Ellen AF, Albers SV, Huibers W, Pitcher A, Hobel CFV, Schwarz H, Folea M, Schouten S, Boekema EJ, Poolman B, Driessen AJM. 2009. Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components. Extremophiles 13:67-79.

40. Kish A, Miot J, Lombard C, Guigner JM, Bernard S, Zirah S, Guyot F. 2016. Preservation of archaeal surface layer structure during mineralization. Sci Rep 6:1-10.

41. Veith A, Klingl A, Zolghadr B, Lauber K, Mentele R, Lottspeich F, Rachel R, Albers S-V, Kletzin A. 2009. *Acidianus*, *Sulfolobus* and *Metallosphaera* surface layers: structure, composition and gene expression. Mol Microbiol 73:58-72.

42. Gambelli L, Meyer BH, McLaren M, Sanders K, Quax TEF, Gold VAM, Albers SV, Daum B. 2019. Architecture and modular assembly of *Sulfolobus* S-layers revealed by electron cryotomography. Proc Nat Acad Sci U S A 116:25278-25286.

43. Zhang C, Wipfler RL, Li Y, Wang Z, Hallett EN, Whitaker RJ. 2019. Cell structure changes in the hyperthermophilic crenarchaeon *Sulfolobus islandicus* lacking the S-layer. mBio 10.

44. Kandiba L, Guan Z, Eichler J. 2013. Lipid modification gives rise to two distinct *Haloferax volcanii* S-layer glycoprotein populations. Biochim Biophys Acta Biomembr 1828:938-943.

45. Rodrigues-Oliveira T, Souza AA, Kruger R, Schuster B, De Freitas SM, Kyaw CM. 2019. Environmental factors influence the *Haloferax volcanii* S-layer protein structure. PLoS ONE 14.

46. Phipps BM, Huber R, Baumeister W. 1991. The cell envelope of the hyperthermophilic archaebacterium *Pyrobaculum organotrophum* consists of two regularly arrayed protein layers: three-dimensional structure of the outer layer. Mol Microbiol 5:253-265.

47. Janesch B, Messner P, Schäffer C. 2013. Are the surface layer homology domains essential for cell surface display and glycosylation of the S-layer protein from *Paenibacillus alvei* CCM 2051T? J Bacteriol 195:565-575.

48. Mesnage S, Fontaine T, Mignot T, Delepierre M, Mock M, Fouet A. 2000. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation | The EMBO Journal, p 4473-4484.

49. Zhao G, Ali E, Sakka M, Kimura T, Sakka K. 2006. Binding of S-layer homology modules from *Clostridium thermocellum* SdbA to peptidoglycans. Appl Microbiol Biotechnol 70:464-469.

50. Blackler RJ, López-Guzmán A, Hager FF, Janesch B, Martinz G, Gagnon SML, Haji-Ghassemi O, Kosma P, Messner P, Schäffer C, Evans SV. 2018. Structural

basis of cell wall anchoring by SLH domains in *Paenibacillus alvei*. Nat Commun 9:3120-3120.

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

52. Janesch B, Koerdt A, Messner P, Schäffer C. 2013. The S-layer homology domain-containing protein SlhA from *Paenibacillus alvei* CCM 2051T is important for swarming and biofilm formation. PLoS ONE 8.

53. Huber C, Ilk N, Rünzler D, Egelseer EM, Weigert S, Sleytr UB, Sára M. 2004. The three S-layer-like homology motifs of the S-layer protein SbpA of *Bacillus sphaericus* CCM 2177 are not sufficient for binding to the pyruvylated secondary cell wall polymer. Mol Microbiol 55:197-205.

54. Tarlovsky Y, Fabian M, Solomaha E, Honsa E, Olson JS, Maresso AW. 2010. A *Bacillus anthracis* S-layer homology protein that binds heme and mediates heme delivery to IsdC. J Bacteriol 192:3503-3511.

55. Zhao G, Li H, Wamalwa B, Sakka M, Kimura T, Sakka K. Different binding specificities of S-layer homology modules from *Clostridium thermocellum* AncA, Slp1, and Slp2. Biosci Biotechnol Biochem doi:10.1271/bbb.50699.

56. Kang HJ, Baker EN. 2012. Structure and assembly of Gram-positive bacterial pili: unique covalent polymers. Curr Opin Struct Biol 22:200-207.

57. Proft T, Baker E. 2009. Pili in Gram-negative and Gram-positive bacteria—structure, assembly and their role in disease. Cell Mol Life Sci 66:613-635.

58. Egelman EH. 2017. Cryo-EM of bacterial pili and archaeal flagellar filaments. Curr Opin Struct Biol 46:31-37.

59. Ton-That H, Schneewind O. 2003. Assembly of pili on the surface of *Corynebacterium diphtheriae*. Mol Microbiol 50:1429-1438.

60. Pansegrau W, Bagnoli F. 2017. Pilus assembly in gram-positive bacteria, p 203-233, vol 404. Springer Verlag.

61. Swierczynski A, Ton-That H. 2006. Type III pilus of corynebacteria: Pilus length is determined by the level of its major pilin subunit. J Bacteriol 188:6318-6325.

62. Ton-That H, Marraffini LA, Schneewind O. 2004. Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. Mol Microbiol 53:251-261.

63. Ton-That H, Schneewind O. 2003. Assembly of pili on the surface of *Corynebacterium diphtheriae*. Mol Microbiol 50:1429-1438.

64. Hospenthal MK, Costa TR, Waksman G. 2017. A comprehensive guide to pilus biogenesis in Gram-negative bacteria. Nat Rev Microbiol 15:365-379.

65. Geibel S, Waksman G. 2014. The molecular dissection of the chaperone–usher pathway. Biochim Biophys Acta - Mol Cell Res 1843:1559-1567.

66. Werneburg GT, Thanassi DG. 2018. Pili Assembled by the chaperone/usher pathway in *Escherichia coli* and *Salmonella*. EcoSal Plus 8.

67. Craig L, Li J. 2008. Type IV pili: paradoxes in form and function. Curr Opin Struct Biol 18:267-277.

68. Bhoite S, van Gerven N, Chapman MR, Remaut H. 2019. Curli biogenesis: bacterial amyloid assembly by the type VIII secretion pathway. EcoSal Plus 8.

69. Evans ML, Chapman MR. 2014. Curli biogenesis: order out of disorder. Biochim Biophys Acta - Mol Cell Res 1843:1551-1558.

70. Klein RD, Shu Q, Cusumano ZT, Nagamatsu K, Gualberto NC, Lynch AJL, Wu C, Wang W, Jain N, Pinkner JS, Amarasinghe GK, Hultgren SJ, Frieden C, Chapman MR. 2018. Structure-function analysis of the curli accessory protein CsgE defines surfaces essential for coordinating amyloid fiber formation. mBio 9.

71. Shibata S, Shoji M, Okada K, Matsunami H, Matthews MM, Imada K, Nakayama K, Wolf M. 2020. Structure of polymerized type V pilin reveals assembly mechanism involving protease-mediated strand exchange. Nat Microbiol doi:10.1038/s41564-020-0705-1:1-8.

72. Burrows LL. 2012. *Pseudomonas aeruginosa* twitching motility: type IV pili in action. Ann Rev Microbiol 66:493-520.

73. Melville S, Craig L. 2013. Type IV pili in Gram-positive bacteria. Microbiol Mol Biol Rev 77:323-341.

74. Albers S-V, Pohlschröder M. 2009. Diversity of archaeal type IV pilin-like structures. Extremophiles 13:403-410.

75. Karaolis DK, Somara S, Maneval DR, Johnson JA, Kaper JB. 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature 399:375-379.

76. Varga JJ, Nguyen V, O'Brien DK, Rodgers K, Walker RA, Melville SB. 2006. Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other *Clostridia*. Mol Microbiol 62:680-694.

77. Piepenbrink KH, Maldarelli GA, De La Peña CFM, Dingle TC, Mulvey GL, Lee A, Von Rosenvinge E, Armstrong GD, Donnenberg MS, Sundberg EJ. 2015. Structural and evolutionary analyses show unique stabilization strategies in the type IV pili of *Clostridium difficile*. Structure 23:385-396.

78. Albers SV, Pohlschröder M. 2009. Diversity of archaeal type IV pilin-like structures, vol 13, p 403-410. Springer.

79. Hospenthal MK, Costa TRD, Waksman G. 2017. A comprehensive guide to pilus biogenesis in Gram-negative bacteria, vol 15, p 365-379. Nature Publishing Group.

80. Averhoff B, Friedrich A. 2003. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. Arch Microbiol 180:385-393.

81. Piepenbrink KH, Maldarelli GA, De La Peña CFM, Mulvey GL, Snyder GA, De Masi L, Von Rosenvinge EC, Günther S, Armstrong GD, Donnenberg MS, Sundberg EJ. 2014. Structure of *Clostridium difficile* pilj exhibits unprecedented divergence from known type iv pilins. J Biol Chem 289:4334-4345.

82. Khan AMAM, Hauk VJ, Ibrahim M, Raffel TR, Blumer-Schuette SE. 2020. *Caldicellulosiruptor bescii* adheres to polysaccharides via a type IV pilin-dependent mechanism. Appl Environ Microbiol 86.

83. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL, Adams MWW, Hamilton-Brehm SD, Elkins JG, Larimer FW, Land ML, Hauser LJ, Cottingham RW, Hettich RL, Kelly RM. 2012. *Caldicellulosiruptor* core and pangenomes reveal determinants for

noncellulosomal thermophilic deconstruction of plant biomass. J Bacteriol 194:4015-4028.

84. Chang YW, Rettberg LA, Treuner-Lange A, Iwasa J, Søgaard-Andersen L, Jensen GJ. 2016. Architecture of the type IV pilus machine. Science 351.

85. Pohlschroder M, Esquivel RN. 2015. Archaeal type IV pili and their involvement in biofilm formation. Front Microbiol 6:190.

86. Wang YA, Yu X, Ng SYM, Jarrell KF, Egelman EH. 2008. The structure of an archaeal pilus. J Mol Biol 381:456-466.

87. Ng SY, Chaban B, Jarrell KF. 2006. Archaeal flagella, bacterial flagella and type IV pili: a comparison of genes and posttranslational modifications. Microbial Physiology 11:167-191.

88. Peabody CR, Chung YJ, Yen M-R, Vidal-Ingigliardi D, Pugsley AP, Saier Jr MH. 2003. Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. Microbiology 149:3051-3072.

89. Braun T, Vos MR, Kalisman N, Sherman NE, Rachel R, Wirth R, Schröder GF, Egelman EH. 2016. Archaeal flagellin combines a bacterial type IV pilin domain with an Ig-like domain. Proc Nat Acad Sci U S A 113:10352-10357.

90. Banerjee A, Neiner T, Tripp P, Albers SV. 2013. Insights into subunit interactions in the *Sulfolobus acidocaldarius* archaellum cytoplasmic complex. FEBS J 280:6141-6149.

91. Banerjee A, Tsai C-L, Chaudhury P, Tripp P, Arvai AS, Ishida JP, Tainer JA, Albers S-V. 2015. FlaF is a β-sandwich protein that anchors the archaellum in the archaeal cell envelope by binding the S-layer protein. Structure 23:863-872.

92. Bardy SL, Ng SY, Jarrell KF. 2004. Recent advances in the structure and assembly of the archaeal flagellum. Microb Physiol 7:41-51.

93. Namba K, Yamashita I, Vonderviszt F. 1989. Structure of the core and central channel of bacterial flagella. Nature 342:648-654.

94. O'Brien E, Bennett PM. 1972. Structure of straight flagella from a mutant Salmonella. J Mol Biol 70:133-152.

95. Nakamura S, Minamino T. 2019. Flagella-driven motility of bacteria. Biomolecules 9.

96. Subramanian S, Kearns DB. 2019. Functional regulators of bacterial flagella. Ann Rev Microbiol 73:225-246.

97. Khan S, Scholey JM. 2018. Assembly, functions and evolution of archaella, flagella and cilia. Curr Biol 28:R278-R292.

98. Green ER, Mecsas J. 2016. Bacterial secretion systems: an overview. Microbiol Spectr 4.

99. Freudl R. 2018. Signal peptides for recombinant protein secretion in bacterial expression systems. Microb Cell Fact 17:1-10.

100. Tsirigotaki A, De Geyter J, Economou A, Karamanou S. 2017. Protein export through the bacterial Sec pathway. Nat Rev Microbiol 15:21-36.

101. Jamshad M, Knowles TJ, White SA, Ward DG, Mohammed F, Rahman KF, Wynne M, Hughes GW, Kramer G, Bukau B, Huber D. 2019. The C-terminal tail of the bacterial translocation ATPase SecA modulates its activity. eLife 8.

102. Ma C, Wu X, Sun D, Park E, Catipovic MA, Rapoport TA, Gao N, Li L. 2019. Structure of the substrate-engaged SecA-SecY protein translocation machine. Nat Commun 10.

103. Freudl R. 2018. Signal peptides for recombinant protein secretion in bacterial expression systems, vol 17. BioMed Central Ltd.

104. Komarudin AG, Driessen AJM. 2019. SecA-mediated protein translocation through the SecYEG channel. Microbiol Spectr 7.

105. Mori H, Ito K. 2001. The Sec protein-translocation pathway. Trends Microbiol 9:494-500.

106. Müller M, Bernd Klösgen R. 2005. The Tat pathway in bacteria and chloroplasts. Mol Membr Biol 22:113-121.

107. Palmer T, Sargent F, Berks BC. 2005. Export of complex cofactor-containing proteins by the bacterial Tat pathway. Trends Microbiol 13:175-180.

108. Ize B, Stanley NR, Buchanan G, Palmer T. 2003. Role of the *Escherichia coli* Tat pathway in outer membrane integrity. Mol Microbiol 48:1183-1193.

109. Patel R, Smith SM, Robinson C. 2014. Protein transport by the bacterial Tat pathway. Biochim Biophys Acta Mol Cell Res 1843:1620-1628.

110. Thomas JD, Daniel RA, Errington J, Robinson C. 2001. Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*. Mol Microbiol 39:47-53.

111. Frain KM, van Dijl JM, Robinson C. 2019. The twin-arginine pathway for protein secretion. EcoSal Plus 8.

112. Ulfig A, Freudl R. 2018. The early mature part of bacterial twin-arginine translocation (Tat) precursor proteins contributes to TatBC receptor binding. J Biol Chem 293:7281-7299.

113. Frain KM, Robinson C, van Dijl JM. 2019. Transport of folded proteins by the tat system. Protein J 38:377-388.

114. Liu L, Huang W-C, Liu Y, Li M. 2021. Diversity of cellulolytic microorganisms and microbial cellulases. Int Biodeterior Biodegradation 163:105277.

115. Sahoo K, Sahoo RK, Gaur M, Subudhi E. 2020. Cellulolytic thermophilic microorganisms in white biotechnology: a review. Folia Microbiol 65:25-43.

116. Burns RG. 2010. How do microbial extracellular enzymes locate and degrade natural and synthetic polymers in soil, p 294-297 doi:10.1007/978-3-642-05297-2_85. Springer Berlin Heidelberg.

117. Fenice M, Gallo AM, Juarez-Jimenez B, Gonzalez-Lopez J. 2007. Screening for extracellular enzyme activities by bacteria isolated from samples collected in the Tyrrhenian Sea. Ann Microbiol 57:93-99.

118. Li Y, Sun L-L, Sun Y-Y, Cha Q-Q, Li C-Y, Zhao D-L, Song X-Y, Wang M, McMinn A, Chen X-L, Zhang Y-Z, Qin Q-L. 2019. Extracellular enzyme activity and its implications for organic matter cycling in northern Chinese marginal seas. Front Microbiol 10:2137-2137.

119. Loperena L, Soria V, Varela H, Lupo S, Bergalli A, Guigou M, Pellegrino A, Bernardo A, Calviño A, Rivas F, Batista S. 2012. Extracellular enzymes produced by microorganisms isolated from maritime Antarctica. World J Microbiol Biotechnol 28:2249-2256.

120. Pohlon E, Fandino AO, Marxsen J. 2013. Bacterial community composition and extracellular enzyme activity in temperate streambed sediment during drying and rewetting. PLoS ONE 8.

121. Wilson JW, Schurr MJ, LeBlanc CL, Ramamurthy R, Buchanan KL, Nickerson CA. 2002. Mechanisms of bacterial pathogenicity. Postgrad Med Jl 78:216-224.

122. Traving SJ, Thygesen UH, Riemann L, Stedmon CA. 2015. A model of extracellular enzymes in free-living microbes: which strategy pays off? Applied and environmental microbiology 81:7385-7393.

123. Chróst RJ. 1990. Microbial ectoenzymes in aquatic environments, p 47-78, Aquatic microbial ecology. Springer.

124. Traving SJ, Thygesen UH, Riemann L, Stedmon CA. 2015. A model of extracellular enzymes in free-living microbes: Which strategy pays off? Appl Environ Microbiol 81:7385-7393.

125. Chen C-C, Dai L, Ma L, Guo R-T. 2020. Enzymatic degradation of plant biomass and synthetic polymers. Nat Rev Chem 4:114-126.

126. Aehle W. 2007. Enzymes in industry: production and applications. John Wiley & Sons.

127. Liu X, Kokare C. 2017. Microbial enzymes of use in industry, p 267-298, Biotechnology of microbial enzymes. Elsevier.

128. Thapa S, Li H, OHair J, Bhatti S, Chen F-C, Nasr KA, Johnson T, Zhou S. 2019. Biochemical characteristics of microbial enzymes and their significance from industrial perspectives. Mol Biotechnol 61:579-601.

129. Bajaj P, Mahajan R. 2019. Cellulase and xylanase synergism in industrial biotechnology. Appl Microbiol Biotechnol 103:8711-8724.

130. Banerjee G, Scott-Craig JS, Walton JD. 2010. Improving enzymes for biomass conversion: a basic research perspective. Bioenergy Res 3:82-92.

131. Dhiman SS, Sharma J, Battan B. 2008. Industrial applications and future prospects of microbial xylanases: a review. BioResources 3:1377-1402.

132. Himmel ME, Xu Q, Luo Y, Ding S-y, Lamed R, Bayer EA. 2010. Microbial enzyme systems for biomass conversion: emerging paradigms. Biofuels 1:323-341.

133. Singh A, Kuhad RC, Ward OP. 2007. Industrial application of microbial cellulases.

134. Hendriks A, Zeeman G. 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresour Technol 100:10-18.

135. Limayem A, Ricke SC. 2012. Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects. Prog Energy Combust Sci 38:449-467.

136. Fan Z. 2014. Consolidated bioprocessing for ethanol production, p 141-160, Biorefineries. Elsevier.

137. Bajaj P, Mahajan R. 2019. Cellulase and xylanase synergism in industrial biotechnology, vol 103, p 8711-8724. Springer Verlag.

138. Robak K, Balcerek M. 2018. Review of second generation bioethanol production from residual biomass. Food Technol Biotechnol 56:174.

139. Raveendran S, Parameswaran B, Ummalyma SB, Abraham A, Mathew AK, Madhavan A, Rebello S, Pandey A. 2018. Applications of microbial enzymes in food industry. Food Technol Biotechnol 56:16.

140. Choudhury P, Bhunia B. 2015. Industrial application of lipase: a review. Biopharm J 1:41-47.

141. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 50:D571-D577.

142. Han Y, Agarwal V, Dodds D, Kims J, Bae B, Mackie RI, Nair SK, Cann IKO. 2012. Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. J Biol Chem 287:34946-34960.

143. Lee S-P, Morikawa M, Takagi M, Imanaka T. 1994. Cloning of the aapT gene and characterization of its product, alpha-amylase-pullulanase (AapT), from thermophilic and alkaliphilic *Bacillus* sp. strain XAL601. Appl Environ Microbiol 60:3764-3773.

144. Cheng YM, Hong TY, Liu CC, Meng M. 2009. Cloning and functional characterization of a complex endo-β-1,3- glucanase from *Paenibacillus* sp. Appl Microbiol Biotechnol 81:1051-1061.

145. Fukuda M, Watanabe S, Yoshida S, Itoh H, Itoh Y, Kamio Y, Kaneko J. 2010. Cell surface xylanases of the glycoside hydrolase family 10 are essential for xylan utilization by *Paenibacillus* sp. W-61 as generators of xylo-oligosaccharide inducers for the xylanase genes. J Bacteriol 192:2210-2219.

146. Ito Y, Tomita T, Roy N, Nakano A, Sugawara-Tomita N, Watanabe S, Okai N, Abe N, Kamio Y. 2003. Cloning, expression, and cell surface localization of *Paenibacillus* sp. strain W-61 xylanase 5, a multidomain xylanase. Appl Environ Microbiol 69:6969-6978.

147. Itoh T, Sugimoto I, Hibi T, Suzuki F, Matsuo K, Fujii Y, Taketo A, Kimoto H. 2014. Overexpression, purification, and characterization of *Paenibacillus* cell surface-expressed chitinase ChiW with two catalytic domains. Biosci Biotechnol Biochem 78:624-634.

148. Dassa B, Borovok I, Lamed R, Henrissat B, Coutinho P, Hemme CL, Huang Y, Zhou J, Bayer EA. 2012. Genome-wide analysis of Acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. BMC Genom 13:1-13.

149. Shin E-S, Yang M-J, Jung KH, Kwon E-J, Jung JS, Park SK, Kim J, Yun HD, Kim H. 2002. Influence of the transposition of the thermostabilizing domain of *Clostridium thermocellum* xylanase (XynX) on xylan binding and thermostabilization. Appl Environ Microbiol 68:3496-3501.

150. Kim H, Jung KH, Pack MY. 2000. Molecular characterization of xynX, a gene encoding a multidomain xylanase with a thermostabilizing domain from *Clostridium thermocellum*. Appl Microbiol Biotechnol 54:521-527.

151. Fuchs K-P, Zverlov VV, Velikodvorskaya GA, Lottspeich F, Schwarz WH. 2003. Lic16A of *Clostridium thermocellum*, a non-cellulosomal, highly complex endo-β-1, 3-glucanase bound to the outer cell surface. Microbiology 149:1021-1031.

152. Ali MK, Fukumura M, Sakano K, Karita S. 1999. Cloning, sequencing, and expression of the gene encoding the *Clostridium stercorarium* xylanase C in *Escherichia coli*. Biosci Biotechnol Biochem 63:1596-1604.

153. Ali MK, Kimura T, Sakka K, Ohmiya K. 2001. The multidomain xylanase Xyn10B as a cellulose-binding protein in *Clostridium stercorarium*. FEMS Microbiol Lett 198:79-83.

154. Jia-Xun F, Shuichi K, Emi F, Tsuchiyoshi F, Tetsuya K, Kazuo S, Kunio O. 2000. Cloning, sequencing, and expression of the gene encoding a cell-bound multi-domain xylanase from *Clostridium josui*, and characterization of the translated product. Biosci Biotechnol Biochem 64:2614-2624.

155. Huang X, Li Z, Du C, Wang J, Li S. 2015. Improved expression and characterization of a multidomain xylanase from *Thermoanaerobacterium aotearoense* SCUT27 in *Bacillus subtilis*. J Agri Food Chem 63:6430-6439.

156. Hung KS, Liu SM, Fang TY, Tzou WS, Lin FP, Sun KH, Tang SJ. 2011. Characterization of a salt-tolerant xylanase from *Thermoanaerobacterium saccharolyticum* NTOU1. Biotechnol Lett 33:1441-1447.

157. Liu S-Y, Gherardini FC, Matuschek M, Bahl H, Wiegel J. 1996. Cloning, sequencing, and expression of the gene encoding a large S-layer-associated endoxylanase from *Thermoanaerobacterium* sp. strain JW/SL-YS 485 in *Escherichia coli*. J Bacteriol 178:1539-1547.

158. Matuschek M, Burchhardt 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.

159. Shao W, Deblois S, Wiegel J. 1995. A high-molecular-weight, cell-associated xylanase isolated from exponentially growing *Thermoanaerobacterium* sp. strain JW/SL-YS485. Appl Environ Microbiol 61:937-940.

160. Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, Zurawski JV, Lee LL, Blumer-Schuette SE, Kelly RM. 2016. Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. J Biol Chem 291:6732-6747.

CHAPTER 2

Enzymes from extremely thermophilic bacteria and archaea: current status and future prospects

Tunyaboon Laemthong[#], April M. Lewis[#], James R. Crosby, Ryan G. Bing, William H. Schneider, Daniel J. Willard, James A. Counts, and Robert M. Kelly*

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

[#]Contributed equally

Published in: *Extremozymes and their industrial applications* (Mishra, J., Arora, N., Agnihotri, S. ed.), UK. pp 41-84 (2022)

Abstract

Extremely thermophilic bacteria and archaea ($T_{opt} \geq 70°C$) inhabit thermal biotopes that are distributed globally in both terrestrial and marine settings. As a group, these microorganisms utilize a very wide array of substrates to support their growth, ranging from lignocellulose to CO_2. The enzymes produced by these microorganisms to process their energy and carbon sources are not only intrinsically thermostable and thermoactive, but they have novel characteristics that go beyond their unusually high functional temperature range. There are many potential applications of these enzymes as biocatalysts for biotechnology that go beyond the use of thermophilic DNA polymerases for the Polymerase Chain Reaction (PCR). Here, we will consider the basic features of enzymes from extreme thermophiles with an eye towards their broader use as industrial biocatalysts as well as components in metabolic engineering pathways. Genome sequencing and availability of molecular genetic tools have impacted extreme thermophile biotechnology and will play important roles in future efforts to deploy and utilize the biocatalytic potential of these interesting enzymes.

Introduction

The prospect for application of thermostable and thermoactive biocatalysts came immediately to mind upon the discovery and isolation of extremely thermophilic bacteria and archaea ($T_{opt} \geq 70°C$) isolated from terrestrial and marine biotopes more than a half century ago. While mesophilic enzymes had long been considered as an alternative to chemical catalysts for a range of industrially important transformations [1], their relatively narrow thermal range was problematic for the processing conditions best suited for their intended application. Many enzymes from extreme thermophiles have temperature optima approaching and exceeding 100°C [2], as well as long half-lives under otherwise denaturing conditions. These performance characteristics suited the thermal cycling needs for the Polymerase Chain Reaction (PCR) [3], but also had potential large-scale use in applications such as starch processing [4], glucose isomerization [5], and hydraulic fracturing (fracking) [6]. In the early 1990s, the need to produce native versions of these enzymes (a major impediment to their industrial use) gave way to recombinant overexpression in mesophilic hosts [7], which was further enabled by the availability of genome sequences for extreme thermophiles (e.g., *Thermotoga maritima* [8], *Pyrococcus furiosus* [9]). Emerging techniques with directed evolution [10] enabled the development of biocatalysts customized for specific applications (e.g., amylase for starch processing [11]). Now, enzymes from extreme thermophiles can be considered as components of the biocatalytic toolbox for a range of commercial uses.

The ever-expanding database of genome and metagenome sequences for extreme thermophiles has opened up prospects for finding thermophilic versions of industrially significant enzymes where strategic use of high temperatures is crucial or advantageous. More challenging, but potentially more significant, is the identification of enzymes from these microorganisms that

can carry out novel biotransformations, either as a single step or as part of a metabolically engineered pathway. The creation of these new biotransformations is enabled by the genetic systems developed for several extreme thermophiles.

There have been several recent reviews covering the industrial potential of enzymes from extreme thermophiles [12-19]. Here, we provide additional perspective on the potential uses of these biocatalysts.

Microbial Diversity at High Temperatures

While naturally inhospitable to most living systems, high temperature environments teem with microbial life, an observation dating back to the early 1900's [20], and are represented significantly in phylogenetic analyses (see **Figure 2.1**). The term 'thermophile' is often used to describe microorganisms capable of growing above mesophilic (20–40˚C) temperatures. This lower thermal boundary results in the broad application of this term to encompass an incredibly diverse set of organisms spanning the tree of life. However, at extremely high temperatures (≥ 70˚C), many organisms, coming almost exclusively from the prokaryotic domains, thrive. In fact, there are many rationales for the diminished presence of eukaryotic organisms at extremely high temperatures that are primarily physiological, related to the function of organelles; to this point, a temperature limit of ~60˚C [21]. There is recent evidence that some protists and phototrophs may survive at higher temperatures (near 70˚C) [22, 23].

Extremely Thermophilic Bacteria

Within the Bacteria, many lineages contain moderately thermophilic representatives (T_{opt} 40-70 °C); however, only a few phyla contain extreme thermophiles (e.g., *Aquificae, Firmicutes, Deinococcus-Thermus,* and *Thermotogae*). The phylum *Aquificae* contains several families of extreme thermophiles found either in terrestrial or marine hydrothermal features [24]. These include some of the most thermophilic species within the Bacteria: "*Aquifex aeolicus"* (T_{opt} ~ 85-90˚C) and *Aquifex pyrophilus* (T_{opt} ~ 85˚C) [25]. While some evidence exists for limited amounts of organoheterotrophy and facultative chemolithotrophy in the *Aquificiae,* most of the phylum is composed of obligate chemolithoautotrophs [24]. Of interest for biotechnological applications is their capacity for sulfur and hydrogen oxidation [26], particularly within terrestrial environments, to power autotrophy via the reductive TCA cycle that is widespread in the phylum [27].

In contrast, many of the other extremely thermophilic bacterial lineages are composed of organoheterotrophs. In the case of the lineages of Thermotogae and Clostridia, two groups of anaerobic, plant biomass-degrading microorganisms come from the genera *Caldicellulosiruptor* (T_{opt} 70-78˚C) and *Thermotoga* (T_{opt} 65-80˚C) [28]. In the case of the former, a lack of carbon catabolite repression, and highly efficient cellulases and other glucanases geared toward a broad array of polysaccharides and specifically lignocellulosic substrates [29, 30], have piqued biotechnological interest. Meanwhile, species in the genus *Thermotoga* are largely marine anaerobes that also produce a variety of glycosidic enzymes (but not true cellulases) [31-33]; the *Thermotogae* have a "toga"-like extracellular sheath, which can contain singular or multiple *Thermotogae* cells. Additionally, there are other members of the order *Thermotogales* that are also extremely thermophilic belonging to the genera *Fervidobacterium* (terrestrial) and *Thermosipho* (marine) [34]. Another bacterial lineage of interest is the species *Thermus* belonging to the

Deinoccoccus-Thermus phylum. These are among the most well studied thermophilic organisms, owing to the facile genetics of *Thermus thermophilus* HB27 [35], and its robust growth under aerobic conditions and with organoheterotrophic media [36]. In fact, this organism has been the backbone of protein structure studies, while another member, *Thermus aquaticus*, served as the source of the earliest polymerase for PCR-based technologies [37, 38].

Extremely Thermophilic Archaea

In contrast, numerous archaeal extreme thermophiles can be found within the phyla *Crenarchaea* and *Euryarchaea*. In fact, the former contains some of the most thermophilic microorganisms discovered to date, such as the Crenarchaeotes *Pyrolobus fumarii* (T_{max} 113˚C, T_{opt} 106˚C) [39] and "*Geogemma barossii*" (T_{max} 121 ˚C) [40]. Meanwhile, the *Euryarchaea* include a number of extremely thermophilic organisms, such as the marine organism *Pyrococcus furiosus* (T_{opt} 100˚C) [41] and the methanogen *Methanopyrus kandleri* (T_{opt} 98 ˚C) [42].

Within the phylum *Euryarchaea* are a number of thermophilic lineages. However, the extremely thermophilic microorganisms studied to date are largely from the organoheterotrophic *Thermococcaceae*, the chemolithoautotrophic/heterotrophic *Archaeoglobi*, and the methanotrophic *Methanopyraceae, Methanococcaceae,* and *Methanothermaceae*. In the case of the *Thermococcaceae,* there is a single family composed of three genera (*Pyrococcus*, *Thermococcus*, and *Palaeococcus*), all originating from marine environments and subsisting on sugars and peptides as carbon sources, with many utilizing sulfur or hydrogen as electron acceptors [43]. The two most studied organisms in this family are *P. furiosus* and *Thermococcus kodakarensis*, driven by the availability of genetics systems for these two archaea [44, 45]. The

organisms also contain novel membrane respiratory complexes that can couple fermentation to the production of hydrogen and hydrogen sulfide gases [46, 47].

Unlike the previous *Euryarchaea* lineages, the *Archaeglobi* (containing *Archaeoglobus, Ferroglobus,* and *Geoglobus*) grow at much lower temperatures (typically 70-88˚C), and demonstrate remarkable metabolic diversity [48]. While most of these organisms are chemolithoautotrophs, *Archaeoglobus profundusi* and *Archaeoglobus infectus* are capable of utilizing organic carbon sources [49]. The final *Euryarchaeal* lineage of interest is the thermophilic methanogens, including the two families, *Methanopyraceae* and *Methanothermaceae*, that contain just one and two named species, respectively. However, the family *Methanocaldococcaceae* contains numerous named species, all of which utilize hydrogen and carbon dioxide to produce methane [50]. In fact, this degree of metabolic specialization relates to their syntrophic relationships with fermentative organoheterotrophic extreme thermophiles producing carbon dioxide and hydrogen, such as those from *Thermococcaceae* [51] and *Thermotogae* (e.g. *Thermotoga maritima* [52]).

The broadest group of extremely thermophilic organisms yet identified are those within the class *Thermoprotei* (within *Crenarchaea*). The class contains four orders (*Acidolobales, Desulfurococcales, Sulfolobales,* and *Thermoproteales*) and 6 separate families with extremely thermophilic microorganisms (*Acidolobaceae, Desulfurococcaceae, Pyroditiaceae, Sulfolobaceae, Thermofilaceae,* and *Thermoproteaceae*). As previously mentioned, some of the most thermophilic organisms known to date originate from the order *Desulfurococcales*. Within its two families, *Desulfurococaeae* and *Pyrodictiaceae,* are several lineages of both terrestrial and marine organisms with varied modes of life that range from chemolithotrophy (*Ignicoccus spp.*) to heterotrophy and mixotrophy, under mostly neutrophilic and, in some cases, slightly acidic

conditions [53]. Furthermore, the phylum also contains the largest number of known extreme thermoacidophiles (T_{opt} > 70˚C; pH_{opt} < 4.5), with representatives in each of the other orders. For example, the *Acidilobales* contains just three named species, *Acidilobus aceticus, Acidilobus sacchaarovorans* and *Caldisphaera laguensis*. These are all anaerobic organoheterotrophs, capable of using sulfur as an electron acceptor, but grow at pH between 3.5-4.0 and temperatures between 70-85˚C [54, 55]. Another order of interest is the *Thermoproteales*, containing the families *Thermoproteaceae* and *Thermofilaceae*, the former containing five named genera, while the latter contains just one named species [56]. Most of the organisms within this order are strictly anaerobic and capable of growth on complex organic media. They can also use oxidized sulfur species to produce hydrogen sulfide, as well as nitrates, and in a few cases exhibit facultative aerobic growth with oxygen as an electron acceptor.

Another order of broad interest is the extremely thermoacidophilic *Sulfolobales.* These organisms were first identified in sulfur springs of Yellowstone, hence the names *Sulfolobales* (order), *Sulfolobaceae* (family), and *Sulfolobus* (genus) [57]. They are also one of the best studied orders of archaea, with eight named genera of organisms ranging from the facultative aerobic chemolithoautotrophs of the genera *Acidianus* [58] (also containing the most acidophilic *Sulfolobales* member to-date *Acidianus sulfidivorans* (pH_{opt} 0.8) [59]) and *Sulfurisphaera* [60, 61], to mostly organoheterotrophic aerobes from the genera *Saccharolobus* [62] and *Sulfolobus*. The order also contains the only known strict anaerobe *Stygiolobus azoricus [63]* and the oddly sulfur-inhibited *Sulfodiicoccus acidiphilus* [64]. There are several organisms from this group that serve as models for the various chemolithoautotrophic metabolisms, including sulfur reduction, oxidation, and disproportionation in *Acidianus ambivalens* [65], iron biooxidation in *Sulfuracidifex metallicus* and *Metallosphaera spp.* [66-68], and autotrophy through the 3-

hydroxypropionate/4-hydroxybutyrate cycle in *Metallosphaera sedula*; this cycle is nearly universally conserved within the order [69, 70]. In addition, numerous genomes have been sequenced for species within this order [70]. Genetic systems have been developed for organoheterotrophic species, such as *Saccharolobus islandicus* (f. *Sulfolobus islandicus*) [71, 72], *Saccharolobus solfataricus* (f. *Sulfolobus solfataricus*) [73], and *Sulfolobus acidocaldarius* [74], the availability of which has facilitated rapid expansion of investigations into the features of these thermoacidophiles.

DNA Polymerases

The fields of molecular genetics and genomics were made possible by thermostable DNA polymerases used for the Polymerase Chain Reaction (PCR) [3]. These DNA polymerases made genetic engineering possible, which allowed for the molecular study of genes and their associated phenotypes. The original PCR polymerase was isolated from the thermophilic bacterium *Thermus aquaticus* [38]. However, the value of this thermally stable polymerase was not fully realized until 1985 when the invention of the temperature cycling polymerase chain reaction (PCR) to amplify DNA was reported [75]. The use of the thermostable family A *Taq* polymerase increased the temperature at which PCR could be performed, thereby improving the specificity and output of the reaction [3]. DNA polymerases from extreme thermophiles have not only revolutionized how researchers approach scientific problems in the life sciences, they are also important for medical applications. For example, PCR is used to screen individuals as carriers for methicillin-resistant *Staphylococcus aureus* [76].

Though the widely used *Taq* polymerase is thermally stable with a reported optimal temperature of 80°C, it lacks the 3’ to 5’ exonuclease activity (proofreading capability) to identify

and correct mistakes in PCR products [77, 78]. Furthermore, *Taq* polymerase has low processivity (the ability to incorporate nucleotides before dissociating from the DNA), rendering it unable to generate larger segments of DNA. Therefore, *Taq* polymerase makes error-prone PCR products of only a moderate length. Despite these drawbacks, *Taq* polymerase is still in use today in applications, such as colony PCR screening, where producing accurate sequences is unnecessary. Polymerases from other thermophiles have emerged for PCR applications, such as the family B polymerase from *Pyrococcus furiosus (Pfu),* to increase fidelity. Due to its 3' to 5' exonuclease proofreading activity, *Pfu* polymerase is a high-fidelity polymerase with a native fidelity 30-fold higher than *Taq* [79]. This 3' to 5' exonuclease domain allows proofreading polymerases to recognize nucleotide mismatches and excise the incorrect base from the primer extension, unlike nonproofreading polymerases like *Taq* that often dissociate from the DNA when a misincorporated base is encountered [80] **(Figure 2.2**). Although *Pfu* polymerase can amplify DNA with a much higher accuracy, it lacks the processivity to make longer DNA segments.

Advances in PCR technology have made use of thermophilic DNA binding proteins to create chimeric polymerases with improved processivity. For instance, fusion of the non-specific dsDNA binding protein Sso7d from the thermoacidophile *Saccharolobus solfataricus* (f. *Sulfolobus solfataricus*) to the *Taq* polymerase improved the polymerase's processivity and increased the average primer extension length ~ 5-fold. Similarly, fusion of the Sso7d dsDNA binding protein to the C-terminus of the *Pfu* polymerase significantly increased processivity and resulted in an average primer extension length ~ 9-fold higher in the chimeric polymerase. Furthermore, Sso7d fusion polymerases have also demonstrated a broader salt tolerance than the associated wild-type polymerase [81].

As cloning practices have moved away from the traditional method of using restriction enzymes toward PCR-based cloning methods, such as Gibson assembly, splicing by overlap extension, or Golden Gate cloning, the need for high-fidelity polymerases with high processivity is in increasingly high demand. When accurate long sequences need to be generated, such as for vector building or site-directed mutagenesis, high-fidelity designer polymerase are used. The commercially available polymerases Phusion® (ThermoFisher Scientific Catalog # F530S/L) and Q5® (New England BioLabs Catalog # M0491S/L)) are widely used high-fidelity polymerase for cloning applications. Both Phusion® and Q5® are chimeric polymerases fused to the Sso7d dsDNA binding protein; Phusion® is a chimera of the *Pfu* polymerase, and Q5 is a fusion with a novel polymerase. Unlike *Taq* polymerase that generates PCR products with single-base 3' overhangs, the Phusion® and Q5® polymerases each generate blunt end PCR products. Additionally, the fidelity of Phusion® is 39-fold higher than that of the *Taq* polymerase [79], making it ideal for the creation of accurate DNA segments. However, for longer products, the Q5® high-fidelity polymerase is often used, and has been reported to be 280-fold less error-prone than the *Taq* polymerase [79]. Protein engineering efforts will continue to focus on improving the fidelity and processivity of DNA polymerases, aided by the discovery of new thermostable enzymes isolated from extreme thermophiles.

Proteases

Proteases are a major part of the enzyme market with applications not only in scientific research (removal of protein from DNA preparations), but also in the food and the laundry industries [13, 82, 83]. Extremely thermophilic proteases for the heat-stable degradation of proteins are potentially useful biotechnological tools. Notable among the many thermophilic

proteases studied to date is the heteromultimeric proteasome. The archaeal proteasome consists of a 19S proteasome-activating nucleotidase (PAN) and a 20S proteolytic core particle. The 20S core particle is composed of four heptameric rings, two β subunit rings sandwiched between two α subunit rings, that form an internal channel where proteolysis occurs. Proteins destined for degradation by the proteasome are first urmylated [84]. Marked proteins are then unfolded by PAN, which is an AAA^+ ATPase regulatory particle that associates with the outer α subunit rings of the core particle. The unfolded protein is then degraded by the β subunit rings of the core particle [85, 86].

In addition to the proteasome, extreme thermophiles have a range of other important thermostable proteases, such as the *Pyrococcus furiosus* protease I (PfpI) that was originally characterized in 1990 as a multi-subunit serine protease [87-89]. Crystal structures of PfpI homologs in *Pyrococcus horikoshii* [90] and *Thermococcus thioreducens* [91] revealed that this enzyme is actually a cysteine protease with a cysteine residue as major component of the active site. The PfpI protease, like the proteasome, is made up of multiple hexameric ring structures with an active site in an internal channel [91], and will even cross-react with bovine proteasome antibodies [92]. The PfpI protease has been classified as a member of the DJ-1/ThiJ/PfpI superfamily [93], which also contains the medically relevant human DJ-1 protease, whose mutation or altered expression is tied to several human diseases, including Parkinson's [94, 95], tumors [96], and infertility [97].

Serine is the major catalytic residue in the active site of the subtilisin-like serine proteases of *Thermococcus kodakarensis*. In fact, *T. kodakarensis* has three subtilisin-like serine proteases encoded in its genome, Tk-SP (Tk-1689), Tk-0076, and Tk-subtilisin [98]. Subtilisin proteases are first produced in a precursor pre-pro-subtilisin immature form that is composed of a signal peptide

for protein secretion, a pro-peptide, and the mature subtilisin domain. The pro-peptide is auto-processed and denatured after protein secretion to produce the mature active subtilisin. Bacterial pro-peptides typically function as both a subtilisin inhibitor and a chaperone for proper folding of the protease. However, the Tk-subtilisin protease only requires Ca^{2+} for proper folding, and Tk-SP does not require its pro-peptide sequence nor Ca^{2+} to form an active subtilisin protease [99]. Subtilisin-like proteases are also found in the other extreme thermophiles, such as *Thermus thermophilus* [100] and *P. furiosus* [101, 102].

Another extracellular protease, thermopsin, originally discovered in the thermoacidophile *Sulfolobus acidocaldarius*, is an acid protease that is optimally active at 75°C and pH 2 [103, 104]. Additionally, a thermopsin-like acid protease has been identified in another thermoacidophilic crenarcheaote, *S. solfataricus* [105]. Interestingly, thermopsin is the only peptidase that has homologs exclusively in archaea [106]. Thermopsin is designated an aspartic peptidase as it is inhibited by aspartic protease inhibitors, such as pepstatin [104]. Although the structure of thermopsin has yet to be solved, this protease could have important biotechnological applications when low pH conditions are required. Similarly, extreme thermophiles are salient reservoirs of thermostable proteases that can be tapped for high temperature processes.

Thermostable amylases from extreme thermophiles

Amylases hydrolyze polysaccharides into sugars by breaking glycosidic linkages. Amylases are categorized into three types, which are **α**-, **β**- and **γ**-acting, and classified into endo- and exo-enzymes based on their mode of action [107]. **α**-amylase (EC 3.2.1.1) belongs to family GH13 glycosyl hydrolases (GHs), which randomly cleave **α**-1,4 glycosidic bonds of polysaccharide containing at least three **α**-1,4 glycosidic linkages. **α**-amylase has been extensively

used in starch processing applications [108], where thermoactivity and thermostability requirements have driven interest in extremely thermophile **α**-amylases [109] (see **Table 2.1**). For example, gelatinization, one of the crucial steps in industrial starch processing, benefits from an **α**-amylase with high thermal stability to breakdown starch into oligosaccharides above 70^{o}C [110].

α-Amylases from extreme thermophiles, such as *Thermococcus sp*. [111, 112], *Pyrococcus sp.* [4, 113], *Sulfolobus sp*. [114], and *Thermotoga maritima* [115] (**Table 2.1**), have been examined in some detail. While early efforts necessarily focused on native versions of these enzymes purified from large cultures of extreme thermophiles, expression in heterologous hosts, such as *E. coli*, furthered characterization efforts for these enzymes. For example, the **α**-amylase encoding gene (*amy*A) from *T. maritima* (a lipoprotein [115]), when cloned and expressed in *Escherichia coli*, retained its thermophilic features with activity in the range of 85 – 90°C. However, native forms of these enzymes have also been examined. For example, the intracellular **α**-amylase from *P. furiosus* functions over the range 60 – 100°C, but was inactive below 60°C and above 100°C, due to cold and thermal denaturation, respectively. **α**-Amylases (including recombinant *P. furiosus* **α**-amylase) from extreme thermophiles are typically metalloenzymes, since their thermal stability and enzyme activity increase with Ca^{2+} addition [113, 116].

Improving thermal stability of α-amylases from extreme thermophiles

There have been several attempts to improve the properties of **α**-amylases from extreme thermophiles through protein engineering. Based on the crystal structure of AmyC from *T. maritima*, methionine residues at position 43, 44, 55, and 62 were replaced by the oxidative-resistant alanine. The M55A mutant was the most oxidation-resistant, retaining 50% of its activity; note that the wild-type **α**-amylase was inactivated by H_2O_2 [117].

Site-directed mutagenesis was used to examine the thermostability of **α**-amylase from *Thermococcus onnurineus* NA1. By mutating residues (His175 and Cys189) involved in zinc binding, thermostability decreased [118]. *T. maritima* AmyC, which belongs to family GH57, is a glycogen-branching branching enzyme involved in glycogen biosynthesis. Moreover, the structure of *T. maritima* AmyC is closely related to mesophilic and thermophilic homologs in bacteria, such as *Mesotoga prima*, *Kosmotoga olearia*, and *Kosmotoga pacifica*, indicating the widespread occurrence of GH57 homologs across the microbial world [119]. The GH57 **α**-amylase from *Pyrococcus sp*. ST04 has a unique catalytic activity – it only recognizes maltose (G2) unit **α**-1,4 and **α**-1,6 linkages in polysaccharides. This enzyme is essential for **α**-1,4 and **α**-1,6 linkage hydrolysis in glycogen, whose products later are used as metabolites in glycolysis [120].

β-glucanases

β-Glucans are polymers of β-D-glucose or derivatives of β-D-glucose, with cellulose and chitin β-(1,4)-glucans being the two most abundant polymers on Earth. β-Glucans are produced by a diverse set of organisms, including plants, algae, fungi, yeast, bacteria, and insects [121-124]. As such, they have wide ranging impacts across many fields, including food, medicine, and energy [124]. β-glucans, which polymerize via three main linkages: (1,3), (1,4), and (1,6) β-glucose bonds (and rarely (1,2)), are linear or branched, and some contain multiple types of linkages [124]. Many organisms, including some extreme thermophiles (**Table 2.2**), possess enzymes that hydrolyze β-glucans releasing glucose or disaccharide glucans (or their derivatives) from the non-reducing ends (exo-β-glucanases) or polyglucans via random cleavage in the middle of polymer chains (endo-β-glucanases) [122, 125].

Extremely thermophilic β-glucanases are highly desirable for industrial application, as the high temperatures provide several advantages, including alleviation of contamination concerns and reduction in cooling requirements after thermal pretreatment of biomass prior to enzymatic treatment. There is an industrial need for highly thermostable β-glucanases in the food/beverage and biofuels industry [125, 126]. This has driven efforts to find β-glucanases with high stability, high activity, and broad substrate preference that also function at extreme temperatures. One example is the exo-β-glucanase (NfBGL1) from the moderately thermophilic fungi, *Neosartorya fischeri* P1, which has broad substrate specificity and extraordinarily high optimal activity at 80°C. *N. fischeri* P1 prefers (1,2), (1,6), and (1,3) linkages, but is also capable of cleaving (1,4) linkages [127].

β-(1,4)-glucanases have received attention for bioenergy applications, due to their importance in developing the next generation of bioproducts. Degradation of cellulose and/or chitin into simpler sugars via β-(1,4)-glucan hydrolysis is the first step in next generation bioproduct production [125]. Several extremely thermophilic cellulases and chitinases have been studied. Among the cellulases, the most studied come from two organisms, *Caldicellulosiruptor bescii* [128] and *Hungateiclostridium thermocellum* (f. *Clostridium thermocellum* [129]); although the latter is not an extreme thermophile, it does produce a number of β-glucanases with optimal activity ≥ 70°C [128, 130]. *C. bescii* and *H. thermocellum* approach cellulose degradation in two very distinct ways. *C. bescii* employs several highly thermostable (up to 90°C), multidomain, secreted cellulases, the most prominent being CelA, which contain multiple catalytic domains and several cellulose binding domains [128, 131, 132]. The secreted cellulases encoded by the Glucan Degradation Locus (GDL) of *C. bescii* have a high degree of synergy and have one of the highest thermostabilities of any known cellulases [128, 132]. In contrast, *H. thermocellum*

utilizes much larger complexes, known as cellulosomes (T_{opt} = 70˚C) [133], which are formed by the complex of a scaffoldin with several catalytic domains, linked via type I cohesins and dockerins. The scaffoldin also contains carbohydrate binding domains and a type II dockerin to attach to type II cohesins contained on the cell surface layer [134].

H. thermocellum encodes for a myriad of enzymes with type I dockerin domains, including cellulases and other β-glucanases, some of which are more thermostable than others [130]. For example, this bacterium produces a lichenase (CtLic16A, β-(1,3)(1,4)-glucanase) that has optimal activity at 80˚C, or 10˚C above *H. thermocellum* cellulosome's optimal growth temperature on microcrystalline cellulose [133, 135]. Subsequent research has tried to identify what makes CtLic16A extraordinarily thermostable via crystal structure comparisons to mesophilic β-(1,3)(1,4)-glucanases (38-57% sequence identity); several changes outside the catalytic and binding domains were found to stabilize the protein [136]. The basis for thermostability has been investigated in other highly thermostable β-glucanases as well. The crystal structure of a laminarinase (β-(1,3)-glucanases) from *Rhodothermus marinus* (T_{opt} = 88˚C), which also displays significant activity on β-(1,3)(1,4)-glucans, showed that salt bridges connecting β-sheets of the catalytic domain and the preservation of active site flexibility were key to its stability [137]. Despite several efforts to determine the mechanisms for thermostability of β-glucanases, there does not appear to be many commonalities among this group of enzymes, indicating that there are many types of interactions that can contribute to thermostability.

Hemicellulases

Hemicellulases are a broad class of enzymes responsible for cleaving hemicellulose from cellulose. Of the enzymes that hydrolyze hemicellulose, xylanases and mannanases are

sought after for application, due to the high xylose and mannose content present in hemicellulose. Other hemicellulases include α-L-arabinofuranosidases, α-D-glucuronidases, and β-xylosidases, all of which can be added to xylanase and mannanase cocktails for improved function [138].

Xylanases are GHs that catalyze the hydrolysis of 1,4 β-D-xylose linked polysaccharides, primarily belonging to GH families 10 and 11. Xylanase or xyloglucanase activity has been reported in families 5, 7, 8, and 43 [139]. While GH11 enzymes are generally more active xylanases, GH10s have a higher tolerance for xylan backbone modifications [140, 141]. Xylanases from GH families 10 and 11 are found primarily in bacteria and eukaryotes; however, GH10s have been identified in some thermophilic haloarchaea [142] as well as *S. solfataricus [143]* and *Thermococcus zilligii* [144]. Xylanases are localized both intracellularly and extracellularly and can have binding domains to aid in increasing substrate affinity, which are primarily from CBM families 6, 9 and 22 [145-148]. Xylanases from several genera of extreme thermophiles, including *Caldicellulosiruptor* and *Thermotoga*, have been expressed recombinantly and characterized [149-155].

Mannanases are glycoside hydrolases often from GH families 5, 26, and 116 that cleave endo-β-1,4 linked mannose residues, primarily found in hemicellulose [156]. While mannanases are common in bacteria and eukaryotes, there are limited reports of mannanases in archaea. Like xylanases, mannanases can be localized intracellularly or extracellularly. GH family 5 enzymes are active on a broad range of substrates, including cellulose; several characterized cellulases from *Caldicellulosiruptor* species have significant activity on mannans [132, 147].

Numerous commercial applications for hemicellulases exist, including preparation of sugar for biofuel production, pulp bleaching, grain processing, brewing, food and feed processing, and oil extraction [156, 157]. One of the limiting factors of enzymatic hydrolysis is the heterogeneity

of the substrates being hydrolyzed. An enzymatic cocktail consisting of two GH10 enzymes with debranching enzymes from GH51 and GH67 and a β-xylosidase (GH3) from *C. bescii* was developed. This cocktail was able to completely hydrolyze oat spelt xylan to monomeric sugars at 75°C, suggesting that these enzymes can potentially be used in high temperature applications [149]. Addition of arabinofuranosidase and galactanase to a xylanase/mannanase mixture from *T. maritima* improved sugar yields on barley straw and corn bran [158]. Recently, printing paper waste biobleaching efficiency by a GH10 from *Caldicoprobacter algeriensis* was tested at 70°C and had a 1.5-fold improvement in sugar and phenolic release compared to a thermophilic fungal xylanase [159].

One of the earliest applications for extremely thermophilic hemicellulases was in hydraulic fracturing [6]. Guar gum (β-1,4-galactomannan) is a common thickener added to hydraulic fracking fluid to prevent well collapse during oil extraction. Mannanases from extreme thermophilies can act as "breakers" allowing for easier extraction as the increased viscosity prevents oil flow after the hydraulic fracture is complete [160]. Most notably, β-mannanases, in conjunction with β-mannosidases and α-galactosidases from *Thermotoga neopolitana*, were used in a breaker cocktail, primarily due to their low activity at ambient temperature and high thermostability [154, 161]. Characterization of an extremely thermophilic mannanase from *Dictyloglobum thermophilum* CGMCC 7283 showed that the enzyme significantly decreased the viscosity of crosslinked guar gum gel at 80°C [161], potentially being an alternative to *Thermotoga* enzymes.

Surface (S)-Layer Homology Domain Glycoside Hydrolases

Some plant biomass degrading microorganisms not only employ secretome-bound GHs to degrade plant biomass, they also use cellulosomes, to deconstruct plant biomass [33]. Cellulosomal bacteria have attached Surface Layer Homology (SLH) domain-containing proteins that interact with the scaffoldin protein of the cellulosomal multienzyme complex to attach to the cell surface [162]. SLH GHs have also been identified in extreme thermophiles [163-166]. The genomes of species in the genus *Caldicellulosiruptor* collectively produce six SLH GHs that have catalytic domains suggesting β-glucanase and xylanase activities (see **Table 2.3**). *Caldicellulosiruptor saccharolyticus*, produces an SLH GH containing a catalytic domain (GH5) and carbohydrate binding domain (CBM28), as well as SLH domains (Csac_0678) [167]. Very close homologs to this enzyme can be found in most *Caldicellulosiruptor* species. Recombinant Csac_0678 exhibited endoglucanase and xylanase activity at 75°C. Deletion of SLH domains in Casc_0678 did not affect catalytic activity, implying that the function of these domains was for cell surface attachment. Calkro_0402 and Calkro_0111, SLH GHs found in *C. kronotskyensis*, have been studied in some detail [168]. Calkro_0402 is a GH5-containing enzyme, which also exhibited xylanase activity (T_{opt} 70°C); this corresponds to the features of a Calkro_0402 homolog found in another extreme thermophile, *T. maritima* [169]. Calkro_0111, a GH16- and GH55-containing enzyme, functions as a laminarinase, although the substrates for this enzyme in the terrestrial environments inhabited by *C. kronotskyensis* are unclear.

Xylose Isomerases

Glucose (xylose) isomerase (XI, EC 5.3.1.5), an intracellular bacterial enzyme, has been widely used industrially to catalyze the reversible isomerization of D-glucose (D-xylose) to D-

fructose and D-xylulose found in high fructose corn syrup (HFCS) [170]. HFCS is used as an alternative sweetener to sucrose in a number of industrial applications, such as pharmaceuticals, baking, and beverages (especially in soft drinks) [5, 171]. XIs are categorized into two classes based on the length of the polypeptide. Class I XIs typically contain ~390 amino acids, whereas Class II XIs have an additional 40-50 amino acids at their N-terminus. XIs can be found in many microorganisms, ranging from mesophiles to hyperthermophiles [172].

The conversion of D-glucose to D-fructose is a temperature-dependent reaction because the thermodynamic equilibrium favors D-fructose at elevated temperature. The isomerization reaction achieves 42% conversion at mesophilic to moderately thermophilic temperatures. However, 55% conversion occurs at temperatures above 85°C, which is important for biotechnology since this is the desired fructose to glucose ratio [173, 174]. Thus, XIs with higher thermostability and thermoactivity are strategic for HFCS production to minimize additional downstream concentration processing steps. The XI from *T. maritima* has a temperature optimum of 110°C, and an estimated half-life at 120°C of 10 minutes [5]. Furthermore, the enzyme maintains its thermostability under slightly acidic conditions, which is beneficial for some industrial applications [175]. The thermostability of XIs from *T. maritima* (TMXI) and *T. neapolitana* (TNXI) compare favorably with commercial XIs from mesophiles, such as *Streptomyces murinus* (SMXI) [176].

XIs typically require divalent cations for activity and structural stability. In the presence of Mg^{2+}, temperature optima for SMXI, TMXI, and TNXI were 85°C, > 95°C, and > 100°C, respectively. In the presence of Mg^{2+}, the productivity of SMXI was higher than TMXI and TNXI at 60°C and 70°C. However, TMXI and TNXI catalyzed glucose or xylose isomerization at temperature above 80°C in the presence of Mg^{2+} and Co^{2+}. In another study, a strategy was

proposed to improve the thermostability of extreme thermophilic XI from *Thermoanaerobactor ethanolicus* (TEXI) [173]; site-directed mutagenesis (W139F/V186T), based on the molecular docking and structure modeling, generated a mutant TEXI that exhibited 1.2-fold higher thermostability than the wild-type. In addition, the specific activity of TEXI was improved 230% compared to the wild-type.

Carbon Dioxide Processing Enzymes

Carbon dioxide fixation by biological systems is of great interest for biotechnology, as it presents an opportunity to generate industrial chemicals using a "green" feedstock. Although the cornerstone of plant metabolism, CO_2 fixation is not commonly observed in microorganisms studied to date. However, early microbial life likely utilized CO_2 as a carbon source, given the apparent anaerobic nature of primordial Earth and the lack of photosynthesis. This physiological feature is evident today in many species of chemolithoautotrophic extreme thermophiles. Exploiting this characteristic provides a biotechnological route to address increasing levels of atmospheric CO_2, thereby minimizing global warming [19, 177-179].

Carboxylases

In extremely thermophilic organisms, carboxylases carry the burden of fixing CO_2 for many purposes, including anaplerosis, generation of biomass, and redox balancing [180]. These enzymes utilize a wide variety of substrates and cofactors, which can prove challenging when considering carboxylases for biotechnological purposes (see **Table 2.4**). Utilizing carboxylases within a metabolic pathway, with multiple substrate-specific carboxylation reactions, can be an effective way to generate biosynthetic precursors [181].

The most well-known carboxylase, which acts as the core of the Calvin-Benson-Bassham (CBB) cycle, is ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). A thermophilic variant of Type III RuBisCO from *T. kodakarensis*, which did not function in the CBB cycle natively [182], was used to complete the CBB cycle in *Rhodopseudomonas palustris*, demonstrating activity even at mesophilic temperatures [183]. Means of improving this enzyme's activity and stability at mesophilic temperatures have been explored through site-directed mutagenesis [184] and subsequent expression in *R. palustris* [185].

Exploration of other autotrophic carboxylases has considered multiple enzymes in an effort to overcome substrate specificity of the carboxylases. The 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) cycle in *M. sedula* has been used to generate 3-hydroxypropionic acid, a monomer used in polymerization reactions. The first three enzymes in this cycle, which include acetyl-CoA/propionyl-CoA carboxylase and two reductases, were introduced into *P. furiosus* to convert CO_2 and acetyl-CoA into 3-hydroxypropionate with hydrogen as an electron donor [179].

An important aspect of carboxylases is the trade-off between catalytic efficiency and oxygen sensitivity. Catalytic efficiency impacts the cost of driving the carboxylation reaction, incurred via the use of cofactors or energy-storing molecules (e.g., ATP). RuBisCO, which does not require cofactors or energy input, is well-known for catalyzing a deleterious side reaction with oxygen. Meanwhile, acetyl-CoA/propionyl-CoA carboxylase is insensitive to oxygen, but requires biotin as a cofactor and consumes ATP during the reaction [180]. Optimizing these apparently opposing characteristics is important for biotechnological application of carboxylases.

Carbonic Anhydrases

Carbon capture is an intriguing alternative to direct fixation of CO_2. While fixation requires specific substrates, carbon capture focuses only on the conversion of CO_2 to bicarbonate by carbonic anhydrase. CO_2 solubility limitations and the slow abiotic conversion to bicarbonate make carbonic anhydrases advantageous both for biological CO_2 fixation and for carbon sequestration technologies [186]. Carbonic anhydrases fall into one of six categories (α, β, γ, δ, ζ, or η), each of which is structurally distinct from the others, but all possess an essential zinc ion active site [187]. Thus far, α-carbonic anhydrases (α-CAs) have received the most attention in biotechnological applications.

Thermophilic α-CAs with high activity have been purified and characterized from *Thermovibrio ammonificans* [188] and *Sulfurihydrogenibium yellowstonense* [189], both of which were produced recombinantly in *E. coli*. The purified α-CA from *S. yellowstonense* (SspCA) has garnered much attention for its thermostability and subsequent application in enzyme immobilization for carbon capture. Indeed, these enzymes have been immobilized on myriad surfaces, including the S-layer of *E. coli* [190], magnetic nanoparticles [191], ionic liquid membranes [192], and polyacrylonitrile/polyethylene terephthalate nanofibers [193], for the purposes of carbon sequestration and selective CO_2 transport. Both experimental and computational methods have been employed, using SspCA as a template, to improve α-CA thermostability and activity, either through high-throughput screening platforms [194] or *in silico* modifications of a mesophilic α-CA to improve thermostability [195]. Thermostability of CAs plays a particularly important role when applied to CO_2 scrubbing, where high temperatures are necessary to regenerate the column [196]. Numerous patents exist in this particular application [197-202], highlighting the economic intrigue of thermostable carbonic anhydrases.

Carbonic Anhydrases and Carboxylases in Concert

The potential synergy between carbonic anhydrases and carboxylases in biotechnology has not been overlooked. Coupling enhanced CO_2 solubility from CAs with CO_2 fixation by autotrophic carboxylases has obvious benefits for the cell. Further investigation of the engineered *P. furious* 3-hydroxypropionate pathway showed that product titers were limited by mass transfer of CO_2 [177]. Addition of a CA from *Metallosphaera sedula* improved product titers 3-fold [203, 204]. An anaplerotic phosphenolpyruvate carboxylase from *Thermosynechococcus elongatus* has also been coupled with a thermophilic α-CA from *Sulphurhydrogenibium azorense*; this combination increased production of oxaloacetate compared to the carboxylase alone [205]. While each of these enzymes has its own intrinsic value, the combined application of carboxylases and carbonic anhydrases will continue to be a focus of biotechnological research.

Hydrogenases and Hydrogen Production

Hydrogenases are metalloenzymes that reversibly catalyze the formation of hydrogen gas from protons and an electron carrier. These enzymes are further classified into three types based on their metal centers, [FeFe], [NiFe], and [Fe], and can be localized to either the cell membrane or cytoplasm [206, 207]. [NiFe] hydrogenases are present in bacteria, eukaryotes, and archaea, [FeFe] hydrogenases are present only in bacteria and eukaryotes, and [Fe] hydrogenases are limited to methanogenic archaea [208]. In addition to the metal center present in the hydrogenase, the electron carrying cofactor is also important for function. The standard potential of proton reduction is -414 mV, so that electrons need to be transferred onto an energy carrier with a reduction potential less than NAD(P)H (E° -320 mV) in order to drive hydrogen production; often this will involve ferredoxin, a small iron-sulfur protein [207].

Hydrogenases have been identified and characterized in numerous organisms [209], with the most well characterized extremely thermophilic hydrogenases coming from *P. furiosus* and *Thermotoga* species. *P. furiosus* has three hydrogenases, two soluble hydrogenases (SHI and SHII) and a membrane bound hydrogenase (MBH), all of which have been characterized biochemically [47, 210, 211]. Interestingly, the two soluble hydrogenases transfer electrons using nicotinamide cofactors as a method for reducing $NADP^+$ for SHI and NAD^+ for SHII [212]. Additionally, the first bifurcating hydrogenase was discovered in the extremely thermophilic bacterium *T. maritima,* which transfers electrons from both reduced ferredoxin and NADH to allow electrons to be at an electronic potential suitable for proton reduction [213]. These bifurcating hydrogenases have been identified in other extreme thermophiles, including *Caldicellulosiruptor* sp. [214], and have also been reported to require tungsten for function, an uncommon requirement for biological systems [215].

Development of a renewable biohydrogen economy requires overcoming technical barriers, such as low yield and volumetric productivity, and practical concerns such as transport and storage [216]. Although several anaerobic extreme thermophiles produce hydrogen metabolically, carbohydrate fermentation via the EMP pathway is limited to 4 mol H_2/mol glucose, despite the potential for generating 12 mol H_2/mol glucose [217]. Using elevated temperatures can help drive hydrogen production by decreasing the solubility of the gas. However, overcoming the so-called “Thauer limit” requires an alternative strategy, including fermentation of an alternative carbon source (e.g. formate), diverting carbon flux alternative redox pathways, or using cell free systems. The 18 subunit formate hydrogen lyase from *Thermococcus oneurensis* was integrated into the genome of *Pyrococcus furiosus*, allowing for near stoichiometric conversion of hydrogen from formate at 80°C and doubling the hydrogen production rate compared to the parent strain

[218]. Disruption of the lactate dehydrogenase gene in *T. maritima* allowed for hydrogen production greater than 4 mol H_2/mol glucose through increased flux through the oxidative pentose phosphate pathway [219]. Enhancing microbial fermentation capabilities can be an attractive strategy for hydrogen production.

While metabolic engineering of extreme thermophiles can increase hydrogen production, cell-free systems ultimately allow for obtaining the most hydrogen from complex carbohydrates. Of the characterized extremely thermophilic hydrogenases, only the soluble hydrogenase I (SHI) from *P. furiosus* has been shown to oxidize NADPH to produce hydrogen [220] (see **Figure 2.3**). Despite the normal thermodynamic limitations, hydrogen has low solubility in aqueous solution, allowing for nicotinamide cofactors to drive hydrogen production under certain conditions. Recombinant expression and maturation in *E. coli* has expanded potential applications of this enzyme in numerous *in vitro* systems by providing an easier expression host [221]

Cell-free hydrogen production using *P. furiosus* SHI is accomplished by combining mesophilic metabolic enzymes to generate glucose-6-phosphate from a carbon source and the pentose phosphate pathway to generate NADPH, from which hydrogen is generated. This strategy exploits the basal activity of SHI at mesophilic temperatures, despite being from an extremely thermophilic archaeon. While several carbon sources have been reported using this pathway [222-224], replacing the mesophilic enzymes with equivalent thermophilic enzymes and increasing the temperature to 50°C increased H_2 generation by 200-fold [225]. Additionally, SHI has been added as a component to a photocatalytic CdS quantum dot system, allowing for light to serve as the electron source in hydrogen production at room temperature; further improvements resulted when viologen donors were added to facilitate electron transfer [226, 227]. Ultimately, generating hydrogen at high productivities will require complementation with additional thermophilic

enzymes to quickly volatilize hydrogen and drive the consumption of target carbon sources so that the enzymes are working under more optimized conditions.

Genetic Tools for Extreme Thermophiles

To leverage the enzymes from extreme thermophiles for biotechnological purposes, there must be robust genetic tools for the genetic manipulation and native expression of these thermophilic enzymes. However, due to the difficulties of culturing extremophiles and their more recent discovery, the development of established genetic systems has been slower than for their mesophilic counterparts. Particularly, the lack of thermostable antibiotic selection has been a significant hindrance and has led those developing genetic systems for extreme thermophiles to largely depend on auxotrophic selection. For instance, selecting for uracil prototrophy in an uracil auxotrophic mutant, followed by 5-fluororotic acid (5-FOA) counterselection, is now a common technique that is used in several extreme thermophiles, including *Sulfolobus acidocaldarius* [74], *Thermococcus kodakaraensis* KOD1 [228], *C. bescii* [229], *P. furiosus* [230], and *Thermotoga sp.* Strain RQ7 [231]. When making unmarked mutants, this method first selects for integration of a plasmid carrying the uridine monophosphate synthesis pathway genes *pyrEF* using medium lacking uracil. The plasmid backbone is then selected against with a 5-FOA counterselection that can be converted to the toxic product 5-fluorouracil if the ability to synthesize uracil is retained. Each selection step drives a homologous recombination event that will result in either generation of the desired mutant or reversion to the parent genotype (**Figure 2.4**).

The use of uracil auxotrophy and 5-FOA counterselection is the most pervasive selection technique in extreme thermophiles, but there are a few reported variations to this theme. Selection for agmatine prototrophy has been successfully implemented in *Sa. islandicus* (f. *Sulfolobus*

islandicus) and *T. kodakaraensis* [232]. Also, a 6-methylpurine counterselection (a purine analog) in an adenine phosphoribosyltransferase (*apt*) mutant has been utilized in *Sa. islandicus* [233] and *Thermococcus barophilus* [234]. Furthermore, a member of the Crenarcheaota phylum, *S. solfataricus*, depends on lactose prototrophy for mutant generation [73, 235]. While genetics for the euryarchaeotes, *P. furiosus* [236] and *T. kodakaraensis* [228] have used tryptophan auxotrophic selection, this has been accompanied by 5-fluoroanthranilic acid (5-FAA) counterselection in *P. furiosus* [236].

Although antibiotic selection is scarcely used in extreme thermophile genetics, there are a few thermostable antibiotic markers available. A highly thermostable kanamycin marker in *C. bescii* [237] is now an established method of selection. With the use of this antibiotic marker and deletion of *pyrE*, the recent engineered *C. bescii* strain (MACB1018) with high genomic stability has become more suitable for further metabolic engineering [238]. Simvastatin selection, alongside a *hmgA* (3-hydroxy-3-methylglutaryl coenzyme A reductase) gene marker, has been successfully used in *Sa. islandicus* [239, 240], *T. kodakaraensis* [241], and *P. furiosus* [44], but spontaneous generation of simvastatin resistant cells make this a less desirable method.

For controlled expression of thermophilic enzymes, inducible promoters are required that are not yet widely available for extremely thermophilic organisms. Most of the available inducible promoter systems reported so far are for species in the thermoacidophilic order Sulfolobales, such as the *araS* arabinose inducible promoter in *Sa. solfataricus* [242] and the *mal* maltose inducible promoter in *S. acidocaldarius* [243]. Recently, a XylR activator based xylose/arabinose inducible promoter has been developed for *S. acidocaldarius* and has demonstrated greater control over gene expression than its predecessors [244]. Outside of the Sulfolobales, a xylose inducible promoter

has been utilized in *C. bescii* [245] and a fluoride-inducible riboswitch has been used in *T. kodakarensis* [246].

Reporter systems are important tools when studying the expression and localization of biomolecules within extreme thermophiles. However, there are only a few thermostable reporters available, such as the β-galactosidase in *S. acidocaldarius*, that was used to develop the previously mentioned maltose inducible promoter [243]. Similarly, two β-galactosidase dependent reporter systems have been developed for *T. kodakarensis* [232, 247].

Other requirements for genetic manipulations are methods for transformation, to force the DNA of interest inside the cell, and DNA methylation. While electroporation is the method of choice for most extreme thermophile genetics, some are naturally competent, which can be a significant advantage for transforming cells. The naturally competent extreme thermophile strains include *T. kodakarensis KOD1* [228] and *P. furiosus* COM1, the latter of which was genetically manipulated to be competent [44]. Furthermore, DNA methylation can be critical for transformation efficiency. Fortunately, there are known methylation patterns for *S. acidocaldarius* [248, 249] and *C. bescii* [250], knowledge of which is required for effective transformation. However, in some extreme thermophiles, such as *Sa. solfataricus*, unmethylated DNA can be effectively transformed [235, 251]. The current collection of extreme thermophile genetic tools is summarized in **Table 2.5**.

In summary, while there are established methods for the genetic manipulation of extreme thermophiles, efforts continue to address limitations that restrict even more widespread application of these microorganisms and their enzymes.

Metabolic Engineering

Use of extreme thermophiles as hosts for metabolic engineering has garnered significant attention to produce industrially relevant compounds, particularly biofuels [15, 17]. These hosts can make and utilize enzymes suitable for their native environments. Extreme thermophile hosts are usually chosen for desirable features not exhibited by model mesophilic organisms, which in many cases go beyond thermostability or thermoactivity. Examples of this include hydrogen production, lignocellulose deconstruction, unique central metabolism pathways, and highly tractable genetic systems.

Beside the intrinsic difficulties associated with developing genetic systems for non-model organisms (see above) [252], extreme thermophiles also encounter challenges that must be addressed to function as metabolic engineering hosts. For example, finding and recruiting key enzymes with sufficient thermoactivity to build novel synthetic metabolic pathways can limit the scope of applications.

One attractive feature in some extreme thermophiles is the presence of modified Embden–Meyerhof–Parnas (EMP) glycolytic pathways. Unique energy generation and utilization modifications are present, such as the bypass of 1,3-bisphosphoglycerate (1,3-BPG) from the direct conversion of glyceraldehyde-3-phosphate (GAP) to 3-phosphoglycerate (3-PG) using ferredoxin or $NADP^+$ dependent enzymes, eliminating NADH and ATP generation (present in some extremely thermophilic archaea like *P. furiosus* and bacteria like *Caldicellulosiruptor* species) [253-260]. Some species regenerate their electron carriers (ferredoxin, NAD(P), etc.) through hydrogenases, allowing for significant hydrogen production [261]. Hydrogen generation also has potential as a valuable co-product. *C. bescii* was metabolically engineered to make acetone with

hydrogen as a co-product; *C. bescii*'s ability to regenerate reducing equivalents through hydrogen production eliminated the need for by-products such as lactate, ethanol, or other alcohols [262].

C. bescii (T_{opt} 78˚C) has been extensively used as a metabolic host, because of its native biomass degradation capabilities. *C. bescii* has a cadre of cellulolytic and hemicellulolytic enzymes, which can eliminate or greatly reduce the need for pretreatments (enzymatic, chemical, mechanical) to breakdown biomass [128, 132, 263-266]. *C. bescii* has been engineered to produce desirable chemicals from non-native pathways, resulting in strains for acetone and ethanol production [262, 267]. However, as mentioned above, a reoccurring metabolic engineering issue is the mismatch between an organism's optimal growth temperature (*e.g. C. bescii* T_{opt} 78˚C) and the optimal temperature of a biosynthetic pathway's enzymes (*e.g.* ethanol at 60˚C or acetone at 70˚C) [262, 268]. This mismatch highlights the difficulty of locating enzymes that match the extremely thermophilic characteristic of *C. bescii*, a challenge that exists for all extremely thermophilic hosts.

P. furiosus (T_{opt} 100˚C) has also been the focus of significant metabolic engineering efforts due to its high growth temperature and genetic tractability of the naturally competent COM1 strain [44]. *P. furiosus* is a prime example of the difficulty associated with sourcing highly thermostable non-native enzymes due to its hyperthermophilic nature. Ethanol production has been demonstrated in *P. furiosus* [265, 268], enabled by a variety of non-native alcohol dehydrogenases. But, due to thermostability limitations, ethanol production was only possible at suboptimal growth temperatures. Recently progress has been made in elevating the optimal temperature for ethanol production from 72˚C to 80˚C through the use of a new alcohol dehydrogenase and modification of a central glycolytic pathway to enable improved redox balancing [253, 269, 270], but further advances are needed. It is possible to grow *P. furiosus* at its optimal temperature to generate cell

mass, then reduce the temperature, thereby triggering a cold-shock promotor to activate a desirable pathway; this was shown for lactate and 3-hydroxypropionate (3HP) production [179, 271].

The incompatibility between enzymes needed for metabolically engineered pathways in extreme thermophiles and availability of homologs that are sufficiently thermostable and thermoactive raises issues. If the enzyme has not been characterized in an extreme thermophile, amino acid homology searches based on less thermostable versions of the desired enzyme can be conducted. However, the homology between a mesophilic and thermophilic enzyme with the same function can be very low, usually with high conservation of active site sequences. A few reported examples show that amino acid identity homologies as low as ~30% are sometimes considered [270, 272, 273]. Extensive expression and screening efforts with these proteins are needed to identify their functionality. For example, in *P. furiosus* ethanol production, 8 bifunctional alcohol dehydrogenase (AdhE) homologs were expressed and compared for ethanol production. AdhE from *H. thermocellum* [129] was used as a basis for a BLAST search; the homology to *H. thermocellum* AdhE among the enzymes identified ranged from 47%-71%, while homology to the *E. coli* AdhE ranged from 46-63%. Many of the source organisms natively produced at least some ethanol, but only two of these AdhE candidates produced ethanol *in vivo* in *P. furiosus* [270]. To complicate the situation further, the ideal enzyme can sometimes be sourced from a mesophile, where the enzyme has exceptional thermostability; this was seen for the acetoacetate decarboxylase (sourced from *Clostridium acetobutylicum*, active up to 80°C) used in the *C. bescii* acetone pathway [262, 272].

There are cases where enzymes from extreme thermophiles can benefit mesophilic organisms. One example of this is the superoxide reductase from *P. furiosus* (PfSOR) that was used as an alternative to native superoxide dismutase to improve stress tolerance in plants (heat,

light, chemical) and silkworms (heat) [274-277]. PfSOR allows for the reduction of superoxide (O_2^-) without the formation of oxygen (O_2) that would cause additional reactive oxygen species to form. PfSOR also has a higher affinity for oxygen than typical superoxide dismutases, making a more efficient detoxification system [278, 279].

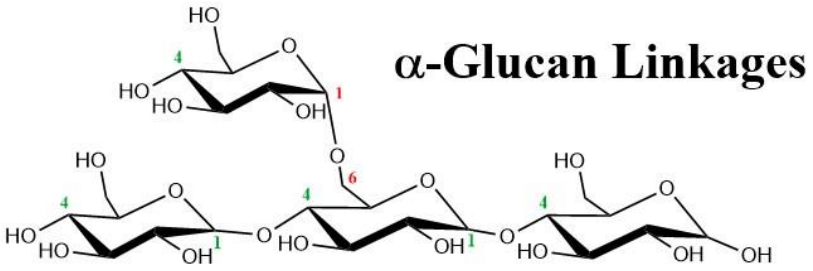
Future Directions

The expansion of the thermal range for biocatalysis through the discovery of extremely thermophilic microorganisms has certainly opened up many opportunities for biotechnological application. Of course, characteristics besides thermostability and thermoactivity can be crucial for specific intended uses, but adjusting active site features to tune biocatalysis is a smaller task than significant improvement in thermophily. An important opportunity going forward is exploiting multi-step biocatalytic transformations in either *in vitro* settings or through metabolic engineering. Improvements in the genetic toolbox for extreme thermophiles will be critical for realizing these potential opportunities. As mathematical models for protein structure prediction improve, these could be pivotal in making subtle adjustments in enzyme specificity, thereby opening up more possibilities to tweak biocatalytic function to meet application needs. Of course, searching the ever-expanding genomics database with improved tools for identifying new enzymes is an exciting path for biocatalyst discovery, especially as more is understood about enzyme structure:function. There are still many thermal environments that have not been surveyed for new microbes, and this route to discovering new biocatalysts should not be ignored. Perhaps, there will be discoveries of thermophilic life on other solar bodies that will further expand the scope of enzyme applications.

Acknowledgments

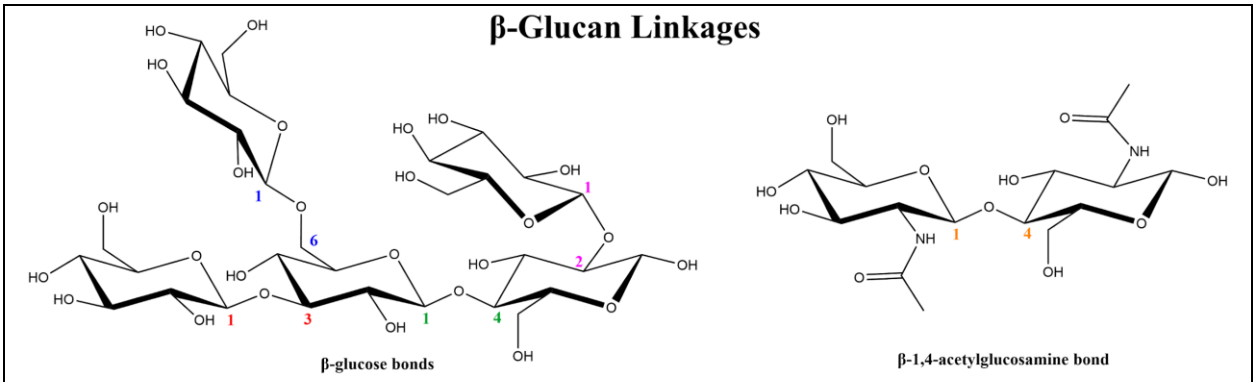
RMK acknowledges support from the US National Science Foundation (CBET-1802939), the US Department of Energy (DE-SC0019391), the US Air Force Office of Sponsored Research (FA9550-17-1-0268), and the US Department of Agriculture (2018-67021-27716). RGB and DJW acknowledge support from US NIH Biotechnology Traineeships (T32 GM008776-16, T32 GM133366-01). JRC acknowledges support from a US Department of Education Graduate Assistance in Areas of National Needs (GAANN) Fellowship (P200A140020).

Table 2.1 - α−Amylolytic enzymes found in extreme thermophiles



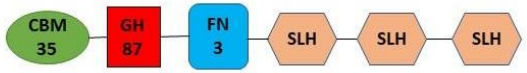
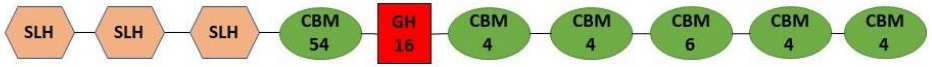
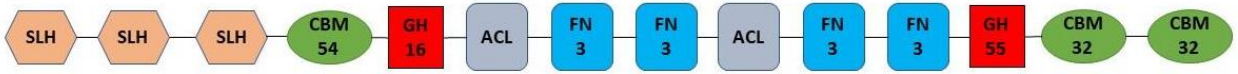
α-amylase type	**Substrate**	**Example Organisms**
α-(1,4)-glucanase-amylase)	soluble starch, amylose, amylopectin, glycogen, dextrin, α-cyclodextrin, p-nitrophenyl-α-D-glucopyranoside, maltose, isomaltose, methyl-α-D-glucopyranoside	*Thermotoga maritima* [115] *Thermoccus profundus* [111] *Pyrococcus woesei* [280] *Sulfolobus solfataricus* [114] *Pyrococcus furiosus* [113]
α-(1,6)-glucanase (pullulanase)	starch, pullulan, isopanose, amylopectin, glycogen, dextran	*Thermococcus kodakarensis* [281] *Thermococcus aggregans* [282] *Thernotoga maritima* [283]
α-(1,4)-(1,6)-glucanase (amylopullulanse)	starch	*Pyrococcus furiosus* [284] *Thermococcus litoralis* [284]
α-(1,4)-glucosidase (maltase, isomaltase)	maltose, isomaltose	*Pyrococcus furiosus* [2]
α -glucohydrolase (sucrase, invertase)	sucrose	*Pyrococcus furiosus* [285]

Table 2.2 - β-Glucanases found in extreme thermophiles



β-glucanase Type	**Substrate**	**Example Organisms**
β-(1,3)-glucanase (laminarinase)	curdlan, laminarin and yeast cell wall (main chains)	*Rhodothermus marinus* [286] *Pyrococcus furiosus* [287]
β-(1,4)-acetylglucosamidase (chitinase)	chitin	*Thermococcus kodakarensis* [286], *Pyrococcus furiosus* [288]
β-(1,3)-(1,4)-glucanase (licheninase)	cereal glucans, lichenin	*Talaromyces emersonii* [289], *Hungateiclostridium thermocellum* [135]
β-(1,4)-glucanase (cellulase)	crystalline cellulose, carboxymethylcellulose (CMC)	*Caldicellulosiruptor bescii* [128], *Hungateiclostridium thermocellum* [290]
β-(1,6)-glucanase (pustulanase)	pustulan, laminarin and yeast cell wall (branches)	*Caldicellulosiruptor kronotskyensis [168], Neosartorya fischeri* P1 [127]
β-(1,2)-glucanase	sophorose, bacterial glucans	*Neosartorya fischeri* P1 [127]

Table 2.3 - Surface Layer Homology (SLH) Domain Glycoside Hydrolases (GHs) encoded in *Caldicellulosiruptor sp.* genomes (adapted from [167, 168])

Domains	#amino acids	Genome Locus
	755-756	Athe_0594, Calkro_2036, Calhy_2064, Csac_0678, Calkr_2007, Calla_0352
	1591-1975	Calkro_0402, Calow_1924, Calkr_2245, Calla_0206
	2007	Calhy_2383
	1724-1732 *526	Calkro_0072, Calhy_0060 *Csac_2549 (partial)
	2435	Calkro_0111
	1440	Calhy_1629

GH Glycoside Hydrolase
CBM Carbohydrate Binding Module
SLH Surface Layer Homology
FN3 Fibronectin type III-like
Cad Cadherin-like, ACL Actin Cross Linking-like RICIN superfamily

Table 2.4 - Autotrophic and assimilatory carboxylases (adapted from [180])

Carboxylase	**Associated Cycle(s)**	**Coupled Reduction**	**Organic substrate**	**Cofactors**	**Oxygen sensitive**
Isocitrate dehydrogenase	rTCA	Yes	α-ketoglutarate	NADPH	No
α-ketoglutarate:ferredoxin oxidoreductase	rTCA	Yes	Succinyl-CoA	Ferredoxin	Yes
Pyruvate:ferredoxin oxidoreductase	rTCA DC/HB	Yes	Acetyl-CoA	Ferredoxin	Yes
Ribulose-1,5-bisphosphate carboxylase/oxygenase Rubisco)	CBB	No	Ribulose-1,5-bisphosphate	N/A	No – side reaction with O_2
Acetyl-coa/propionyl-coa carboxylase	HP HP/HB	No	Acetyl-CoA Propionyl-CoA	Biotin, ATP	No
Phosphoenolpyruvate carboxylase	DC/HB	No	Phosphoenolpyruvate	ATP, Mg^{2+}	No

rTCA = reverse Tricarboxylic Cycle

DC/HB = Dicarboxylate/ 4-Hydroxybutyrate Cycle

CBB = Calvin-Benson-Bassham Cycle

HP = 3-Hydroxypropionate Bicycle

HP/HB = 3-Hydroxypropionate/4-Hydroxybutyrate Cycle

Table 2.5 - Genetic systems for extreme thermophiles

Extreme Thermophile	Auxotrophic Selection						Counter Selection			Antibiotic Markers		Inducible Promoters				β-galactosidase Reporter	Naturally Competent	References
	Uracil	Tryptophan	Agmatine	Adenine	Lactose	Arginine/Citrulline	5-FOA	6-methylpurine	5-FAA	Kanamycin	Simvastatin	Arabinose	Maltose	Xylose	Flouride Inducible Riboswitch			
Firmicutes																		
Caldicellulosiruptor bescii	X						X			X				X				[229, 237, 245]
Crenarcheaota																		
Sulfolobus acidocaldarius	X						X						X	X		X		[243, 244]
Saccharolobus islandicus	X		X	X			X	X			X							[233, 239, 240, 291]
Saccharolobus solfataricus					X							X						[73, 242]
Euryarcheaota																		
Pyrococcus furiosus	X	X					X		X		X						X	[44, 230, 236]
Thermococcus kodakarensis	X	X	X			X	X	X			X				X	X	X	[228, 232, 241, 246, 247]

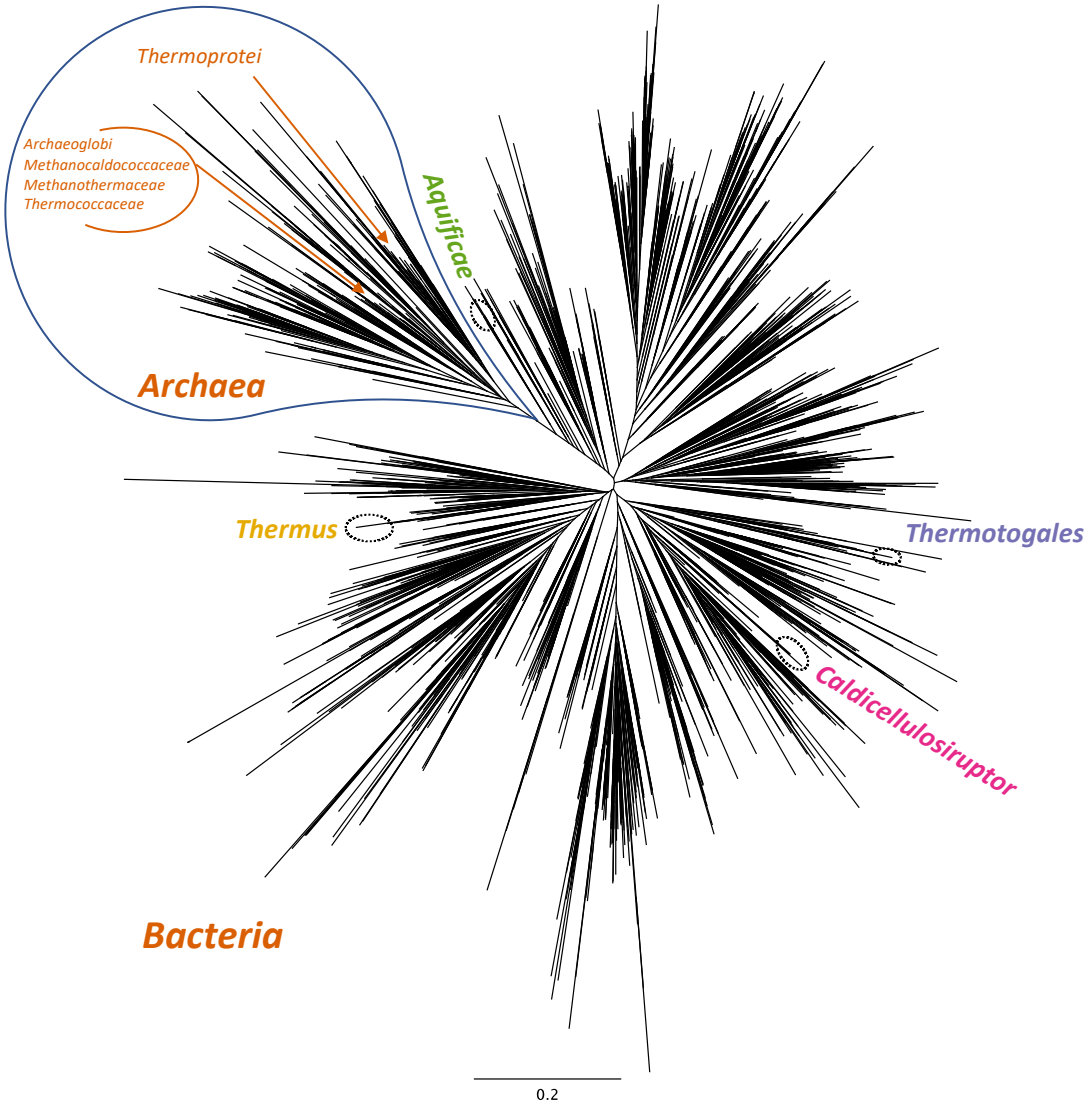


Figure 2.1 - Phylogenetic tree of 16S rDNA sequences from prokaryotic organisms. Sequences derived from the Living Tree Project (v. 138) SSU Database. All non-bacterial and non-archaeal sequences were removed and the dataset was simplified to 80% similarity using CD-HIT-EST. The non-redundant 80% similarity dataset was aligned using MAFFT and sites with greater than 80% gapped bases were removed from a final alignment used to construct a tree with the approximately maximum likelihood FastTree algorithm. Relative locations of taxa of interest are shown in colored regions with corresponding taxonomic naming.

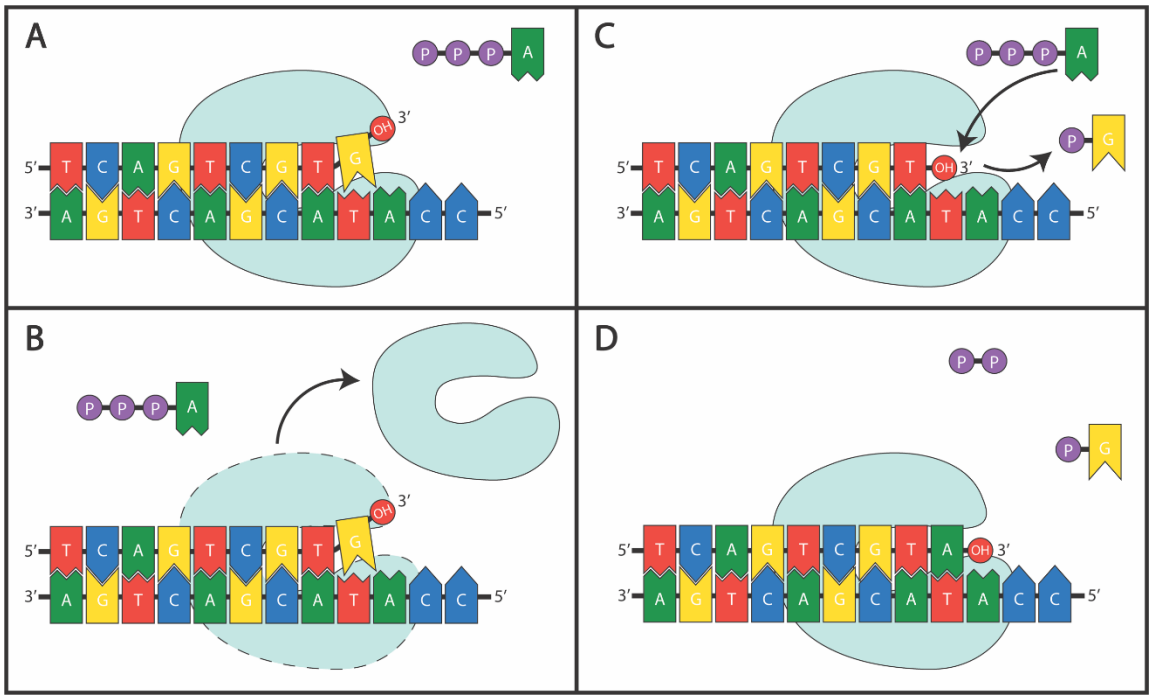


Figure 2.2 - DNA Polymerase Proofreading 3' to 5' Exonuclease Activity (Source: Authors) A) At some frequency DNA polymerases will incorporate a mismatched nucleotide. B) Non-proofreading DNA polymerases, e.g. *Taq*, will often dissociated from the DNA when a mismatched nucleotide is encountered. C) Proofreading DNA polymerases, e.g. *Pfu*, Phusion®, or Q5®, will excise the misincorporated nucleotide when a mismatch is encountered and D) continue adding bases.

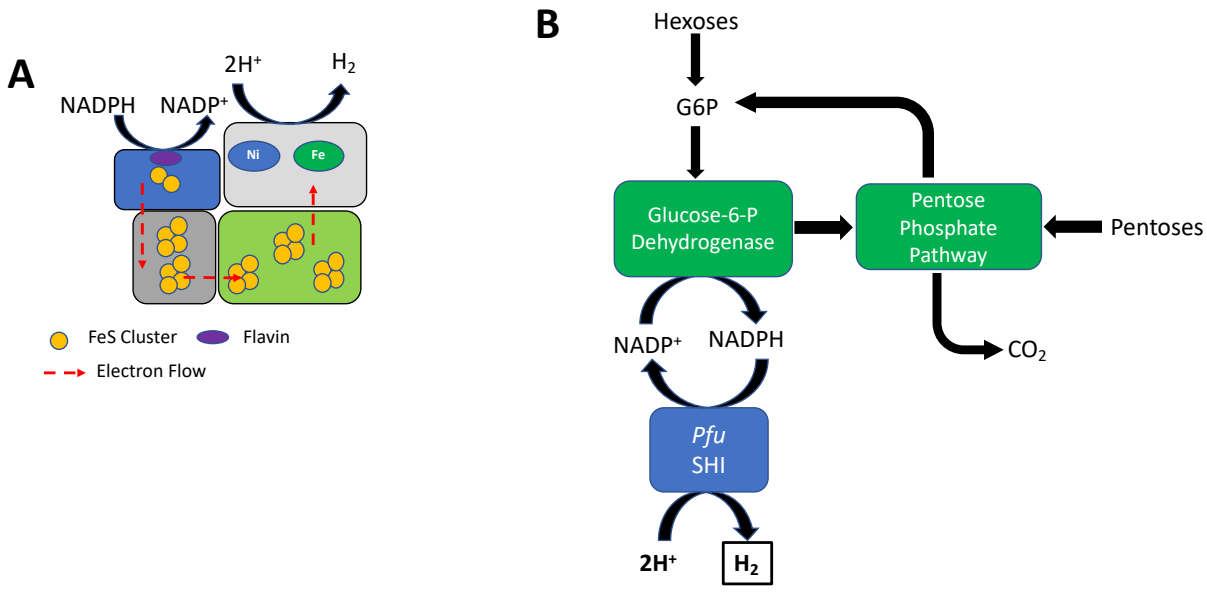


Figure 2.3 - Soluble Hydrogenases and their Applications for Cell Free Hydrogen Production (Source: Authors) A) Domain structure and electron flow in *Pyrococcus furiosus* Soluble Hydrogenase I (SHI). The heterotetramer contains one flavin unit to accept electrons from NADPH, which then flow through several iron sulfur clusters to the Ni-Fe catalytic domain. (Adapted from [220]. B) Application of *Pfu* SHI in cell free hydrogen production. Both pentose and hexose sugars can be converted to glucose-6-phosphate (G6P). Through the glucose-6-phosphate dehydrogenase, NADPH can donate electrons onto protons to produce hydrogen. Independent of carbon sources, the pentose phosphate pathway is used to regenerate G6P to continue to power hydrogen production.

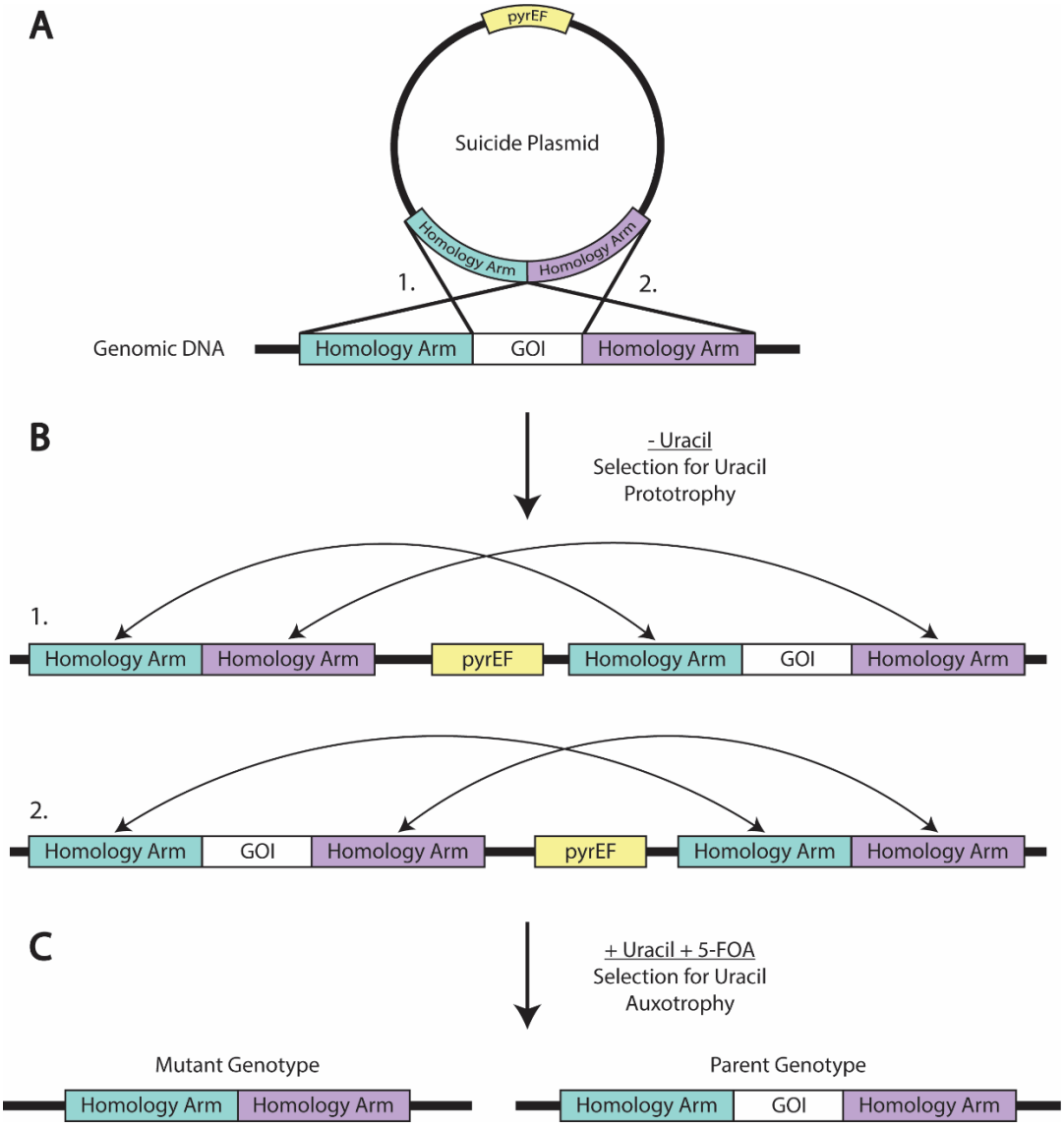


Figure 2.4 - Mutant Generation using Uracil Auxotrophy and 5-FOA Counterselection (Source: Authors) Schematic depicting the creation of an unmarked mutant deleting a gene of interest (GOI) using uracil auxotrophic selection followed by 5-FOA counterselection. A) A non-replicating suicide plasmid carrying the *pyrEF* genes can undergo one of two homologous recombinations to integrate into the genome. B) Media without uracil supplementation selects for uracil prototrophy and the integration of the plasmid into the genome. This first homologous recombination event can create two possible products. C) Media supplemented with uracil and 5-FOA will counterselect for a second homologous recombination event that will either generate the desired mutant or revert the genotype back to the parent. The 5-FOA counterselection at this step selects for uracil auxotrophy and against the presence of the *pyrEF* genes on the plasmid backbone.

References

[1] Adams MW, Perler FB, Kelly RM. Extremozymes: expanding the limits of biocatalysis. Biotechnology (N Y). 1995;13(7):662-8.

[2] Costantino HR, Brown SH, Kelly RM. Purification and characterization of an alpha-glucosidase from a hyperthermophilic archaebacterium, *Pyrococcus furiosus*, exhibiting a temperature optimum of 105 to 115 degrees C. J Bacteriol. 1990;172(7):3654-60.

[3] Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239(4839):487-91.

[4] Brown SH, Costantino HR, Kelly RM. Characterization of amylolytic enzyme activities associated with the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Appl Environ Microbiol. 1990;56(7):1985-91.

[5] Brown SH, Sjøholm C, Kelly RM. Purification and characterization of a highly thermostable glucose isomerase produced by the extremely thermophilic eubacterium, *Thermotoga maritima*. Biotechnol Bioeng. 1993;41(9):878-86.

[6] McCutchen CM, Duffaud GD, Leduc P, Petersen AR, Tayal A, Khan SA, et al. Characterization of extremely thermostable enzymatic breakers (alpha-1,6-galactosidase and beta-1,4-mannanase) from the hyperthermophilic bacterium *Thermotoga neapolitana* 5068 for hydrolysis of guar gum. Biotechnol Bioeng. 1996;52(2):332-9.

[7] Zwickl P, Lottspeich F, Baumeister W. Expression of functional *Thermoplasma acidophilum* proteasomes in *Escherichia coli*. FEBS Lett. 1992;312(2-3):157-60.

[8] Nelson KE, Eisen JA, Fraser CM. Genome of *Thermotoga maritima* MSB8. Methods Enzymol. 2001;330:169-80.

[9] Robb FT, Maeder DL, Brown JR, DiRuggiero J, Stump MD, Yeh RK, et al. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. Methods Enzymol. 2001;330:134-57.

[10] Robertson DE, Chaplin JA, DeSantis G, Podar M, Madden M, Chi E, et al. Exploring nitrilase sequence space for enantioselective catalysis. Appl Environ Microbiol. 2004;70(4):2429-36.

[11] Richardson TH, Tan X, Frey G, Callen W, Cabell M, Lam D, et al. A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. J Biol Chem. 2002;277(29):26501-7.

[12] Atalah J, Caceres-Moreno P, Espina G, Blamey JM. Thermophiles and the applications of their enzymes as new biocatalysts. Bioresour Technol. 2019;280:478-88.

[13] Cabrera MA, Blamey JM. Biotechnological applications of archaeal enzymes from extreme environments. Biol Res. 2018;51(1):37.

[14] Coker JA. Extremophiles and biotechnology: current uses and prospects. F1000Res. 2016;5.

[15] Crosby JR, Laemthong T, Lewis AM, Straub CT, Adams MW, Kelly RM. Extreme thermophiles as emerging metabolic engineering platforms. Curr Opin Biotechnol. 2019;59:55-64.

[16] Kruger A, Schafers C, Schroder C, Antranikian G. Towards a sustainable biobased industry - Highlighting the impact of extremophiles. N Biotechnol. 2018;40(Pt A):144-53.

[17] Schocke L, Brasen C, Siebers B. Thermoacidophilic *Sulfolobus* species as source for extremozymes and as novel archaeal platform organisms. Curr Opin Biotechnol. 2019;59:71-7.

[18] Straub CT, Counts JA, Nguyen DMN, Wu CH, Zeldes BM, Crosby JR, et al. Biotechnology of extremely thermophilic archaea. FEMS Microbiol Rev. 2018;42(5):543-78.

[19] Zeldes BM, Keller MW, Loder AJ, Straub CT, Adams MW, Kelly RM. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Front Microbiol. 2015;6:1209.

[20] Setchell WA. The upper temperature limits of life. Science. 1903;17(441):934-7.

[21] Tansey M, Brock T. The upper temperature limit for eukaryotic organisms. Proc Natl Acad Sci U S A. 1972;69(9):2426-8.

[22] Boyd ES, Fecteau KM, Havig JR, Shock EL, Peters JW. Modeling the habitat range of phototrophs in Yellowstone National Park: toward the development of a comprehensive fitness landscape. Front Microbiol. 2012;3:221.

[23] Brown PB, Wolfe GV. Protist genetic diversity in the acidic hydrothermal environments of Lassen Volcanic National Park, USA. J Eukaryot Microbiol. 2006;53(6):420-31.

[24] Gupta RS. The phylum Aquificae. The Prokaryotes2014. p. 417-45.

[25] Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, König H, et al. *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. Syst Appl Microbiol. 1992;15(3):340-51.

[26] Guiral M, Prunetti L, Aussignargues C, Ciaccafava A, Infossi P, Ilbert M, et al. The hyperthermophilic bacterium *Aquifex aeolicus*: from respiratory pathways to extremely resistant enzymes and biotechnological applications. Adv Microb Physiol. 2012;61:125-94.

[27] Hügler M, Huber H, Molyneaux SJ, Vetriani C, Sievert SM. Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. Environ Microbiol. 2007;9(1):81-92.

[28] Counts JA, Zeldes BM, Lee LL, Straub CT, Adams MWW, Kelly RM. Physiological, metabolic and biotechnological features of extremely thermophilic microorganisms. Wiley Interdiscip Rev Syst Biol Med. 2017;9(3).

[29] VanFossen A, Ozdemir I, Zeline S, Kelly R. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Biotechnol Bioeng. 2011;108(7):1559-69.

[30] Vanfossen AL, Verhaart MR, Kengen SM, Kelly RM. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl Environ Microbiol. 2009;75(24):7718-24.

[31] VanFossen A, Lewis D, Nichols J, Kelly R. Polysaccharide degradation and synthesis by extremely thermophilic anaerobes. Ann NY Acad Sci. 2008;1125(1):322-37.

[32] Conners SB, Montero CI, Comfort DA, Shockley KR, Johnson MR, Chhabra SR, et al. An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. J Bacteriol. 2005;187(21):7267-82.

[33] Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, et al. Thermophilic lignocellulose deconstruction. FEMS Microbiol Rev. 2014;38(3):393-448.

[34] Bhandari V, Gupta RS. The phylum Thermotogae. The Prokaryotes2014. p. 989-1015.

[35] Koyama Y, Hoshino T, Tomizuka N, Furukawa K. Genetic transformation of the extreme thermophile *Thermus thermophilus* and of other *Thermus* spp. J Bacteriol. 1986;166(1):338-40.

[36] Oshima T, Imahori K. Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. Int J Syst Bacteriol. 1974;24(1):102-12.

[37] Brock TD, Freeze H. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. J Bacteriol. 1969;98(1):289-97.

[38] Chien A, Edgar DB, Trela JM. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. J Bacteriol. 1976;127(3):1550-7.

[39] Blöchl E, Rachel R, Burggaf S, Hafenbradl D, Jannasch H, Stetter K. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit of life to 113˚C. Extremophiles. 1997(1):14-21.

[40] Kashefi K, Lovley DR. Extending the upper temperature limit for life. Science. 2003;301(5635):934.

[41] Fiala G, Stetter K. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100~ Arch Microbiol. 1986;145:56061.

[42] Kurr M, Huber R, König H, Jannasch H, Fricke H, Trincone A, et al. *Methanopyrus kandleri*, gen. and sp. nov. represents a novel group of hyperthermophilic methanogens, growing at 110 ° C. Arch Microbiol. 1991;156:239-47.

[43] Schut GJ, Lipscomb GL, Han Y, Notey JS, Kelly RM, Adams MMW. The order *Thermococcales* and the family *Thermococcaceae*. The Prokaryotes2014. p. 363-83.

[44] Lipscomb GL, Stirrett K, Schut GJ, Yang F, Jenney FE, Jr., Scott RA, et al. Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. Appl Environ Microbiol. 2011;77(7):2232-8.

[45] Sato T, Fukui T, Atomi H, Imanaka T. Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. Appl Environ Microbiol. 2005;71(7):3889-99.

[46] Wu CH, Schut GJ, Poole FL, 2nd, Haja DK, Adams MWW. Characterization of membrane-bound sulfane reductase: A missing link in the evolution of modern day respiratory complexes. J Biol Chem. 2018;293(43):16687-96.

[47] Bryant FO, Adams MW. Characterization of hydrogenase from the hyperthermophilic archaebacterium, *Pyrococcus furiosus*. J Biol Chem. 1989;264(9):5070-9.

[48] Brileya K, Reysenbach A-L. The class Archaeoglobi. The Prokaryotes2014. p. 15-23.

[49] Steinsbu BO, Thorseth IH, Nakagawa S, Inagaki F, Lever MA, Engelen B, et al. *Archaeoglobus sulfaticallidus* sp. nov., a thermophilic and facultatively lithoautotrophic sulfate-reducer isolated from black rust exposed to hot ridge flank crustal fluids. Int J Syst Evol Microbiol. 2010;60(Pt 12):2745-52.

[50] Oren A. The family Methanothermaceae. The Prokaryotes2014. p. 291-5.

[51] Ver Eecke HC, Butterfield DA, Huber JA, Lilley MD, Olson EJ, Roe KK, et al. Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. Proc Natl Acad Sci U S A. 2012;109(34):13674-9.

[52] Johnson MR, Conners SB, Montero CI, Chou CJ, Shockley KR, Kelly RM. The *Thermotoga maritima* phenotype is impacted by syntrophic interaction with *Methanococcus jannaschii* in hyperthermophilic coculture. Appl Environ Microbiol. 2006;72(1):811-8.

[53] Itoh T. Taxonomy of nonmethanogenic hyperthermophilic and related thermophilic archaea. J Biosci and Bioeng. 2003;96(3):203-12.

[54] Itoh T, Suzuki K, Sanchez PC, Nakase T. *Caldisphaera lagunensis* gen. nov., sp. nov., a novel thermoacidophilic crenarchaeote isolated from a hot spring at Mt Maquiling, Philippines. Int J Syst Evol Microbiol. 2003;53(Pt 4):1149-54.

[55] Prokofeva M, Merkel A, Lebedinsky A, Bonch-Osmolovskaya E. The family *Acidilobaceae*. The Prokaryotes2014. p. 9-14.

[56] Itoh T. The family Thermoproteaceae. The Prokaryotes2014. p. 389-401.

[57] Brock TD, Brock KM, Belly RT, Weiss RL. *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Mikrobiol. 1972;84(1):54-68.

[58] Segerer A, Neuner A, Kristjansson JK, Stetter KO. *Acidianus infernus* gen. nov., sp. nov., and *Acidianus brierleyi* Comb. nov. - facultatively aerobic, extremely acidophilic thermophilic sulfur-metabolizing archaebacteria. Int J Syst Evol Microbiol. 1986;36(4):559-64.

[59] Plumb JJ, Haddad CM, Gibson JaE, Franzmann PD. *Acidianus sulfidivorans* sp. nov., an extremely acidophilic, thermophilic archaeon isolated from a solfatara on Lihir Island, Papua New Guinea, and emendation of the genus description. Int J Syst Evol Microbiol. 2007;57:1418-23.

[60] Kurosawa N, Itoh YH, Iwai T, Sugai A, Uda I, Kimura N, et al. *Sulfurisphaera ohwakuensis* gen. nov., sp. nov., a novel extremely thermophilic acidophile of the order Sulfolobales. Int J Syst Evol Microbiol. 1998;48(2):451-6.

[61] Tsuboi K, Sakai HD, Nur N, Stedman KM, Kurosawa N, Suwanto A. *Sulfurisphaera javensis* sp. nov., a hyperthermophilic and acidophilic archaeon isolated from Indonesian hot spring, and reclassification of *Sulfolobus tokodaii*

Suzuki et al. 2002 as *Sulfurisphaera tokodaii* comb. nov. Int J Syst Evol Microbiol. 2018;68(6):1907-13.

[62] Sakai HD, Kurosawa N. *Saccharolobus caldissimus* gen. nov., sp. nov., a facultatively anaerobic iron-reducing hyperthermophilic archaeon isolated from an acidic terrestrial hot spring, and reclassification of *Sulfolobus solfataricus* as *Saccharolobus solfataricus* comb. nov. and *Sulfolobus shibatae* as *Saccharolobus shibatae* comb. nov. Int J Syst Evol Microbiol. 2018;68(4):1271-8.

[63] Segerer AH, Trincone A, Gahrtz M, Stetter KO. *Stygiolobus azoricus* gen. nov., sp. nov. represents a novel genus of anaerobic, extremely thermoacidophilic archaebacteria of the order Sulfolobales. Int J Syst Bacteriol. 1991;41:495-501.

[64] Sakai HD, Kurosawa N. *Sulfodiicoccus acidiphilus* gen. nov., sp. nov., a sulfur-inhibited thermoacidophilic archaeon belonging to the order Sulfolobales isolated from a terrestrial acidic hot spring. Int J Syst Evol Microbiol. 2017;67(6):1880-6.

[65] Kletzin A, Urich T, Muller F, Bandeiras TM, Gomes CM. Dissimilatory oxidation and reduction of elemental sulfur in thermophilic archaea. J Bioenerg Biomembr. 2004;36(1):77-91.

[66] Kozubal MA, Dlakic M, Macur RE, Inskeep WP. Terminal oxidase diversity and function in "*Metallosphaera yellowstonensis*": gene expression and protein modeling suggest mechanisms of Fe(II) oxidation in the Sulfolobales. Appl Environ Microbiol. 2011;77(5):1844-53.

[67] Auernik KS, Kelly RM. Identification of components of electron transport chains in the extremely thermoacidophilic crenarchaeon *Metallosphaera sedula* through

iron and sulfur compound oxidation transcriptomes. Appl Environ Microbiol. 2008;74(24):7723-32.

[68] Bathe S, Norris PR. Ferrous iron- and sulfur-induced genes in *Sulfolobus metallicus*. Appl Environ Microbiol. 2007;73:2491-7.

[69] Berg IA, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea. Science. 2007;318(5857):1782-6.

[70] Counts JA, Willard DJ, Kelly RM. Life in hot acid: a genome-based reassessment of the archaeal order Sulfolobales. Environ Microbiol. 2020.

[71] She Q, Zhang C, Deng L, Peng N, Chen Z, Liang YX. Genetic analyses in the hyperthermophilic archaeon *Sulfolobus islandicus*. Biochem Soc Trans. 2009;37(1):92-6.

[72] Zhang C, Tian B, Li S, Ao X, Dalgaard K, Gökce S, et al. Genetic manipulation in *Sulfolobus islandicus* and functional analysis of DNA repair genes. Biochem Soc Trans. 2013;41(1):405-10.

[73] Worthington P, Hoang V, Perez-Pomares F, Blum P. Targeted disruption of the α-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Bacteriol. 2003;185(2):482-8.

[74] Wagner M, van Wolferen M, Wagner A, Lassak K, Meyer BH, Reimann J, et al. Versatile genetic tool box for the Crenarchaeote *Sulfolobus acidocaldarius*. Front Microbiol. 2012;3:214-.

[75] Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985;230(4732):1350-4.

[76] Wolk D, Picton E, Johnson D, Davis T, Pancholi P, Ginocchio C, et al. Multicenter evaluation of the Cepheid Xpert methicillin-resistant *Staphylococcus aureus* (MRSA) test as a rapid screening method for detection of MRSA in nares. J Clin Microbiol. 2009;47(3):758-64.

[77] Tindall KR, Kunkel TA. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry. 1988;27(16):6008-13.

[78] Lawyer FC, Stoffel S, Saiki RK, Chang S-Y, Landre PA, Abramson RD, et al. High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. Genome Res. 1993;2(4):275-87.

[79] Potapov V, Ong JL. Examining sources of error in PCR by single-molecule sequencing. PloS One. 2017;12(1):e0169774.

[80] Huang M-M, Arnheim N, Goodman MF. Extension of base mispairs by *Taq* DNA polymerase: implications for single nucleotide discrimination in PCR. Nucleic Acids Res. 1992;20(17):4567-73.

[81] Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. Nucleic Acids Res. 2004;32(3):1197-207.

[82] Sarmiento F, Peralta R, Blamey JM. Cold and hot extremozymes: industrial relevance and current trends. Front Bioeng Biotech. 2015;3:148.

[83] Dumorné K, Córdova DC, Astorga-Eló M, Renganathan P. Extremozymes: a potential source for industrial applications. J Microbiol Biotechnol. 2017;27(4):649-59.

[84] Anjum RS, Bray SM, Blackwood JK, Kilkenny ML, Coelho MA, Foster BM, et al. Involvement of a eukaryotic-like ubiquitin-related modifier in the proteasome pathway of the archaeon *Sulfolobus acidocaldarius*. Nat Commun. 2015;6(1):8163.

[85] Maupin-Furlow J. Proteasomes and protein conjugation across domains of life. Nat Rev Microbiol. 2012;10(2):100-11.

[86] Bar-Nun S, Glickman MH. Proteasomal AAA-ATPases: structure and function. Biochim Biophys Acta Mol Cell Res. 2012;1823(1):67-82.

[87] Blumentals II, Robinson AS, Kelly RM. Characterization of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Appl Environ Microbiol. 1990;56(7):1992-8.

[88] Halio SB, Blumentals II, Short SA, Merrill BM, Kelly RM. Sequence, expression in *Escherichia coli*, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon *Pyrococcus furiosus*. J Bacteriol. 1996;178(9):2605-12.

[89] Halio SB, Bauer MW, Mukund S, Adams M, Kelly RM. Purification and characterization of two functional forms of intracellular protease PfpI from the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl Environ Microbiol. 1997;63(1):289-95.

[90] Du X, Choi I-G, Kim R, Wang W, Jancarik J, Yokota H, et al. Crystal structure of an intracellular protease from *Pyrococcus horikoshii* at 2-Å resolution. Proc Natl Acad Sci U S A. 2000;97(26):14079-84.

[91] Larson SB, McPherson A. The structure of the Pfp1 protease from the hyperthermophilic archaeon *Thermococcus thioreducens* in two crystal forms. Acta Crystallogr D. 2017;73(9):749-56.

[92] Snowden LJ, Blumentals II, Kelly RM. Regulation of proteolytic activity in the hyperthermophile *Pyrococcus furiosus*. Appl Environ Microbiol. 1992;58(4):1134-41.

[93] Bandyopadhyay S, Cookson MR. Evolutionary and functional relationships within the DJ1 superfamily. BMC Evol Biol. 2004;4(1):6.

[94] Wilson MA, Collins JL, Hod Y, Ringe D, Petsko GA. The 1.1-Å resolution crystal structure of DJ-1, the protein mutated in autosomal recessive early onset Parkinson's disease. Proc Natl Acad Sci U S A. 2003;100(16):9256-61.

[95] Olzmann JA, Brown K, Wilkinson KD, Rees HD, Huai Q, Ke H, et al. Familial Parkinson's disease-associated L166P mutation disrupts DJ-1 protein folding and function. J Biol Chem. 2004;279(9):8506-15.

[96] Le Naour F, Misek DE, Krause MC, Deneux L, Giordano TJ, Scholl S, et al. Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. Clin Cancer Res. 2001;7(11):3328-35.

[97] An C-N, Jiang H, Wang Q, Yuan R-P, Liu J-M, Shi W-L, et al. Down-regulation of DJ-1 protein in the ejaculated spermatozoa from Chinese asthenozoospermia patients. Fertil Steril. 2011;96(1):19-23. e2.

[98] Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res. 2005;15(3):352-63.

[99] Foophow T, Tanaka S-I, Koga Y, Takano K, Kanaya S. Subtilisin-like serine protease from hyperthermophilic archaeon *Thermococcus kodakaraensis* with N- and C-terminal propeptides. Protein Eng Des Sel. 2010;23(5):347-55.

[100] Xie G, Shao Z, Zong L, Li X, Cong D, Huo R. Heterologous expression and characterization of a novel subtilisin-like protease from a thermophilic *Thermus thermophilus* HB8. Int J Biol Macromol. 2019;138:528-35.

[101] Voorhorst WG, Eggen RI, Geerling AC, Platteeuw C, Siezen RJ, de Vos WM. Isolation and characterization of the hyperthermostable serine protease, pyrolysin, and its gene from the hyperthermophilic archaeon *Pyrococcus furiosus*. J Biol Chem. 1996;271(34):20426-31.

[102] Gao X, Zeng J, Yi H, Zhang F, Tang B, Tang X-F. Four inserts within the catalytic domain confer extra stability and activity to hyperthermostable pyrolysin from *Pyrococcus furiosus*. Appl Environ Microbiol. 2017;83(5):e03228-16.

[103] Lin X-l, Tang J. Purification, characterization, and gene cloning of thermopsin, a thermostable acid protease from *Sulfolobus acidocaldarius*. J Biol Chem. 1990;265(3):1490-5.

[104] Fusek M, Lin X-l, Tang J. Enzymic properties of thermopsin. J Biol Chem. 1990;265(3):1496-501.

[105] Gogliettino M, Riccio A, Cocca E, Rossi M, Palmieri G, Balestrieri M. A new pepstatin-insensitive thermopsin-like protease overproduced in peptide-rich cultures of *Sulfolobus solfataricus*. Int J Mol Sci. 2014;15(2):3204-19.

[106] Rawlings ND. Evolution of the thermopsin peptidase family (A5). PLoS One. 2013;8(11):e78998.

[107] Mehta D, Satyanarayana T. Bacterial and archaeal α-amylases: diversity and amelioration of the desirable characteristics for industrial applications. Front Microbiol. 2016;7:1129.

[108] Lim SJ, Oslan SNH, Oslan SN. Purification and characterisation of thermostable α-amylases from microbial sources. Bioresources. 2020;15(1):2005-29.

[109] George R, Georrge JJ. Thermostable alpha-amylase and its activity, stability and industrial relevance studies. Stability and Industrial Relevance Studies 2020.

[110] Rollings JE, Thompson RW. Kinetics of enzymatic starch liquefaction: simulation of the high-molecular-weight product distribution. Biotechnol Bioeng. 1984;26(12):1475-84.

[111] Chung YC, Kobayashi T, Kanai H, Akiba T, Kudo T. Purification and properties of extracellular amylase from the hyperthermophilic archaeon *Thermococcus profundus* DT5432. Appl Environ Microbiol. 1995;61(4):1502-6.

[112] Janeček Š, Lévêque E, Belarbi A, Haye B. Close evolutionary relatedness of α-amylases from archaea and plants. J Mol Evol. 1999;48(4):421-6.

[113] Dong G, Vieille C, Zeikus JG. Cloning, sequencing, and expression of the gene encoding amylopullulanase from *Pyrococcus furiosus* and biochemical

characterization of the recombinant enzyme. Appl Environ Microbiol. 1997;63(9):3577-84.

[114] Haseltine C, Rolfsmeier M, Blum P. The glucose effect and regulation of alpha-amylase synthesis in the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Bacteriol. 1996;178(4):945-50.

[115] Liebl W, Stemplinger I, Ruile P. Properties and gene structure of the *Thermotoga maritima* alpha-amylase AmyA, a putative lipoprotein of a hyperthermophilic bacterium. J Bacteriol. 1997;179(3):941-8.

[116] Ballschmiter M, Fütterer O, Liebl W. Identification and characterization of a novel intracellular alkaline α-amylase from the hyperthermophilic bacterium *Thermotoga maritima* MSB8. Appl Environ Microbiol. 2006;72(3):2206-11.

[117] Ozturk H, Ece S, Gundeger E, Evran S. Site-directed mutagenesis of methionine residues for improving the oxidative stability of α-amylase from *Thermotoga maritima*. J Biosci Bioeng. 2013;116(4):449-51.

[118] Lim JK, Lee HS, Kim YJ, Bae SS, Jeon JH, Kang SG, et al. Critical factors to high thermostability of an alpha-amylase from hyperthermophilic archaeon *Thermococcus onnurineus* NA1. J Microbiol Biotechnol. 2007;17(8):1242-8.

[119] Zhang X, Leemhuis H, Janeček Š, Martinovičová M, Pijning T, van der Maarel MJ. Identification of *Thermotoga maritima* MSB8 GH57 α-amylase AmyC as a glycogen-branching enzyme with high hydrolytic activity. Appl Microbiol Biotechnol. 2019;103(15):6141-51.

[120] Jung J-H, Seo D-H, Holden JF, Park C-S. Maltose-forming α-amylase from the hyperthermophilic archaeon *Pyrococcus* sp. ST04. Appl Microbiol Biotechnol. 2014;98(5):2121-31.

[121] Lange JP. Lignocellulose conversion: an introduction to chemistry, process and economics. Biofuel Bioprod Biorefin. 2007;1(1):39-48.

[122] Martin K, McDougall BM, McIlroy S, Jayus, Chen J, Seviour RJ. Biochemistry and molecular biology of exocellular fungal β-(1, 3)-and β-(1, 6)-glucanases. FEMS Microbiol Rev. 2007;31(2):168-92.

[123] Tharanathan RN, Kittur FS. Chitin—the undisputed biomolecule of great potential. Crit Rev Food Sci Nutr. 2003;43(1):61-87.

[124] Zhu F, Du B, Xu B. A critical review on production and industrial applications of beta-glucans. Food Hydrocoll. 2016;52:275-88.

[125] Patel AK, Singhania RR, Sim SJ, Pandey A. Thermostable cellulases: current status and perspectives. Bioresour Technol. 2019;279:385-92.

[126] Turner P, Mamo G, Karlsson EN. Potential and utilization of thermophiles and thermostable enzymes in biorefining. Microb Cell Fact. 2007;6(1):9.

[127] Yang X, Ma R, Shi P, Huang H, Bai Y, Wang Y, et al. Molecular characterization of a highly-active thermophilic β-glucosidase from *Neosartorya fischeri* P1 and its application in the hydrolysis of soybean isoflavone glycosides. PLoS One. 2014;9(9):e106785.

[128] Conway JM, McKinley BS, Seals NL, Hernandez D, Khatibi PA, Poudel S, et al. Functional analysis of the glucan degradation locus in *Caldicellulosiruptor bescii*

reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. Appl Environ Microbiol. 2017;83(24):e01828-17.

[129] Zhang X, Tu B, Dai L-r, Lawson PA, Zheng Z-z, Liu L-Y, et al. *Petroclostridium xylanilyticum* gen. nov., sp. nov., a xylan-degrading bacterium isolated from an oilfield, and reclassification of clostridial cluster III members into four novel genera in a new *Hungateiclostridiaceae* fam. nov. Int J Syst Evol Microbiol. 2018;68(10):3197-211.

[130] Zverlov VV, Kellermann J, Schwarz WH. Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. Proteomics. 2005;5(14):3646-53.

[131] Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, et al. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science. 2013;342(6165):1513-6.

[132] Conway JM, Crosby JR, McKinley BS, Seals NL, Adams MW, Kelly RM. Parsing in vivo and in vitro contributions to microcrystalline cellulose hydrolysis by multidomain glycoside hydrolases in the *Caldicellulosiruptor bescii* secretome. Biotechnol Bioeng. 2018;115(10):2426-40.

[133] Johnson EA, Sakajoh M, Halliwell G, Madia A, Demain AL. Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. Appl Environ Microbiol. 1982;43(5):1125-32.

[134] Gilbert HJ. Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. Mol Microbiol. 2007;63(6):1568-76.

[135] Schimming S, Schwarz WH, Staudenbauer WL. Properties of a thermoactive β-1, 3-1, 4-glucanase (lichenase) from *Clostridium thermocellum* expressed in *Escherichia coli*. Biochem Biophys Res Commun. 1991;177(1):447-52.

[136] Chen C-C, Huang J-W, Zhao P, Ko T-P, Huang C-H, Chan H-C, et al. Structural analyses and yeast production of the β-1, 3-1, 4-glucanase catalytic module encoded by the licB gene of *Clostridium thermocellum*. Enzyme Microb Technol. 2015;71:1-7.

[137] Bleicher L, Prates ET, Gomes TC, Silveira RL, Nascimento AS, Rojas AL, et al. Molecular basis of the thermostability and thermophilicity of laminarinases: x-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. J Phys Chem B. 2011;115(24):7940-9.

[138] Shallom D, Shoham Y. Microbial hemicellulases. Curr Opin Microbiol. 2003;6(3):219-28.

[139] Collins T, Gerday C, Feller G. Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol Rev. 2005;29(1):3-23.

[140] Paës G, Berrin JG, Beaugrand J. GH11 xylanases: structure/function/properties relationships and applications. Biotechnol Adv. 2012;30(3):564-92.

[141] Biely P, Vršanská M, Tenkanen M, Kluepfel D. Endo-β-1, 4-xylanase families: differences in catalytic properties. J Biotechnol. 1997;57(1-3):151-66.

[142] Anderson I, Scheuner C, Göker M, Mavromatis K, Hooper SD, Porat I, et al. Novel insights into the diversity of catabolic metabolism from ten haloarchaeal genomes. PLoS One. 2011;6(5):e20237.

[143] Cannio R, Di Prizito N, Rossi M, Morana A. A xylan-degrading strain of *Sulfolobus solfataricus*: isolation and characterization of the xylanase activity. Extremophiles. 2004;8(2):117-24.

[144] Uhl AM, Daniel RM. The first description of an archaeal hemicellulase: the xylanase from *Thermococcus zilligii* strain AN1. Extremophiles. 1999;3(4):263-7.

[145] Notenboom V, Boraston AB, Kilburn DG, Rose DR. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. Biochemistry. 2001;40(21):6248-56.

[146] Jia X, Han Y. The extracellular endo-β-1,4-xylanase with multidomain from the extreme thermophile *Caldicellulosiruptor lactoaceticus* is specific for insoluble xylan degradation. Biotechnol Biofuels. 2019;12(1):143.

[147] Conway JM, Crosby JR, Hren AP, Southerland RT, Lee LL, Lunin VV, et al. Novel multidomain, multifunctional glycoside hydrolases from highly lignocellulolytic *Caldicellulosiruptor* species. AIChE J. 2018;64(12):4218-28.

[148] Tajwar R, Shahid S, Zafar R, Akhtar MW. Impact of orientation of carbohydrate binding modules family 22 and 6 on the catalytic activity of *Thermotoga maritima* xylanase XynB. Enzyme Microb Technol. 2017;106:75-82.

[149] Su X, Han Y, Dodd D, Moon YH, Yoshida S, Mackie RI, et al. Reconstitution of a thermostable xylan-degrading enzyme mixture from the bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol. 2013;79(5):1481-90.

[150] Jia X, Qiao W, Tian W, Peng X, Mi S, Su H, et al. Biochemical characterization of extra- and intracellular endoxylanse from thermophilic bacterium *Caldicellulosiruptor kronotskyensis*. Sci Rep. 2016;6:21672.

[151] Zhang Y, An J, Yang G, Zhang X, Xie Y, Chen L, et al. Structure features of GH10 xylanase from *Caldicellulosiruptor bescii*: implication for its thermophilic adaption and substrate binding preference. Acta Biochim Biophys Sin (Shanghai). 2016;48(10):948-57.

[152] Yu T, Anbarasan S, Wang Y, Telli K, Aslan AS, Su Z, et al. Hyperthermostable *Thermotoga maritima* xylanase XYN10B shows high activity at high temperatures in the presence of biomass-dissolving hydrophilic ionic liquids. Extremophiles. 2016;20(4):515-24.

[153] Zhengqiang J, Kobayashi A, Ahsan MM, Lite L, Kitaoka M, Hayashi K. Characterization of a thermostable family 10 endo-xylanase (XynB) from *Thermotoga maritima* that cleaves p-nitrophenyl-β-D-xyloside. J Biosci Bioeng. 2001;92(5):423-8.

[154] Duffaud GD, McCutchen CM, Leduc P, Parker KN, Kelly RM. Purification and characterization of extremely thermostable beta-mannanase, beta-mannosidase, and alpha-galactosidase from the hyperthermophilic eubacterium *Thermotoga neapolitana* 5068. Appl Environ Microbiol. 1997;63(1):169-77.

[155] Parker KN, Chhabra SR, Lam D, Callen W, Duffaud GD, Snead MA, et al. Galactomannanases Man2 and Man5 from Thermotogaspecies: Growth physiology on galactomannans, gene sequence analysis, and biochemical properties of recombinant enzymes. Biotechnol Bioeng. 2001;75(3):322-33.

[156] Aulitto M, Fusco S, Limauro D, Fiorentino G, Bartolucci S, Contursi P. Galactomannan degradation by thermophilic enzymes: a hot topic for

biotechnological applications. World J of Microbiol and Biotechnol. 2019;35(2):32.

[157] Srivastava PK, Kapoor M. Production, properties, and applications of endo-β-mannanases. Biotechnol Adv. 2017;35(1):1-19.

[158] Benedetti M, Vecchi V, Betterle N, Natali A, Bassi R, Dall'Osto L. Design of a highly thermostable hemicellulose-degrading blend from *Thermotoga neapolitana* for the treatment of lignocellulosic biomass. J Biotechnol. 2019;296:42-52.

[159] Mhiri S, Bouanane-Darenfed A, Jemli S, Neifar S, Ameri R, Mezghani M, et al. A thermophilic and thermostable xylanase from *Caldicoprobacter algeriensis*: Recombinant expression, characterization and application in paper biobleaching. Int J Biol Macromol. 2020.

[160] Comfort DA, Chhabra SR, Conners SB, Chou C-J, Epting KL, Johnson MR, et al. Strategic biocatalysis with hyperthermophilic enzymes. Green Chemistry. 2004;6(9):459-65.

[161] Hu K, Li C-X, Pan J, Ni Y, Zhang X-Y, Xu J-H. Performance of a new thermostable mannanase in breaking guar-based fracturing fluids at high temperatures with little premature degradation. Appl Biochem Biotechnol. 2014;172(3):1215-26.

[162] Lemaire M, Ohayon H, Gounon P, Fujino T, Béguin P. OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope. J Bacteriol. 1995;177(9):2451-9.

[163] Ali MK, Kimura T, Sakka K, Ohmiya K. The multidomain xylanase Xyn10B as a cellulose-binding protein in *Clostridium stercorarium*. FEMS Microbiol Lett. 2001;198(1):79-83.

[164] Franche N, Tardif C, Ravachol J, Harchouni S, Ferdinand P-H, Borne R, et al. Cel5I, a SLH-containing glycoside hydrolase: characterization and investigation on its role in *Ruminiclostridium cellulolyticum*. PLoS One. 2016;11(8):e0160812.

[165] Liu S-Y, Gherardini FC, Matuschek M, Bahl H, Wiegel J. Cloning, sequencing, and expression of the gene encoding a large S-layer-associated endoxylanase from *Thermoanaerobacterium* sp. strain JW/SL-YS 485 in *Escherichia coli*. J Bacteriol. 1996;178(6):1539-47.

[166] Waeonukul R, Pason P, Kyu KL, Sakka K, Kosugi A, Mori Y, et al. Cloning, sequencing, and expression of the gene encoding a multidomain endo-β-1, 4-xylanase from *Paenibacillus curdlanolyticus* B-6, and characterization of the recombinant enzyme. J Microbiol Biotechnol. 2009;19(3):277-85.

[167] Ozdemir I, Blumer-Schuette SE, Kelly RM. 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. 2012;78(3):768-77.

[168] Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, et al. Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. J Biol Chem. 2016;291(13):6732-47.

[169] Liebl W, Winterhalter C, Baumeister W, Armbrecht M, Valdez M. Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. J Bacteriol. 2008;190(4):1350-8.

[170] Bhosale SH, Rao MB, Deshpande VV. Molecular and industrial aspects of glucose isomerase. Microbio Rev. 1996;60(2):280-300.

[171] Jia D-X, Zhou L, Zheng Y-G. Properties of a novel thermostable glucose isomerase mined from *Thermus oshimai* and its application to preparation of high fructose corn syrup. Enzyme Microb Technol. 2017;99:1-8.

[172] Fatima B, Javed MM. Production, purification and physicochemical characterization of D-xylose/glucose isomerase from *Escherichia coli* strain BL21. 3 Biotech. 2020;10(2):39.

[173] Liu Z-Q, Zheng W, Huang J-F, Jin L-Q, Jia D-X, Zhou H-Y, et al. Improvement and characterization of a hyperthermophilic glucose isomerase from *Thermoanaerobacter ethanolicus* and its application in production of high fructose corn syrup. J Ind Microbiol Biotechnol. 2015;42(8):1091-103.

[174] Hartley BS, Hanlon N, Jackson RJ, Rangarajan M. Glucose isomerase: insights into protein engineering for increased thermostability. Biochim Biophys Acta Protein Struct Mol Enzymol. 2000;1543(2):294-335.

[175] Starnes RL, Kelly RM, Brown SH, inventorsMethod for glucose isomerization using xylose isomerase purified from *Thermotoga Maritima* and *Thermotoga Neapolitana*1993 issued December, 1993.

[176] Bandlish RK, Michael Hess J, Epting KL, Vieille C, Kelly RM. Glucose-to-fructose conversion at high temperatures with xylose (glucose) isomerases from *Streptomyces murinus* and two hyperthermophilic *Thermotoga* species. Biotechnol Bioeng. 2002;80(2):185-94.

[177] Hawkins AB, Lian H, Zeldes BM, Loder AJ, Lipscomb GL, Schut GJ, et al. Bioprocessing analysis of *Pyrococcus furiosus* strains engineered for CO_2-based 3-hydroxypropionate production. Biotechnol Bioeng. 2015;112(8):1533-43.

[178] Loder AJ, Han Y, Hawkins AB, Lian H, Lipscomb GL, Schut GJ, et al. Reaction kinetic analysis of the 3-hydroxypropionate/4-hydroxybutyrate CO_2 fixation cycle in extremely thermoacidophilic archaea. Metab Eng. 2016;38:446-63.

[179] Keller MW, Schut GJ, Lipscomb GL, Menon AL, Iwuchukwu IJ, Leuko TT, et al. Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. Proc Natl Acad Sci U S A. 2013;110(15):5840-5.

[180] Erb TJ. Carboxylases in natural and synthetic microbial pathways. Appl Environ Microbiol. 2011;77(24):8466-77.

[181] von Borzyskowski LS, Rosenthal RG, Erb TJ. Evolutionary history and biotechnological future of carboxylases. J Biotechnol. 2013;168(3):243-51.

[182] Ezaki S, Maeda N, Kishimoto T, Atomi H, Imanaka T. Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, *Pyrococcus kodakaraensis* KOD1. J Biol Chem. 1999;274(8):5078-82.

[183] Yoshida S, Inui M, Yukawa H, Kanao T, Tomizawa KI, Atomi H, et al. Phototrophic growth of a Rubisco-deficient mesophilic purple nonsulfur bacterium harboring a Type III Rubisco from a hyperthermophilic archaeon. J Biotechnol. 2006;124(3):532-44.

[184] Nishitani Y, Yoshida S, Fujihashi M, Kitagawa K, Doi T, Atomi H, et al. Structure-based Catalytic Optimization of a Type III Rubisco from a Hyperthermophile. J Biol Chem. 2010;285(50):39339-47.

[185] Fujihashi M, Nishitani Y, Kiriyama T, Aono R, Sato T, Takai T, et al. Mutation design of a thermophilic Rubisco based on three-dimensional structure enhances its activity at ambient temperature. Proteins. 2016;84(10):1339-46.

[186] Pearson A, Hurley SJ, Elling FJ, Wilkes EB. CO_2-dependent carbon isotope fractionation in archaea, part I: modeling the 3HP/4HB pathway. Geochim Cosmochim Ac. 2019;261:368-82.

[187] Littlechild JA. Improving the 'tool box' for robust industrial enzymes. J Ind Microbiol Biotechnol. 2017;44(4-5):711-20.

[188] Jo BH, Seo JH, Cha HJ. Bacterial extremo-alpha-carbonic anhydrases from deep-sea hydrothermal vents as potential biocatalysts for CO_2 sequestration. J Mol Catal B-Enzym. 2014;109:31-9.

[189] Capasso C, De Luca V, Carginale V, Cannio R, Rossi M. Biochemical properties of a novel and highly thermostable bacterial alpha-carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* YO3AOP1. J Enzym Inhib Med Ch. 2012;27(6):892-7.

[190] Del Prete S, Perfetto R, Rossi M, Alasmary FAS, Osman SM, AlOthman Z, et al. A one-step procedure for immobilising the thermostable carbonic anhydrase (SspCA) on the surface membrane of *Escherichia coli*. J Enzym Inhib Med Ch. 2017;32(1):1120-8.

[191] Perfetto R, Del Prete S, Vullo D, Sansone G, Barone CMA, Rossi M, et al. Production and covalent immobilisation of the recombinant bacterial carbonic anhydrase (SspCA) onto magnetic nanoparticles. J Enzym Inhib Med Ch. 2017;32(1):759-66.

[192] Abdelrahim MYM, Martins CF, Neves LA, Capasso C, Supuran CT, Coelhoso IM, et al. Supported ionic liquid membranes immobilized with carbonic anhydrases for CO_2 transport at high temperatures. J Memb Sci. 2017;528:225-30.

[193] Effendi SSW, Chiu CY, Chang YK, Ng IS. Crosslinked on novel nanofibers with thermophilic carbonic anhydrase for carbon dioxide sequestration. Int J Biol Macromol. 2020;152:930-8.

[194] Hsu KP, Tan SI, Chiu CY, Chang YK, Ng IS. ARduino-pH Tracker and screening platform for characterization of recombinant carbonic anhydrase in *Escherichia coli*. Biotechnol Progr. 2019;35(5).

[195] Bharatiy SK, Hazra M, Paul M, Mohapatra S, Samantaray D, Dubey RC, et al. In silico designing of an industrially sustainable carbonic anhydrase using molecular dynamics simulation. Acs Omega. 2016;1(6):1081-103.

[196] Savile CK, Lalonde JJ. Biotechnology for the acceleration of carbon dioxide capture and sequestration. Curr Opin Biotechnol. 2011;22(6):818-23.

[197] Bhattacharya S, inventorEnzyme facilitated solubilization of carbon dioxide from emission streams in novel attachable reactors/devices2004 filed December 24, 2004.

[198] Borchert MS, inventorHeat-stable persephonella carbonic anhydrases and their use 2018 issued March 6, 2018.

[199] Cha HJ, Jo BH, Seo JH, inventorsCarbonic anhydrase with stability at high temperature and capturring agent for carbon dioxide comprising the same2016 filed August 4, 2016.

[200] Daigle R, Madore É, Fradette S, inventorsTechniques for CO_2 capture using *Sulfurihydrogenibium* sp. carbonic anhydrase2018 issued May 15, 2018.

[201] Newman LM, Clark L, Ching C, Zimmerman S, inventorsCarbonic anhydrase polypeptides and uses thereof2010 filed August 19, 2010.

[202] Saunders P, Salmon S, Borchert M, inventorsModular reactor and process for carbon-dioxide extraction 2011 filed July 21, 2011.

[203] Lian H, Zeldes BM, Lipscomb GL, Hawkins AB, Han YJ, Loder AJ, et al. Ancillary contributions of heterologous biotin protein ligase and carbonic anhydrase for CO_2 incorporation into 3-hydroxypropionate by metabolically engineered *Pyrococcus furiosus*. Biotechnol Bioeng. 2016;113(12):2652-60.

[204] Adams MW, Kelly RM, Hawkins AB, Menon AL, Lipscomb GLP, Schut GJ, inventorsSequestration of carbon dioxide with hydrogen to useful products2017 issued March 7, 2017.

[205] Del Prete S, De Luca V, Capasso C, Supuran CT, Carginale V. Recombinant thermoactive phosphoenolpyruvate carboxylase (PEPC) from *Thermosynechococcus elongatus* and its coupling with mesophilic/thermophilic bacterial carbonic anhydrases (CAs) for the conversion of CO_2 to oxaloacetate. Bioorgan Med Chem. 2016;24(2):220-5.

[206] Lubitz W, Ogata H, Rüdiger O, Reijerse E. Hydrogenases. Chem Rev. 2014;114(8):4081-148.

[207] Schuchmann K, Chowdhury NP, Müller V. Complex Multimeric [FeFe] Hydrogenases: Biochemistry, Physiology and New Opportunities for the Hydrogen Economy. Front Microbiol. 2018;9(2911).

[208] Kim DH, Kim MS. Hydrogenases for biological hydrogen production. Bioresour Technol. 2011;102(18):8423-31.

[209] Vignais PM, Billoud B. Occurrence, classification, and biological function of hydrogenases: an overview. Chem Rev. 2007;107(10):4206-72.

[210] Ma K, Weiss R, Adams MW. Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction. J Bacteriol. 2000;182(7):1864-71.

[211] McTernan PM, Chandrayan SK, Wu CH, Vaccaro BJ, Lancaster WA, Yang Q, et al. Intact functional fourteen-subunit respiratory membrane-bound [NiFe]-hydrogenase complex of the hyperthermophilic archaeon *Pyrococcus furiosus*. J Biol Chem. 2014;289(28):19364-72.

[212] Schut GJ, Nixon WJ, Lipscomb GL, Scott RA, Adams MWW. Mutational analyses of the enzymes involved in the metabolism of hydrogen by the hyperthermophilic archaeon *Pyrococcus furiosus*. Front Microbiol [Internet]. 2012 2012; 3:[163 p.]. Available from: http://europepmc.org/abstract/MED/22557999

https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/22557999/pdf/?tool=EBI

https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/22557999/?tool=EBI

https://doi.org/10.3389/fmicb.2012.00163

https://europepmc.org/articles/PMC3341082

https://europepmc.org/articles/PMC3341082?pdf=render.

[213] Schut GJ, Adams MWW. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. J Bacteriol. 2009;191(13):4451-7.

[214] Cha M, Chung D, Elkins JG, Guss AM, Westpheling J. Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass. Biotechnol Biofuels. 2013;6(1):85.

[215] Juszczak A, Aono S, Adams MW. The extremely thermophilic eubacterium, *Thermotoga maritima*, contains a novel iron-hydrogenase whose cellular activity is dependent upon tungsten. J Biol Chem. 1991;266(21):13834-41.

[216] Pawar SS, van Niel EW. Thermophilic biohydrogen production: how far are we? Appl Microbiol Biotechnol. 2013;97(18):7999-8009.

[217] Thauer RK, Jungermann K, Decker K. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev. 1977;41(1):100-80.

[218] Lipscomb GL, Schut GJ, Thorgersen MP, Nixon WJ, Kelly RM, Adams MW. Engineering hydrogen gas production from formate in a hyperthermophile by heterologous production of an 18-subunit membrane-bound complex. J Biol Chem. 2014;289(5):2873-9.

[219] Singh R, White D, Demirel Y, Kelly R, Noll K, Blum P. Uncoupling fermentative synthesis of molecular hydrogen from biomass formation in *Thermotoga maritima*. Appl Environ Microbiol. 2018;84(17):e00998-18.

[220] Wu C-H, McTernan PM, Walter ME, Adams MWW. Production and application of a soluble hydrogenase from *Pyrococcus furiosus*. Archaea. 2015;2015:912582.

[221] Sun J, Hopkins RC, Jenney Jr FE, McTernan PM, Adams MW. Heterologous expression and maturation of an NADP-dependent [NiFe]-hydrogenase: a key enzyme in biofuel production. PLoS One. 2010;5(5):e10526.

[222] Ye X, Wang Y, Hopkins RC, Adams MW, Evans BR, Mielenz JR, et al. Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails. Chempluschem. 2009;2(2):149-52.

[223] Martín del Campo JS, Rollin J, Myung S, Chun Y, Chandrayan S, Patiño R, et al. High-yield production of dihydrogen from xylose by using a synthetic enzyme cascade in a cell-free system. Angew Chem Int Ed Engl. 2013;52(17):4587-90.

[224] Rollin JA, Martin del Campo J, Myung S, Sun F, You C, Bakovic A, et al. High-yield hydrogen production from biomass by in vitro metabolic engineering: Mixed sugars coutilization and kinetic modeling. Proc Natl Acad Sci U S A. 2015;112(16):4964-9.

[225] Kim JE, Kim EJ, Chen H, Wu CH, Adams MWW, Zhang YP. Advanced water splitting for green hydrogen gas production through complete oxidation of starch by in vitro metabolic engineering. Metab Eng. 2017;44:246-52.

[226] Chica B, Wu C-H, Liu Y, Adams MW, Lian T, Dyer RB. Balancing electron transfer rate and driving force for efficient photocatalytic hydrogen production in CdSe/CdS nanorod–[NiFe] hydrogenase assemblies. Energy Environ Sci. 2017;10(10):2245-55.

[227] Sanchez ML, Wu C-H, Adams MW, Dyer RB. Optimizing electron transfer from CdSe QDs to hydrogenase for photocatalytic H_2 production. Chem Commun. 2019;55(39):5579-82.

[228] Sato T, Fukui T, Atomi H, Imanaka T. Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J Bacteriol. 2003;185(1):210-20.

[229] Chung D, Farkas J, Huddleston JR, Olivar E, Westpheling J. Methylation by a unique α-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. PLoS One. 2012;7(8):e43844.

[230] Waege I, Schmid G, Thumann S, Thomm M, Hausner W. Shuttle vector-based transformation system for *Pyrococcus furiosus*. Appl Environ Microbiol. 2010;76(10):3308-13.

[231] Han D, Xu Z. Development of a *pyrE*-based selective system for *Thermotoga sp.* strain RQ7. Extremophiles. 2017;21(2):297-306.

[232] Santangelo TJ, Čuboňová Lu, Reeve JN. *Thermococcus kodakarensis* genetics: TK1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. Appl Environ Microbiol. 2010;76(4):1044-52.

[233] Zhang C, She Q, Bi H, Whitaker RJ. The apt/6-methylpurine counterselection system and its applications in genetic studies of the hyperthermophilic archaeon *Sulfolobus islandicus*. Appl Environ Microbiol. 2016;82(10):3070-81.

[234] Birien T, Thiel A, Henneke G, Flament D, Moalic Y, Jebbar M. Development of an effective 6-methylpurine counterselection marker for genetic manipulation in *Thermococcus barophilus*. Genes. 2018;9(2).

[235] Albers S-V, Driessen AJM. Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. Archaea. 2007;2(3):145-9.

[236] Farkas J, Stirrett K, Lipscomb GL, Nixon W, Scott RA, Adams MW, et al. Recombinogenic properties of *Pyrococcus furiosus* strain COM1 enable rapid selection of targeted mutants. Appl Environ Microbiol. 2012;78(13):4669-76.

[237] Lipscomb GL, Conway JM, Blumer-Schuette SE, Kelly RM, Adams MWW. A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii*. Appl Environ Microbiol. 2016;82(14):4421-8.

[238] Williams-Rhaesa AM, Poole FL, Dinsmore JT, Lipscomb GL, Rubinstein GM, Scott IM, et al. Genome stability in engineered strains of the extremely thermophilic lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol. 2017;83(14).

[239] Zheng T, Huang Q, Zhang C, Ni J, She Q, Shen Y. Development of a simvastatin selection marker for a hyperthermophilic acidophile, *Sulfolobus islandicus*. Appl Environ Microbiol. 2012;78(2):568-74.

[240] Zhang C, Whitaker RJ. A broadly applicable gene knockout system for the thermoacidophilic archaeon *Sulfolobus islandicus* based on simvastatin selection. Microbiology. 2012;158(Pt 6):1513-22.

[241] Matsumi R, Manabe K, Fukui T, Atomi H, Imanaka T. Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance. J Bacteriol. 2007;189(7):2683-91.

[242] Lubelska JM, Jonuscheit M, Schleper C, Albers S-V, Driessen AJM. Regulation of expression of the arabinose and glucose transporter genes in the thermophilic archaeon *Sulfolobus solfataricus*. Extremophiles. 2006;10(5):383-91.

[243] Berkner S, Wlodkowski A, Albers SV, Lipps G. Inducible and constitutive promoters for genetic systems in *Sulfolobus acidocaldarius*. Extremophiles. 2010;14(3):249-59.

[244] van der Kolk N, Wagner A, Wagner M, Waßmer B, Siebers B, Albers S-V. Identification of XylR, the activator of arabinose/xylose inducible regulon in *Sulfolobus acidocaldarius* and its application for homologous protein expression. Front Microbiol. 2020;11:1066.

[245] Williams-Rhaesa AM, Awuku NK, Lipscomb GL, Poole FL, Rubinstein GM, Conway JM, et al. Native xylose-inducible promoter expands the genetic tools for the biomass-degrading, extremely thermophilic bacterium *Caldicellulosiruptor bescii*. Extremophiles. 2018;22(4):629-38.

[246] Speed MC, Burkhart BW, Picking JW, Santangelo TJ. An archaeal fluoride-responsive riboswitch provides an inducible expression system for hyperthermophiles. Appl Environ Microbiol. 2018;84(7).

[247] Santangelo TJ, Čuboňová Lu, Matsumi R, Atomi H, Imanaka T, Reeve JN. Polarity in archaeal operon transcription in *Thermococcus kodakaraensis*. J Bacteriol. 2008;190(6):2244-8.

[248] Prangishvili DA, Vashakidze RP, Chelidze MG, Gabriadze I. A restriction endonuclease SuaI from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. FEBS Lett. 1985;192(1):57-60.

[249] Grogan DW. Cytosine methylation by the SuaI restriction-modification system: implications for genetic fidelity in a hyperthermophilic archaeon. J Bacteriol. 2003;185(15):4657-61.

[250] Chung D-H, Huddleston JR, Farkas J, Westpheling J. Identification and characterization of CbeI, a novel thermostable restriction enzyme from

Caldicellulosiruptor bescii DSM 6725 and a member of a new subfamily of HaeIII-like enzymes. J Ind Microbiol Biotechnol. 2011;38(11):1867.

[251] Stedman KM, Schleper C, Rumpf E, Zillig W. Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. Genetics. 1999;152(4):1397-405.

[252] Chung D, Sarai NS, Himmel ME, Bomble YJ. Genetics of unstudied thermophiles for industry. In: Himmel ME, Bomble YJ, editors. Metabolic Pathway Engineering: Springer; 2020. p. 5-19.

[253] Basen M, Schut GJ, Nguyen DM, Lipscomb GL, Benn RA, Prybol CJ, et al. Single gene insertion drives bioalcohol production by a thermophilic archaeon. Proc Natl Acad Sci U S A. 2014;111(49):17618-23.

[254] Bielen AAM, Willquist K, Engman J, Van Der Oost J, Van Niel EWJ, Kengen SWM. Pyrophosphate as a central energy carrier in the hydrogen-producing extremely thermophilic *Caldicellulosiruptor saccharolyticus*. FEMS Microbiol Lett. 2010;307:48-54.

[255] Kengen S, De Bok F, Van Loo N, Dijkema C, Stams A, De Vos W. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. J Biol Chem. 1994;269(26):17537-41.

[256] Mukund S, Adams MW. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. J Biol Chem. 1995;270(15):8389-92.

[257] Sapra R, Bagramyan K, Adams MW. A simple energy-conserving system: proton reduction coupled to proton translocation. Proc Natl Acad Sci U S A. 2003;100(13):7545-50.

[258] Scott IM, Rubinstein GM, Lipscomb GL, Basen M, Schut GJ, Rhaesa AM, et al. A new class of tungsten-containing oxidoreductase in *Caldicellulosiruptor*, a genus of plant biomass-degrading thermophilic bacteria. Appl Environ Microbiol. 2015;81(20):7339-47.

[259] Scott IM, Rubinstein GM, Poole FL, Lipscomb GL, Schut GJ, Williams-Rhaesa AM, et al. The thermophilic biomass-degrading bacterium *Caldicellulosiruptor bescii* utilizes two enzymes to oxidize glyceraldehyde 3-phosphate during glycolysis. J Biol Chem. 2019;294(25):9995-10005.

[260] van de Werken HJ, Verhaart MR, VanFossen AL, Willquist K, Lewis DL, Nichols JD, et al. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Appl Environ Microbiol. 2008;74(21):6720-9.

[261] Eberly JO, Ely RL. Thermotolerant hydrogenases: biological diversity, properties, and biotechnological applications. Crit Rev Microbiol. 2008;34(3-4):117-30.

[262] Straub CT, Bing RG, Otten JK, Keller LM, Zeldes BM, Adams MW, et al. Metabolically engineered *Caldicellulosiruptor bescii* as a platform for producing acetone and hydrogen from lignocellulose. Biotechnol Bioeng. 2020.

[263] Basen M, Rhaesa AM, Kataeva I, Prybol CJ, Scott IM, Poole FL, et al. Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the

thermophilic bacterium *Caldicellulosiruptor bescii*. Bioresour Technol. 2014;152:384-92.

[264] Straub CT, Khatibi PA, Otten JK, Adams MWW, Kelly RM. Lignocellulose solubilization and conversion by extremely thermophilic *Caldicellulosiruptor bescii* improves by maintaining metabolic activity. Biotechnol Bioeng. 2019;116:1901-8.

[265] Straub CT, Khatibi PA, Wang JP, Conway JM, Williams-Rhaesa AM, Peszlen IM, et al. Quantitative fermentation of unpretreated transgenic poplar by *Caldicellulosiruptor bescii*. Nat Commun. 2019;10:1-6.

[266] Zurawski JV, Khatibi PA, Akinosho HO, Straub CT, Compton SH, Conway JM, et al. Bioavailability of carbohydrate content in natural and transgenic switchgrasses for the extreme thermophile *Caldicellulosiruptor bescii*. Appl Environ Microbiol. 2017;83.

[267] Chung D, Cha M, Guss AM, Westpheling J. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. Proc Natl Acad Sci U S A. 2014;111(24):8931-6.

[268] Williams-Rhaesa AM, Rubinstein GM, Scott IM, Lipscomb GL, Poole FL, Kelly RM, et al. Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*. Metab Eng Commun. 2018;7:e00073.

[269] Straub CT, Schut G, Otten JK, Keller LM, Adams MW, Kelly RM. Modification of the glycolytic pathway in *Pyrococcus furiosus* and the implications for metabolic engineering. Extremophiles. 2020.

[270] Keller MW, Lipscomb GL, Nguyen DM, Crowley AT, Schut GJ, Scott I, et al. Ethanol production by the hyperthermophilic archaeon *Pyrococcus furiosus* by expression of bacterial bifunctional alcohol dehydrogenases. Microb Biotechnol. 2017;10(6):1535-45.

[271] Basen M, Sun J, Adams MW. Engineering a hyperthermophilic archaeon for temperature-dependent product formation. mBio. 2012;3(2).

[272] Zeldes BM, Straub CT, Otten JK, Adams MWW, Kelly RM. A synthetic enzymatic pathway for extremely thermophilic acetone production based on the unexpectedly thermostable acetoacetate decarboxylase from *Clostridium acetobutylicum*. Biotechnol Bioeng. 2018;115:2951-61.

[273] Loder AJ, Zeldes BM, Garrison GD, Lipscomb GL, Adams MW, Kelly RM. Alcohol selectivity in a synthetic thermophilic n-butanol pathway is driven by biocatalytic and thermostability characteristics of constituent enzymes. Appl Environ Microbiol. 2015;81(20):7187-200.

[274] Geng X-M, Liu X, Ji M, Hoffmann WA, Grunden A, Xiang Q-YJ. Enhancing heat tolerance of the little dogwood *Cornus canadensis* L. f. with introduction of a superoxide reductase gene from the hyperthermophilic archaeon *Pyrococcus furiosus*. Front Plant Sci. 2016;7:26.

[275] Im YJ, Ji M, Lee AM, Boss WF, Grunden AM. Production of a thermostable archaeal superoxide reductase in plant cells. FEBS Lett. 2005;579(25):5521-6.

[276] Jiang L, Huang C, Wang B, Guo H, Sun Q, Xia F, et al. Enhanced heat tolerance in transgenic silkworm via overexpression of *Pyrococcus furiosus* superoxide reductase. Insect Biochem Mol Biol. 2018;92:40-4.

[277] Im YJ, Ji M, Lee A, Killens R, Grunden AM, Boss WF. Expression of *Pyrococcus furiosus* superoxide reductase in *Arabidopsis* enhances heat tolerance. Plant Physiol. 2009;151(2):893-904.

[278] Jenney FE, Verhagen MF, Cui X, Adams MW. Anaerobic microbes: oxygen detoxification without superoxide dismutase. Science. 1999;286(5438):306-9.

[279] Smith-Moore CM, Grunden AM. Bacteria and archaea as the sources of traits for enhanced plant phenotypes. Biotechnol Adv. 2018;36(7):1900-16.

[280] Koch R, Spreinat A, Lemke K, Antranikian G. Purification and properties of a hyperthermoactive α-amylase from the archaeobacterium *Pyrococcus woesei*. Arch Microbiol. 1991;155(6):572-8.

[281] Ahmad N, Rashid N, Haider MS, Akram M, Akhtar M. Novel maltotriose-hydrolyzing thermoacidophilic type III pullulan hydrolase from *Thermococcus kodakarensis*. Appl Environ Microbiol. 2014;80(3):1108-15.

[282] Niehaus F, Peters A, Groudieva T, Antranikian G. Cloning, expression and biochemical characterisation of a unique thermostable pullulan-hydrolysing enzyme from the hyperthermophilic archaeon *Thermococcus aggregans*. FEMS Microbiol Lett. 2000;190(2):223-9.

[283] Kriegshäuser G, Liebl W. Pullulanase from the hyperthermophilic bacterium *Thermotoga maritima*: purification by β-cyclodextrin affinity chromatography. J Chromatogr B Biomed Sci Appl. 2000;737(1-2):245-51.

[284] Brown SH, Kelly RM. Characterization of amylolytic enzymes, having both alpha-1,4 and alpha-1,6 hydrolytic activity, from the thermophilic archaea *Pyrococcus*

furiosus and *Thermococcus litoralis*. Appl Environ Microbiol. 1993;59(8):2614-21.

[285] Badr H, Sims, KA, Adams, MWW. Purification and characterization of sucrose α-glucohydrolase (invertase) from the hyperthermophilic archaeon *Pyrococcus furiosus*. Syst Appl Microbiol. 1994;17:1-6

[286] Tanaka T, Fukui T, Fujiwara S, Atomi H, Imanaka T. Concerted action of diacetylchitobiose deacetylase and exo-β-d-glucosaminidase in a novel chitinolytic pathway in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J Biol Chem. 2004;279(29):30021-7.

[287] Gueguen Y, Voorhorst WG, van der Oost J, de Vos WM. Molecular and Biochemical Characterization of an Endo-β-1, 3-glucanase of the Hyperthermophilic Archaeon *Pyrococcus furiosus*. J Biol Chem. 1997;272(50):31258-64.

[288] Gao J, Bauer MW, Shockley KR, Pysz MA, Kelly RM. Growth of hyperthermophilic archaeon *Pyrococcus furiosus* on chitin involves two family 18 chitinases. Appl Environ Microbiol. 2003;69(6):3119-28.

[289] Tuohy MG, Walsh DJ, Murray PG, Claeyssens M, Cuffe MM, Savage AV, et al. Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. Biochim Biophys Acta Protein Struct Mol Enzym. 2002;1596(2):366-80.

[290] Freier D, Mothershed CP, Wiegel J. Characterization of *Clostridium thermocellum* JW20. Appl Environ Microbiol. 1988;54(1):204-11.

[291] Zhang C, Cooper TE, Krause DJ, Whitaker RJ. Augmenting the genetic toolbox for *Sulfolobus islandicus* with the stringent positive selectable marker for agmatine prototrophy. Appl Environ Microbiol. 2013.

CHAPTER 3

Engineering *Caldicellulosiruptor bescii* with surface layer homology domain-linked glycoside hydrolases improves plant biomass solubilization

Tunyaboon Laemthong[1], Ryan G. Bing[1], James R. Crosby[1],

Michael W.W. Adams[2], and Robert M. Kelly[1,*]

[1]*Department of Chemical and Biomolecular Engineering,*

North Carolina State University, Raleigh, NC 27695

[2]*Department of Biochemistry and Molecular Biology,*

University of Georgia, Athens, GA 30602

Published in: Appl. Environ. Microbiol. Sep 2022 e01274-22

Abstract

Extremely thermophilic *Caldicellulosiruptor* species solubilize carbohydrates from lignocellulose through glycoside hydrolases (GH) that can be extracellular, intracellular or cell surface layer associated. *Caldicellulosiruptor* genomes sequenced so far encode at least one Surface Layer Homology domain glycoside hydrolase (SLH-GH), representing six different classes of these enzymes; these can have multiple binding and catalytic domains. Biochemical characterization of a representative from each class was done to determine their biocatalytic features: four SLH-GHs from C*aldicellulosiruptor kronotskyensis* (Calkro_0111, Calkro_0402, Calkro_0072, Calkro_2036) and two from *Caldicellulosiruptor hydrothermalis* (Calhy_1629 and Calhy_2383). Calkro_0111, Calkro_0072, and Calhy_2383 exhibited β-1,3-glucanase activity, Calkro_0402 was active on both β-1,3/1,4-glucan and β-1,4-xylan, Calkro_2036 exhibited activity on both β-1,3/1,4-glucan and β-1,4-glucan, and Calhy_1629 was active only on arabinan. *C. bescii*, the only species with molecular genetic tools as well as already a strong cellulose degrader, contains only one SLH-GH, Athe_0594, a glucanase that is a homolog of Calkro_2036; the other 5 classes of SLH-GHs are absent in *C. bescii*. The *C. bescii* secretome, supplemented with individual enzymes or cocktails of SLH-GHs, increased *in vitro* sugar release from sugar cane bagasse and poplar. Expression of non-native SLH-GHs *in vivo*, either associated with the S-layer or as freely secreted enzymes, improved total carbohydrate solubilization of sugar cane bagasse and poplar by up to 45% and 23%, respectively. Most notably, expression of Calkro_0402, a xylanase/glucanase, improved xylose solubilization from poplar and bagasse by over 70% by *C. bescii*. While *Caldicellulosiruptor* species are already prolific lignocellulose degraders, they can be further improved by the strategy described here.

Importance

Caldicellulosiruptor species hold promise as microorganisms that can solubilize the carbohydrate portion of lignocellulose and subsequently convert fermentable sugars into bio-based chemicals and fuels. Members of the genus have Surface (S)-layer homology domain associated glycoside hydrolases (SLH-GHs) that mediate attachment to biomass as well as hydrolysis of carbohydrates. *Caldicellulosiruptor bescii*, the most studied member of the genus, has only one SLH-GH. Expression of SLH-GHs from other *Caldicellulosiruptor* species in *C. bescii* significantly improved degradation of sugar cane bagasse and poplar. This suggests that this extremely thermophilic bacterium can be engineered to further improve its ability to degrade specific plant biomasses by inserting genes encoding SLH-GHs recruited from other *Caldicellulosiruptor* species.

Introduction

Lignocellulose has significant potential as a renewable resource for biofuel and bio-based chemical production due to its abundance and cost (1). The biological conversion of lignocellulose into industrial chemicals requires synergistic action of cellulases, hemicellulases, and other plant biomass degrading enzymes to overcome biomass recalcitrance (2, 3). The simultaneous presence of these enzymes, and the subsequent ability to ferment both C5 and C6 sugars without significant catabolite repression, is rare among microorganisms, but does occur in several extremely thermophilic *Caldicellulosiruptor* species (4). These bacteria (T_{opt} > 70^{o}C) have drawn considerable attention for use in Consolidated BioProcessing (CBP), where lignocellulose solubilization and conversion are achieved with a single microorganism (5). Members of this genus are extremely thermophilic, Gram-positive, fermentative anaerobic bacteria that can grow on glucans and hemicelluloses; some species can also grow on microcrystalline cellulose (6). A genetic system developed for *Caldicellulosiruptor bescii* (7) has been improved significantly (8-11), thereby creating a metabolic engineering platform for bio-based chemical production (12). Recent developments, both experimentally and computationally, have further improved prospects for *C. bescii* (13-16).

Six large multi-domain glycoside hydrolases (GHs), CelA (Athe_1867), CelB (Athe_1859), CelC (Athe_1857), CelD (Athe_1866), CelE (Athe_1865), and CelF (Athe_1860), present in the Glucan Degradation Locus (GDL) of the *C. bescii* genome, play a key role in microcrystalline cellulose hydrolysis (17, 18). These GDL enzymes are naturally optimized for synergistic cellulose degradation (19). *Caldicellulosiruptor* species also use unique carbohydrate binding proteins, called “tāpirins”, and surface-layer homology (SLH) associated GHs (SLH-GHs) to attach specifically to lignocellulosic plant biomass (20, 21). Intracellular and extracellular GHs

across the genus *Caldicellulosiruptor* have been examined in detail, such as the cellulase, CelA (22-24), and the *C. kronotskyensis* xylanase (Xyn10A) (25), but a comprehensive study of the SLH-GHs has not been done.

The Surface layer (S-layer) in microorganisms lies on the outside of the cell envelope and it is non-covalently associated at either the N- or C-terminus of the S-layer proteins (26-28). Bacterial SLH domains in S-layer associated proteins are found in several microorganisms and mainly function as cell wall anchors (29-31). In the genus *Caldicellulosiruptor*, SLH-GHs are displayed on the cell surface and help deconstruct certain types of lignocellulosic plant biomass. For example, a *Caldicellulosiruptor saccharolyticus* SLH endoglucanase (Csac_0678) attaches to and hydrolyzes glucans in switchgrass (32). An SLH-GH in *Caldicellulosiruptor kronotskyensis* (Calkro_0402) was characterized as a xylanase and a knock-in of Calkro_0402 into a *C. bescii* strain improved substrate attachment and degradation of certain xylans compared to wild-type *C. bescii*. Calkro_0402 likely plays a role in helping the cell attach to xylan substrates (20). Furthermore, *Caldicellulosiruptor kristjanssonii* Xyn10C-GE15A, a homolog of Calkro_0402, exhibited thermostable xylanase-glucuronoyl esterase activity; it specifically hydrolyzed glucoranoarabinoxylan, a type of xylan found in corn biomass, made possible by its two different catalytic domains (33). While two SLH-GHs (Calkro_0402, Calkro_2036) and their homologs have been studied, representatives of the other four SLH-GHs have not (20, 32, 34, 35). Here, we consider representatives of all six classes of SLH-GHs from *Caldicellulosiruptor* species: Calkro_0111, Calkro_0402, Calkro_0072, and Calkro_2036 from *C. kronotskyensis*, Calhy_1629 and Calhy_2383 from *C. hydrothermalis* (see **Table 3 . 1)**. The SLH-GHs were produced recombinantly in either *C. bescii* or *E. coli*, and characterized biochemically. Then, *in vitro* lignocellulosic plant biomass hydrolysis was done with cocktails of SLH-GHs. Finally, six strains

of *C. bescii* were created to express each of the non-native SLH-GHs and then evaluated for potential contributions to *C. bescii* biomass solubilization.

Results and Discusssion

Recombinant production of *Caldicellulosiruptor* SLH-GHs. Six different classes of SLH-GHs are represented in *Caldicellulosiruptor* sp., as shown in **Table 3.1**. These SLH-GHs are found across the genus and have different domain architecture and molecular masses. Each SLH-GH consists of at least one glycoside hydrolase (GH) domain, one or more carbohydrate binding modules (CBMs), domains of unknown function, and have repetitive SLH domains at either the N- or C- terminus. Because of the size and complexity of the SLH-GHs, they have been difficult to produce recombinantly (20). However, successful expression of representative SLH-GHs from the six classes was accomplished using either *Escherichia coli* or *C. bescii* as the host (see **Figure 3.1**); SLH domains were removed to facilitate protein expression. In *C. bescii*, this resulted in secreted protein expression, thereby aiding the purification process. The SLH-GHs without their SLH domain were expressed with a C-terminal histidine tag. To determine whether anchoring to the S-layer was important for biomass degradation, SLH-GHs were also expressed in *C. bescii* with their SLH domain intact; in these cases, the expression strain and supernatant were probed with a His-tagged antibody to confirm the expression on the S-layer and in the culture supernatant (**Figure 3.6**). The predominant band (more than 75% of total protein in the lane on SDS PAGE in **Figure 3.1**) indicated that the M_r of the purified, truncated proteins, Calkro_0111, Calkro_0402, Calkro_0072, Calkro_2036, Calhy_1629, and Calhy_2383, were approximately 240 kDa, 160 kDa, 170 kDa, 60 kDa, 110 kDa, and 200 kDa, respectively, consistent with the M_r without the SLH domain determined from amino acid sequence.

Biochemical characterization of SLH-GH enzymes. The optimal pH and temperature of each SLH-GH were determined using the substrates on which the enzyme was most highly active (see **Tables 3.1** and **3.2**). The optimal pH of these SLH-GHs was in the range of 5-7 (see Figure **3.7**), while their optimal temperatures were between 70-80^{o}C (see **Figure 3.8**); these optima correspond to the optimal growth temperatures of *Caldicellulosiruptor* species from which the encoding genes were recruited. **Figure 3.3** provides information on thermostability of these enzymes. All SLH-GHs exhibited high stability at 70^{o}C and, as expected, stability decreased with increasing temperature. At 75^{o}C, all SLH-GHs, with the exception of Calkro_0072 and Calhy_2383, retained more than 50% of their original activity after 5 h of incubation, with half-lives of at least 5 h. The half-lives of Calkro_0072 and Calhy_2383 at 75°C were approximately 2-3 h. At 80^{o}C, Calkro_0111, Calhy_1629 and Calhy_2383 exhibited high thermostability with half-lives ~4 h. However, the other SLH-GHs lost 50% of their original activity within 30 min of incubation at 80°C. At 85^{o}C, most SLH-GHs had ~30 min half-lives, except for Calhy_1629 and Calhy_2383 which had $t_{1/2}$ of ~1.5 h and ~2.5 h, respectively.

Substrate specificity of *Caldicellulosiruptor* SLH-GHs. A variety of polysaccharides were incubated with each SLH-GH to investigate substrate range. As seen in **Figure 3.2**, *Caldicellulosiruptor* SLH-GHs consist of: catalytic domains (Glycoside Hydrolase (GH)), carbohydrate binding domains (Carbohydrate Binding Module (CBM)), and domains of unknown function (e.g., Actin Cross-linking-Like (ACL), Cadherin-like Domain (Cad), Fibronectin type-III (FN3)). Each SLH-GH was active on multiple substrates, reflecting the differences in sugar backbones and side chains.

Calkro_0111, consisting of GH16 and GH55 catalytic domains, was most active on substrates containing β-1,3-linkages in their backbones, with either β-1,4 or β-1,6 side chains (laminarin, barley-β-glucan**, Figure 3.2**). According to the Carbohydrate Active enZYmes (CAZY) database (36), the GH16 family consists of several subgroups, including exo- and endo-β-1,3-glucanases. Moreover, a truncated Calkro_0111, to separate the GH domains, displayed endo- or exo-glucanase activity on laminarin, reflecting the GH16 and GH55 domains, respectively (20). GH16 domains can release oligosaccharides from laminarin (20, 37) and GH55 domains can release glucose and laminaribiose from laminarin (38, 39). Therefore, Calkro_0111 can be classified as a laminarinase.

Calkro_0402, which has a GH10 domain, exhibited high activity on substrates containing β-1,4-xylose linkages, including birchwood xylan, oat spelt xylan, lichenan, wheat arabinoxylan, and barley-β-glucan (**Figure 3.2**). Calkro_0402 and its homologs have been previously described as endo-β-1,4-xylanases (20, 33), representative of the high activity on xylan and wheat arabinoxylan. The results here also show hydrolysis of β-1,3/1,4-glucans, as evidenced by the activity on lichenan and barley-β-glucan, most likely indicating endo-β-1,3/1,4-glucanase (lichenase) activity. Calkro_0402 was not active on substrates with only β-1,3-glucans or only β-1,4-glucans.

Calkro_0072, which contains a GH16 domain, was highly active on curdlan, laminarin, lichenan, and barley-β-glucan, all of which contain β-1,3-glucans (**Figure 3.2**). Calkro_0072 can be described as a β-1,3-glucanase, as observed for similar enzymes from other microorganisms (40, 41).

Calkro_2036 is the smallest enzyme among *Caldicellulosiruptor* SLH-GHs. Its catalytic domain belongs to family GH5 and its homolog in *C. bescii*, Athe_0594, has been previously

described as a cellulase and a lichenase (34). Here, Calkro_2036 was highly active on substrates lichenan, carboxymethyl cellulose (CMC), and barley-β-glucan (**Figure 3.2**). Trace activity on oat spelt xylan was detected. This indicated both cellulase and lichenase activity, with additional activity on oat spelt xylan, likely due to glucose sidechain hydrolysis.

Calhy_1629 belongs to family GH43. It was mostly active on arabinan substrates that have α-1,5-arabinan as their backbone (arabinan and debranched arabinan). GH43 domains are typically found in arabinanases, with homologs in other microorganisms, including *Thermotoga thermarum* (42) and *Bacillus subtilis* (43). Thus, the primary activity of Calhy_1629 is as an arabinanase.

Lastly, Calhy_2383 contains a GH87 catalytic domain, commonly found in fungi and is usually part of a multi-domain enzyme (44, 45). Enzymes containing GH87 domains are mainly active on α-1,3-glucan substrates (46). Here, Calhy_2383 was active on lichenan, laminarin, and barley-β-glucan, which contain β-1,3-glucan mixed with β-1,4 or β-1,6 glucan linkages. Lower activity was detected on curdlan (β-1,3-glucose), and CMC (β-1,4-glucose. No activity was seen on any of the xylose containing substrates (**Figure 3.2**). This suggests a primary activity as a β-1,3-glucanase, with additional secondary activities β-1,4-glucanase. Calhy_2383 activity on CMC is particularly interesting as *C. hydrothermalis* is unable to effectively degrade cellulose, indicating that the β-1,4-glucanase activity seen here is not sufficient enough for *C. hydrothermalis* to degrade crystalline cellulose.

Engineering *C. bescii*'s S-layer to improve solubilization of plant biomass. *C. bescii*, the only *Caldicellulosiruptor* species for which molecular genetic tools are available, has only one SLH-GH (Athe_0594). Given its lack of five of the six SLH-GH types in the genus *Caldicellulosiruptor*, the objective was to determine if *C. bescii* could be a better biomass degrader

with additional SLH-GHs recruited from other species. Previously, the function of SLH domains has been examined in other microorganisms (30, 47, 48). The SLH domains in Gram-positive bacteria are involved in surface attachment using the interaction between peptidoglycan-linked secondary cell wall polymers and SLH domains (49). This is likely the case for the *Caldicellulosiruptor* SLH-GHs studied here. The SLH-GHs were heterologously expressed in *C. bescii* with and without their S-layer domain intact and in all cases with C-terminal 6x His tags. To confirm that the SLH-domain containing enzymes were localized to the cell surface, cells were incubated with Alexa Fluor 488 conjugated Ni-NTA beads and imaged with confocal microscopy. As seen in **Figure 3.4**, these expression strains displayed the SLH-GHs on their cell surface, as indicated by the green fluorescence observed. Meanwhile, strains with deleted SLH domains did not exhibit green fluorescence on the cell surface.

Lignocellulosic plant biomass solubilization by *C. bescii* secretome supplemented with SLH-GH cocktails. The addition of SLH-GH cocktails to the parent strain's secretome, *in vitro*, was done to see if this enhanced reducing sugar release from lignocellulosic plant biomasses (see **Table 3.3**). Sugar cane bagasse and poplar, which have different carbohydrate profiles, were used as growth substrates to generate the secretome. These biomasses each contain more cellulose than hemicellulose, but they differ in sugar composition (glucose, xylose, galactose, arabinose and mannose) (see **Figure 3.5**). Some degree of synergism occurred in all SLH combinations with the parent strain's secretome in vitro. The best case was a mixture of 75% secretome and 25% SLH-GHs for both biomasses. For sugar cane bagasse hydrolysis, the mixture of 75% secretome and 25% SLH-GHs released 79% more reducing sugar compared to secretome only. Other

combinations, 50% secretome or 25% secretome with balance SLH-GHs, were able to release 12 and 32%, respectively, more reducing sugars compared with using secretome only (**Table 3.3**).

Similar trends were observed with poplar biomass hydrolysis, but the SLH-GH addition was more impactful. Incubation with 75% secretome and 25% SLH-GHs improved reducing sugar release from poplar more than 321% compared to secretome only (**Table 3.3**). The differences in sugar cane bagasse and poplar biomass hydrolysis between the combination of secretome and SLH-GHs were likely due to differences in plant biomass structure and polysaccharide components. The cross-linking of ferulate-complex and lignin-polysaccharide is normally formed in sugar cane bagasse making it more recalcitrant (50). Moreover, sugar components in both are varied. Other combinations of secretome and SLH-GHs also showed significant improvements for poplar. Synergistic enzymatic activity between two enzymes has been shown to exhibit cooperativity between the different domains. Adding glucoronoyl esterase (*Ck*GE15) to a commercial xylanase, at its optimal temperature, resulted in increased xylan liberation from lignin in wheat straw (33). Enzymatic synergism between cellulases and hemicellulases has also been reported to improve lignocellulosic plant biomass deconstruction (51-53). Natively, *C. bescii* has only a single SLH-GH, but has a plethora of other secreted cellulases and hemicellulases that are capable of deconstructing significant portions of plant biomasses (35). However, *C. bescii* is unable to solubilize all polysaccharides contained in certain lignocelluloses, especially xylans and arabinans (14, 16). Supplementing with SLH-GH cocktails could overcome this limitation.

Lignocellulosic plant biomass solubilization by *C. bescii* SLH-GH expression strains. To further evaluate the role of each SLH-GH on lignocellulosic plant biomass deconstruction, *C. bescii* expression strains for each SLH-GH, with and without the SLH domain intact, were

compared against the parent strain on sugar cane bagasse and poplar. Recall that Calkro_0111, Calkro_0072, and Calhy_2383 were most active on β-1,3-glucan containing substrates; Calkro_2036, Calkro_0402 and Calhy_1629 were most active on lichenan, β-1,4-xylan and α-1,5-arabinan, respectively. There was no significant difference in lignocellulosic plant biomass solubilization between strains with and without SLH domains. Therefore, the strains with SLH domains were used for biomass solubilization. In natural environments where carbohydrate substrates are scarce, the intact SLH domain is likely important for scavenging nutrients.

As seen in **Table 3.4, Table 3.5** and **Figure 3.5**, the addition of Calkro_0402 (β-1,4-xylanase/lichenase) to the parent strain solubilized 44.5% and 23.0% more carbohydrate than the parent strain for bagasse and poplar, respectively. This strain had significantly increased xylose solubilization compared to all other strains (78.1% and 71.9%, for bagasse and poplar, respectively), representative of the enzyme's xylanase activity. All strains were capable of improving glucose solubilization compared to the parent strain (see **Table 3.4, Table 3.5** and **Figure 3.5**); the most significant solubilization was 48.5% and 35.2% for Calkro_2036 for bagasse and poplar, respectively. Calkro_2036 is homologous to Athe_0594 in *C. bescii*, but the higher expression level in the expression strain indicates that Athe_0594 needs to be at higher levels to be effective against specific biomasses. All of these enzymes have glucanase activity of various types, including β-1,3-glucanase (Calkro_0072, Calhy_2383), β-1,4-glucanase (Calkro_2036, Calhy_2383), lichenase (Calkro_0402, Calkro_2036, Calhy_1629 [trace], Calhy_2383), and laminarinase (Calkro_0111). These glucanase activities support the increases in vivo glucose solubilization from plant biomasses (**Figure 3.5**). Enzymes with the highest activities on lichenan and barley-β-glucan also had the highest glucan solubilization (**Figure 3.2**, **Figure 3.5**, **Table 3.4, Table 3.5**). Alternatively, several SLH-GHs (Calkro_0111, Calkro_0072, Calkro_2036, and

Calhy_2383), which contributed significantly to glucose release, contain CBM domains unique from *C. bescii* glucanases and may be responsible for enhancing glucan solubilization. Additionally, these SLH-GHs also increased mannose solubilization in poplar, indicating that these glucanases may also have activity on mannan; sugar cane bagasse has no significant mannose content (**Figure 3.5**).

The Calhy_1629 strain increased arabinose solubilization by 71.3% from sugar cane bagasse, the highest among strains; however, arabinose is not a major component in sugar cane bagasse, thus not contributing much to overall carbohydrate solubilization. It is noted that there was no significant difference in galactose solubilization between strains (all 64-66% solubilization), although galactose is a minor component in these biomasses (**Figure 3.5**). The results shown here indicate that the addition of surface-associated enzymes to the parent strain of *C. bescii* cooperate with secreted carbohydrate active enzymes in the secretome to better deconstruct lignocellulosic plant biomasses, resulting in more fermentable sugar release for subsequent use in metabolic pathways.

The importance of surface layer association of the enzymes is likely not the primary cause of increased solubilization, as no difference was seen for Calkro_0402 and Calkro_2036, with or without the SLH domain, expressed in *C. bescii* parent strain on sugar cane bagasse (data not shown). It is more likely that the unique CBM domains, catalytic activity, and/or increased expression level (in the case of Calkro_2036, Athe_0594 homolog) that are responsible for improved biomass degradation. The shared glucanase activity of these enzymes suggests that *C. bescii* might benefit from additional β-1,3/1,4-glucanase activity, at least for poplar and sugar cane bagasse. In other studies of *C. bescii*, coexpression of *Themotoga maritima* β-D-xylosidase and *Acidothermus cellulolyticus* xylanase in *C. bescii* substantially improves exproteome activity

towards xylan and cell growth on cellulose (54). Additionally, co-expression of β-glucanases and a cellobiose phosphorylase in *C. bescii* increased cellulolytic activity of *C. bescii* exproteome (55). These are evidence that synergism between the *C. bescii* secretome and non-native enzymes might be key to improving its ability to hydrolyze plant biomass. Future work could further explore the importance of native and non-native β-1,3/1,4-glucanases in *C. bescii*. In any case, these results show that there is room to improve *C. bescii*'s native ability to deconstruct plant biomass by adding or enhancing the level of carbohydrate active enzymes. While previous work has shown that *C. bescii* has fully optimized its native enzymes for crystalline cellulose degradation (19), non-native enzymes can still make significant improvements in solubilization of certain polysaccharides. Addition of these enzymes into *C. bescii* can help make it a better candidate in CBP where plant biomass deconstruction still remains one of the major bottlenecks.

Materials and Methods

Bacterial strains, plasmids, and reagents. Plasmid construction and expression of recombinant proteins was performed using the following *Escherichia coli* strains: NEB 5-alpha and NEB 10-beta (New England Biolabs, Ipswich, MA), Rosetta 2(DE3) pLysS (Fisher Scientific), and Arctic Express (DE3) RIL *E. coli* (Agilent Technologies). Overexpression of full length and truncated SLH proteins in *C. bescii* was performed using strain *C. bescii* MACB1018 (8, 10). Genes of interest were amplified from *Caldicellulosiruptor* species genomic DNA by using NEB Monarch kit) and polymerase chain reaction (PCR) (NEB Phusion High Fidelity DNA polymerase). Plasmid construction was done using the pET46 Ek/LIC vector kit (EMD Millipore) for *E. coli* protein expression and pJMC046 replicating plasmids containing 6 x His tag on the C-terminus were used for *C. bescii* protein expression (19) using NEB HiFi Builder (New England Biolabs, Ipswich, MA). Plasmid sequences were confirmed by Sanger sequencing (Azenta Life

Sciences, Morrisville, NC). Plasmids were maintained at -80°C in LB medium + 15% glycerol and 50 μg/mL kanamycin, 50 μg/mL ampicillin, and/or 34 μg/mL chloramphenicol, as appropriate.

Genetic manipulation of *C. bescii.* Competent cells of MACB1018 strain and methylated plasmids were prepared using a protocol previously described in (8, 18, 21). For each electroporation, 1 μg of methylated plasmids was added directly to 50 μL of competent cells at room temperature. The mixture was then transferred to 1 mm gap glass cuvette (USA Scientific) and electroporated using Gene Pulser II system with a Pulse Controller PLUS module (Bio-Rad). The electroporation was initiated at 2.0 kV, 400 Ω, and 25 microfarads. The voltage and resistance were increased as the size of plasmids increased. After the electroporation, the cells were immediately transferred to low osmolality complex (LOC) medium (55) and incubated at 70^{o}C for 1 h for recovery. After recovery, the transformants were passaged into defined modified medium DSM 516 (18) with 50 μg/mL kanamycin selection and 24 mM NH_4Cl. After 3-4 days, the culture was plated and embedded in DSM516 medium with 1.5% agar and grown anaerobically at 70^{o}C. Single colonies were picked and grown in liquid medium. The transformed DNA was isolated from liquid culture and sequenced. Primers used are listed in **Table 3.6**.

Protein expression and purification. Protein expression from *E. coli* strains was performed in 2x YT medium (10 g/L yeast extract, 16 g/L tryptone, 5g/L sodium chloride) with the appropriate antibiotics (50 μg/mL kanamycin or 50 μg/mL ampicillin plus 34 μg/mL chloramphenicol for Rosetta; 50 μg/mL kanamycin and 50 μg/mL gentamicin for Arctic Express) at 37^{o}C with 250 rpm agitation. Once the OD_{600} reached 0.8, cultures were cooled and then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20^{o}C for 18 hours. Cells were

harvested at 6,000 x g for 10 min. Cell pellets were kept at -20°C prior to lysis and purification. Cell pellets were thawed and then resuspended in 20 mM sodium phosphate, 500 mM sodium chloride, pH 8.1. Cell suspension was lysed twice in a French press (Sim-Aminco) at ~15,000 psi. Because the target proteins are expected to be thermostable, cell lysate was heat-treated at 65°C for 20 min to remove heat labile proteins from *E. coli*. The cell lysate was centrifuged at 25,000 x g for 30 mins at 4°C, and the supernatant was passed through a 0.22 μm filter to obtain cell-free extract.

For Calkro_0111, the *C. bescii* overexpression strain was grown in complex modified DSM 516 medium (18) plus 50 μg/mL kanamycin for 24-36 hours at 75°C and the supernatant was harvested by centrifuging the culture at 6,000 x g for 10 min. The supernatant was filter-sterilized and was kept at 4°C before purification.

Proteins were purified using 5 mL HisTrap HP Ni-Sepharose (Cytiva) immobilized metal affinity chromatography (IMAC) columns, operated according to the manufacturer's instructions. Size exclusion chromatography was used after the IMAC column on a HiLoad 26/600 Superdex 200 pg column (Cytiva) using buffer containing 20 mM sodium phosphate, 500 mM sodium chloride, pH 8.1. All chromatography steps were performed on a Biologic DuoFlow FPLC (Bio-Rad). The purity of purified proteins was assessed via SDS-PAGE using 4-15% Mini-PROTEAN TGX stain-free gels (Bio-Rad) using Laemilli's method with a Benchmark protein ladder (Life Technologies). The gel images were also analyzed using densitometry (ImageJ). The concentration of purified proteins was determined using Pierce BCA protein assay (Thermo Fisher) with Bovine Serum Albumin (BSA) as a protein standard.

Polysaccharide substrate specificity. Proteins (0.5 μM) were incubated with 1% w/v of the following substrates; birchwood xylan, oat spelt xylan (MilliporeSigma, St. Louis, MO), curdlan, laminarin, lichenan, carboxymethyl cellulose, pustulan (Carbosynth), tamarind xyloglucan, Wheat arabinoxylan, barley-β-glucan, arabinan, debranched arabinan (Megazyme, Wicklow, Ireland) in a final assay volume of 100 μL. Assays were performed at 70^{o}C for 20 min in a thermocycler. Released reducing sugars were measured using 3,5-dinitrosalicyclic acid (DNS) assay, as previously described by (19). The DNS reagent used contained 1.6% (w/v) sodium hydroxide, 30% (w/v) sodium potassium nitrate, and 0.1% (w/v) 3,5-dinitrosalicyclic acid. Briefly, 25 μL of sample was incubated with 50 μL of DNS reagent and incubated in PCR tubes using the following program: 95^{o}C for 5 min, 48^{o}C for 1 min, and hold at 20^{o}C. Then, 36 μL of DNS reaction was diluted with 160 μL of distilled water, and absorbance with measure at 540 nm. Xylose or glucose was used as a standard for released xylooligosaccharide and glucooligosaccharide.

pH and temperature optima. pH optima were determined at 70^{o}C for 20 min in buffer pHs ranging from 3-10 (50 mM citrate buffer for pH 3-6, 50 mM sodium phosphate buffer for pH 6.5-8, and sodium bicarbonate buffer for pH 9-10). The temperature optima were evaluated at the optimal pH of each enzyme at 30, 40, 50, 60, 65, 70, 75, 80, 90 and 95^{o}C. All reactions were done in triplicate with 0.5 μM of SLH-GH in 100 μL of 1% w/v substrate and no enzyme controls were used. The DNS assay was used to assess the amount of released sugars from each reaction, as described above.

Enzyme thermostability. Each SLH-GH was incubated at 75°C, 80°C, and 85^{o}C for 5 h in its storage buffer. At different time points (0, 0.5, 1, 2, 3, 4, and 5 hours), 0.5 μM of preheated

sample was taken and the residual activities were measured by incubating enzyme-substrate mixture at 70^{o}C for 20 min. Ten mg/mL of birchwood xylan, curdlan, CMC, arabinan, and lichenan were used as substrates for Calkro_0402, Calkro_0072, Calkro_2036, Calhy_1629, and Calhy_2383, respectively. DNS assay was used to determine the amount of released sugar from each reaction and compared with non-denatured enzyme.

SLH-GH specific activity. Enzyme activity was determined by incubating each enzyme with 1% polysaccharide substrate at the optimal temperature for 20 min in its optimal pH buffer. Substrates used in these assays were picked based on their highest activity found in the substrate specificity. DNS assay was used to measure released sugars. The enzyme activity was reported as μmol of reducing sugar equivalents per min.

Confocal microscopy. The procedure was adapted from (57). All centrifugation steps were performed at 6,000 x g for 10 min at room temperature. Twenty-five mL of cell culture was drawn and pelleted. Cell pellets were washed three times with 1x Phosphate-Buffered Saline (PBS) and were resuspended in 3.6 μM 2-[4-(Aminoiminomethyl)phenyl]-1*H*-Indole-6-carboximidamide hydrochloride (DAPI) (Thermo Fisher) in PBS and were left for 12 hours at 4^{o}C. After staining with DAPI, cells were pelleted and washed three times with 1 x PBS. The cells were incubated with 300 μg/mL of Ni-NTA fluorescent dye (MilliporeSigma, St. Louis, MO) for another 12 h at 4^{o}C. The cells were pelleted and washed three times with 1 x PBS and were resuspended in 1 x PBS prior to imaging. The images were taken on Zeiss LSM880 confocal microscope (Carl Zeiss, Germany) operating in AiryScan Fast Mode (North Carolina State University Cellular and

Molecular Imaging Facility (CMIF), Raleigh, NC). Images were processed by Zen Black 2.3 software.

Lignocellulosic plant biomass preparation. All biomass substrates were prepared by milling and sieving, as previously described in (16). The 40/80 sieve fraction was collected and was washed three times with 75^{o}C distilled water to remove any soluble sugars. The biomass was then dried at 65^{o}C overnight.

Lignocellulosic plant biomass hydrolysis with SLH-GH cocktails. Enzyme synergism was investigated on sugar cane bagasse and wild-type poplar. Briefly, the -80 sieve fractions, as described above, of each lignocellulosic plant biomass were collected and used in the experiment. Different combinations of SLH-GHs and secretome of MACB1018 cells grown on the respective lignocellulosic plant biomasses was incubated at 70^{o}C with 500 rpm agitation. Sugar release was monitored at different time points (0, 2, 4, 8, 12, and 24 hours) by DNS assay. All assays were performed in triplicate.

Biomass solubilization. To assess biomass solubilization, 5 g/L of prepared 40/80 sieve fraction biomass was used in each culture. *Caldicellulosiruptor* strains were passaged three times on modified 671d medium, as described previously (58), on each biomass substrate, supplemented 250 mM uracil and 50 μg/mL kanamycin. Solubilization cultures were prepared by inoculating 5x10^{8} cells/mL, and grown at 75^{o}C with 250 rpm for 7 days. Experiments were done in triplicate. After the fermentation, each culture was cooled to room temperature and harvested by centrifuging at 5,000 rpm for 10 min at room temperature. Soluble fermentation products were determined

using a Waters Arc3 HPLC, as described in (14). using a refractive index detector and Rezex-ROA column with a 5 mM sulfuric acid mobile phase flowing at 0.6 mL/min at 50 °C (Model 2414, (Phenomenex), Waters Corp.). Residual substrate biomass was washed three times and dried for 24 h at 75^{o}C. The final dried plant biomass weight was recorded to calculate biomass solubilization.

Quantitative saccharification. Composition of lignocellulosic plant biomasses was evaluated using Klason method, as described in (16). The amount of sugar was measured using a Waters Arc3 HPLC using a Shodex SP0810 sugar column for separation with a water mobile phase at 0.6 mL/min at 85^{o}C and a Waters Model 2414 refractive index detector for detection.

Acknowledgments

This work was supported in part by the US Department of Agriculture (2018-67021-27716), the US National Science Foundation (CBET-1802939) and the US Department of Energy BER Awards DE-SC0019391 and DE-SC0022192. T Laemthong acknowledges support from the Government of Thailand. RG Bing acknowledges support from an NIH Biotechnology Traineeship (NIH T32 GM008776-16). JR Crosby acknowledges support from a US DoEd GAANN Fellowship (P200A160061). We acknowledge the North Carolina State University Cellular and Molecular Imaging Facility supported by the state of North Carolina and the National Science Foundation and Dr. Mariusz Zareba for assistance with imaging analysis (Zeiss LSM880 confocal microscope).

References

1. Kumar R, Singh S, Singh OV. 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. J Ind Microbiol Biotechnol 35:377-391.
2. Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MW, Kelly RM. 2014. Thermophilic lignocellulose deconstruction. FEMS Microbiol Rev 38:393-448.
3. Girio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S, Bogel-Lukasik R. 2010. Hemicelluloses for fuel ethanol: A review. Bioresour Technol 101:4775-800.
4. Blumer-Schuette SE, Kataeva I, Westpheling J, Adams MW, Kelly RM. 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr Opin Biotechnol 19:210-7.
5. Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE. 2017. Cellulosic ethanol: status and innovation. Curr Opin Biotechnol 45:202-211.
6. Lee LL, Crosby JR, Rubinstein GM, Laemthong T, Bing RG, Straub CT, Adams MWW, Kelly RM. 2020. The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in 'Caldi World'. Extremophiles 24:1-15.
7. Chung D, Cha M, Guss AM, Westpheling J. 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. Proc Natl Acad Sci 111:8931-8936.
8. Lipscomb GL, Conway JM, Blumer-Schuette SE, Kelly RM, Adams MW. 2016. A highly thermostable kanamycin resistance marker expands the tool kit for genetic

manipulation of *Caldicellulosiruptor bescii*. Appl Environ Microbiol 82:4421-4428.

9. Williams-Rhaesa AM, Awuku NK, Lipscomb GL, Poole FL, Rubinstein GM, Conway JM, Kelly RM, Adams MW. 2018. Native xylose-inducible promoter expands the genetic tools for the biomass-degrading, extremely thermophilic bacterium *Caldicellulosiruptor bescii*. Extremophiles 22:629-638.
10. Williams-Rhaesa AM, Poole FL, 2nd, Dinsmore JT, Lipscomb GL, Rubinstein GM, Scott IM, Conway JM, Lee LL, Khatibi PA, Kelly RM, Adams MWW. 2017. Genome stability in engineered strains of the extremely thermophilic lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol 83.
11. Williams-Rhaesa AM, Rubinstein GM, Scott IM, Lipscomb GL, Poole FL, Kelly RM, Adams MW. 2018. Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*. Metab Eng 7:e00073.
12. Straub CT, Bing RG, Otten JK, Keller LM, Zeldes BM, Adams MW, Kelly RM. 2020. Metabolically engineered *Caldicellulosiruptor bescii* as a platform for producing acetone and hydrogen from lignocellulose. Biotechnol Bioeng doi:10.1002/bit.27529.
13. Rodionov DA, Rodionova IA, Rodionov VA, Arzamasov AA, Zhang K, Rubinstein GM, Tanwee TNN, Bing RG, Crosby JR, Nookaew I, Basen M, Brown SD, Wilson CM, Klingeman DM, Poole FL, 2nd, Zhang Y, Kelly RM, Adams MWW. 2021. Transcriptional regulation of plant biomass degradation and carbohydrate

utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. mSystems 6:e0134520.

14. Straub CT, Khatibi PA, Wang JP, Conway JM, Williams-Rhaesa AM, Peszlen IM, Chiang VL, Adams MWW, Kelly RM. 2019. Quantitative fermentation of unpretreated transgenic poplar by *Caldicellulosiruptor bescii*. Nat Commun 10:3548.

15. Zhang K, Zhao W, Rodionov DA, Rubinstein GM, Nguyen DN, Tanwee TNN, Crosby J, Bing RG, Kelly RM, Adams MWW, Zhang Y. 2021. Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for bio-based chemical production. mSystems 6:e0135120.

16. Bing RG, Straub CT, Sulis DB, Wang JP, Adams MW, Kelly RM. 2022. Plant biomass fermentation by the extreme thermophile *Caldicellulosiruptor bescii* for co-production of green hydrogen and acetone: Technoeconomic analysis. Bioresour Technol 348:126780.

17. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL, Adams MW. 2012. *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. J Bacteriol 194:4015-4028.

18. Conway JM, McKinley BS, Seals NL, Hernandez D, Khatibi PA, Poudel S, Giannone RJ, Hettich RL, Williams-Rhaesa AM, Lipscomb GL. 2017. Functional analysis of the glucan degradation locus in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. Appl Environ Microbiol 83:e01828-17.

19. Conway JM, Crosby JR, McKinley BS, Seals NL, Adams MW, Kelly RM. 2018. Parsing in vivo and in vitro contributions to microcrystalline cellulose hydrolysis by multidomain glycoside hydrolases in the *Caldicellulosiruptor bescii* secretome. Biotechnol Bioeng 115:2426-2440.

20. Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, Zurawski JV, Lee LL, Blumer-Schuette SE, Kelly RM. 2016. Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. J Biol Chem 291:6732-6747.

21. Lee LL, Hart WS, Lunin VV, Alahuhta M, Bomble YJ, Himmel ME, Blumer-Schuette SE, Adams MWW, Kelly RM. 2019. Comparative biochemical and structural analysis of novel cellulose binding proteins (tapirins) from extremely thermophilic *Caldicellulosiruptor* species. Appl Environ Microbiol 85.

22. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang S-J, Resch MG, Adams MW, Lunin VV. 2013. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science 342:1513-1516.

23. Chung D, Young J, Bomble YJ, Vander Wall TA, Groom J, Himmel ME, Westpheling J. 2015. Homologous expression of the *Caldicellulosiruptor bescii* CelA reveals that the extracellular protein is glycosylated. PLoS One 10:e0119508.

24. Young J, Chung D, Bomble YJ, Himmel ME, Westpheling J. 2014. Deletion of *Caldicellulosiruptor bescii* CelA reveals its crucial role in the deconstruction of lignocellulosic biomass. Biotechnol Biofuels 7:142.

25. Jia X, Qiao W, Tian W, Peng X, Mi S, Su H, Han Y. 2016. Biochemical characterization of extra- and intracellular endoxylanse from thermophilic bacterium *Caldicellulosiruptor kronotskyensis*. Sci Rep 6:21672.

26. Fagan RP, Janoir C, Collignon A, Mastrantonio P, Poxton IR, Fairweather NF. 2011. A proposed nomenclature for cell wall proteins of *Clostridium difficile*. J Med Microbiol 60:1225-1228.

27. Johnson BR, Hymes J, Sanozky-Dawes R, Henriksen ED, Barrangou R, Klaenhammer TR. 2016. Conserved S-layer-associated proteins revealed by exoproteomic survey of S-layer-forming Lactobacilli. Appl Environ Microbiol 82:134-45.

28. Mesnage S, Fontaine T, Mignot T, Delepierre M, Mock M, Fouet A. 2000. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J 19:4473-84.

29. Blackler RJ, Lopez-Guzman A, Hager FF, Janesch B, Martinz G, Gagnon SML, Haji-Ghassemi O, Kosma P, Messner P, Schaffer C, Evans SV. 2018. Structural basis of cell wall anchoring by SLH domains in *Paenibacillus alvei*. Nat Commun 9:3120.

30. Kern J, Wilton R, Zhang R, Binkowski TA, Joachimiak A, Schneewind O. 2011. Structure of surface layer homology (SLH) domains from *Bacillus anthracis* surface array protein. J Biol Chem 286:26042-9.

31. May A, Pusztahelyi T, Hoffmann N, Fischer RJ, Bahl H. 2006. Mutagenesis of conserved charged amino acids in SLH domains of *Thermoanaerobacterium*

thermosulfurigenes EM1 affects attachment to cell wall sacculi. Arch Microbiol 185:263-9.

32. Ozdemir I, Blumer-Schuette SE, Kelly RM. 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.

33. Krska D, Larsbrink J. 2020. Investigation of a thermostable multi-domain xylanase-glucuronoyl esterase enzyme from *Caldicellulosiruptor kristjanssonii* incorporating multiple carbohydrate-binding modules. Biotechnol Biofuels 13:68.

34. Dam P, Kataeva I, Yang SJ, Zhou F, Yin Y, Chou W, Poole FL, 2nd, Westpheling J, Hettich R, Giannone R, Lewis DL, Kelly R, Gilbert HJ, Henrissat B, Xu Y, Adams MW. 2011. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. Nucleic Acids Res 39:3240-54.

35. Poudel S, Giannone RJ, Basen M, Nookaew I, Poole FL, 2nd, Kelly RM, Adams MWW, Hettich RL. 2018. The diversity and specificity of the extracellular proteome in the cellulolytic bacterium *Caldicellulosiruptor bescii* is driven by the nature of the cellulosic growth substrate. Biotechnol Biofuels 11:80.

36. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 50:D571-D577.

37. Badur AH, Ammar EM, Yalamanchili G, Hehemann J-H, Rao CV. 2020. Characterization of the GH16 and GH17 laminarinases from *Vibrio breoganii* 1C10. Appl Microbiol Biotechnol 104:161-171.

38. Bianchetti CM, Takasuka TE, Deutsch S, Udell HS, Yik EJ, Bergeman LF, Fox BG. 2015. Active site and laminarin binding in glycoside hydrolase family 55. J Biol Chem 290:11819-11832.

39. Wang Y, Li D, Dong C, Zhao Y, Zhang L, Yang F, Ye X, Huang Y, Li Z, Cui Z. 2021. Heterologous expression and characterization of a novel glycoside hydrolase family 55 β-1, 3-glucanase, AcGluA, from *Archangium sp.* strain AC19. Appl Microbiol Biotechnol 105:6793-6803.

40. Mouyna I, Aimanianda V, Hartl L, Prevost Mc, Sismeiro O, Dillies MA, Jagla B, Legendre R, Coppee JY, Latgé JP. 2016. GH16 and GH81 family β-(1, 3)-glucanases in *Aspergillus fumigatus* are essential for conidial cell wall morphogenesis. Cell Microbiol 18:1285-1293.

41. Strohmeier M, Hrmova M, Fischer M, Harvey AJ, Fincher GB, Pleiss J. 2004. Molecular modeling of family GH16 glycoside hydrolases: potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the poaceae. Protein Sci 13:3200-3213.

42. Shi H, Ding H, Huang Y, Wang L, Zhang Y, Li X, Wang F. 2014. Expression and characterization of a GH43 endo-arabinanase from *Thermotoga thermarum*. BMC Biotechnol 14:1-9.

43. Inácio JM, de Sá-Nogueira I. 2008. Characterization of abn2 (yxiA), encoding a *Bacillus subtilis* GH43 arabinanase, Abn2, and its role in arabino-polysaccharide degradation. J Bacteriol 190:4272-4280.

44. Itoh T, Intuy R, Suyotha W, Hayashi J, Yano S, Makabe K, Wakayama M, Hibi T. 2020. Structural insights into substrate recognition and catalysis by glycoside hydrolase family 87 α-1, 3-glucanase from *Paenibacillus glycanilyticus* FH11. FEBS J 287:2524-2543.

45. Adav SS, Ravindran A, Sze SK. 2013. Proteomic analysis of temperature dependent extracellular proteins from *Aspergillus fumigatus* grown under solid-state culture condition. J Proteome Res 12:2715-2731.

46. Hu H, da Costa RR, Pilgaard B, Schiøtt M, Lange L, Poulsen M. 2019. Fungiculture in termites is associated with a mycolytic gut bacterial community. mSphere 4:e00165-19.

47. Janesch B, Messner P, Schäffer C. 2013. Are the surface layer homology domains essential for cell surface display and glycosylation of the S-layer protein from *Paenibacillus alvei* CCM 2051T? J Bacteriol 195:565-575.

48. Kosugi A, Murashima K, Tamaru Y, Doi RH. 2002. Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. J Bacteriol 184:884-888.

49. Legg MS, Hager-Mair FF, Krauter S, Gagnon SM, Lòpez-Guzmán A, Lim C, Blaukopf M, Kosma P, Schäffer C, Evans SV. 2022. The S-layer homology domains of *Paenibacillus alvei* surface protein SpaA bind to cell wall polysaccharide through the terminal monosaccharide residue. J Biol Chem 298.

50. Buranov AU and Mazza G. 2008. Lignin in straw of herbaceous crops. Ind Crops Prod 28(3): 237-259.

51. Malgas S, Chandra R, Van Dyk J, Saddler J, Pletschke B. 2017. Formulation of an optimized synergistic enzyme cocktail, HoloMix, for effective degradation of various pre-treated hardwoods. Bioresour Technol 245:52-65.

52. Yang M, Zhang K-D, Zhang P-Y, Zhou X, Ma X-Q, Li F-L. 2016. Synergistic cellulose hydrolysis dominated by a multi-modular processive endoglucanase from *Clostridium cellulosi*. Front Microbiol 7:932.

53. Zhang K, Chen X, Schwarz WH, Li F. 2014. Synergism of glycoside hydrolase secretomes from two thermophilic bacteria cocultivated on lignocellulose. Appl Environ Microbiol 80:2592-2601.

54. Kim S-K, Russell J, Cha M, Himmel ME, Bomble YJ, Westpheling J. 2021. Coexpression of a β-d-xylosidase from *Thermotoga maritima* and a family 10 xylanase from *Acidothermus cellulolyticus* significantly improves the xylan degradation activity of the *Caldicellulosiruptor bescii* exoproteome. Appl Environ Microbiol 87:e00524-21.

55. Kim S-K, Chung D, Himmel ME, Bomble YJ, Westpheling J. 2019. Heterologous co-expression of two β-glucanases and a cellobiose phosphorylase resulted in a significant increase in the cellulolytic activity of the *Caldicellulosiruptor bescii* exoproteome. J Ind Microbiol Biotechnol 46:687-695.

56. Farkas J, Chung D, Cha M, Copeland J, Grayeski P, Westpheling J. 2013. Improved growth media and culture techniques for genetic analysis and assessment of

biomass utilization by *Caldicellulosiruptor bescii*. J Ind Microbiol Biotechnol 40:41-9.

57. Edwardson TG, Mori T, Hilvert D. 2018. Rational engineering of a designed protein cage for siRNA delivery. J Am Chem Soc 140:10439-10442.

58. Zurawski JV, Conway JM, Lee LL, Simpson HJ, Izquierdo JA, Blumer-Schuette S, Nookaew I, Adams MW, Kelly RM. 2015. Comparative analysis of extremely thermophilic *Caldicellulosiruptor* species reveals common and unique cellular strategies for plant biomass utilization. Appl Environ Microbiol 81:7159-7170.

Lane	Sample	M_r (kDa)
1,8	Benchmark protein ladder	-
2	Calkro_0111_ΔSLH	240
3	Calkro_0402_ΔSLH	160
4	Calkro_0072_ΔSLH	170
5	Calkro_2036_ΔSLH	60
6	Calhy_1629_ΔSLH	110
7	Calhy_2383_ΔSLH	200

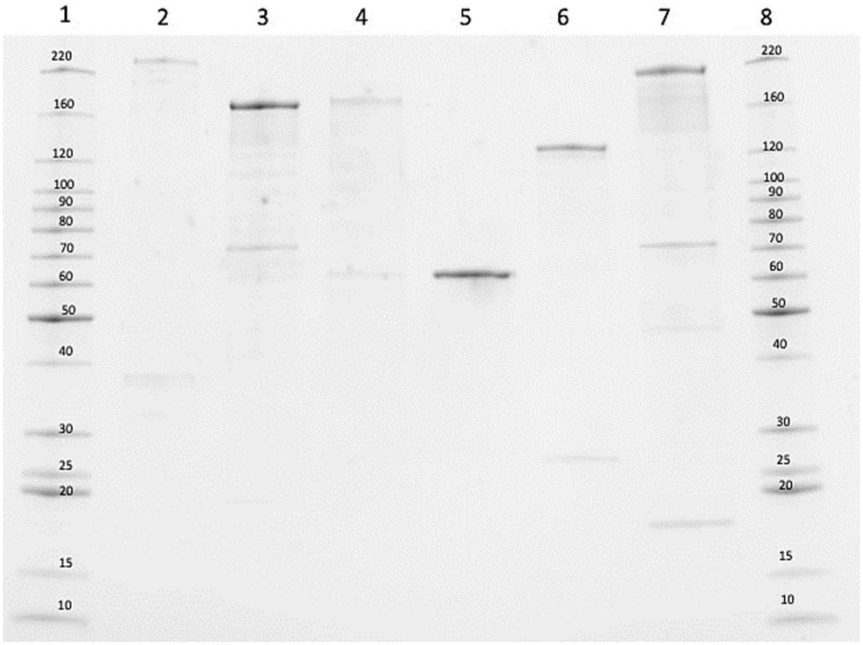


Figure 3.1 - SDS-PAGE gel of purified recombinant SLH-GHs produced in *E. coli* or *C. bescii* host. One hundred μg of each purified truncated SLH-GHs was resolved in 4-20% Mini-PROTEAN TGX stain-free gels. Lane 1,8: Benchmark protein ladder; Lane 2: Calkro_0111_ΔSLH (240 kDa); Lane 3: Calkro_0402_ΔSLH (160 kDa); Lane 4: Calkro_0072_ΔSLH (170 kDa); Lane 5: Calkro_2036_ΔSLH (60 kDa); Lane 6: Calhy_1629_ΔSLH (110 kDa); Lane 7: Calhy_2383_ΔSLH (200 kDa)

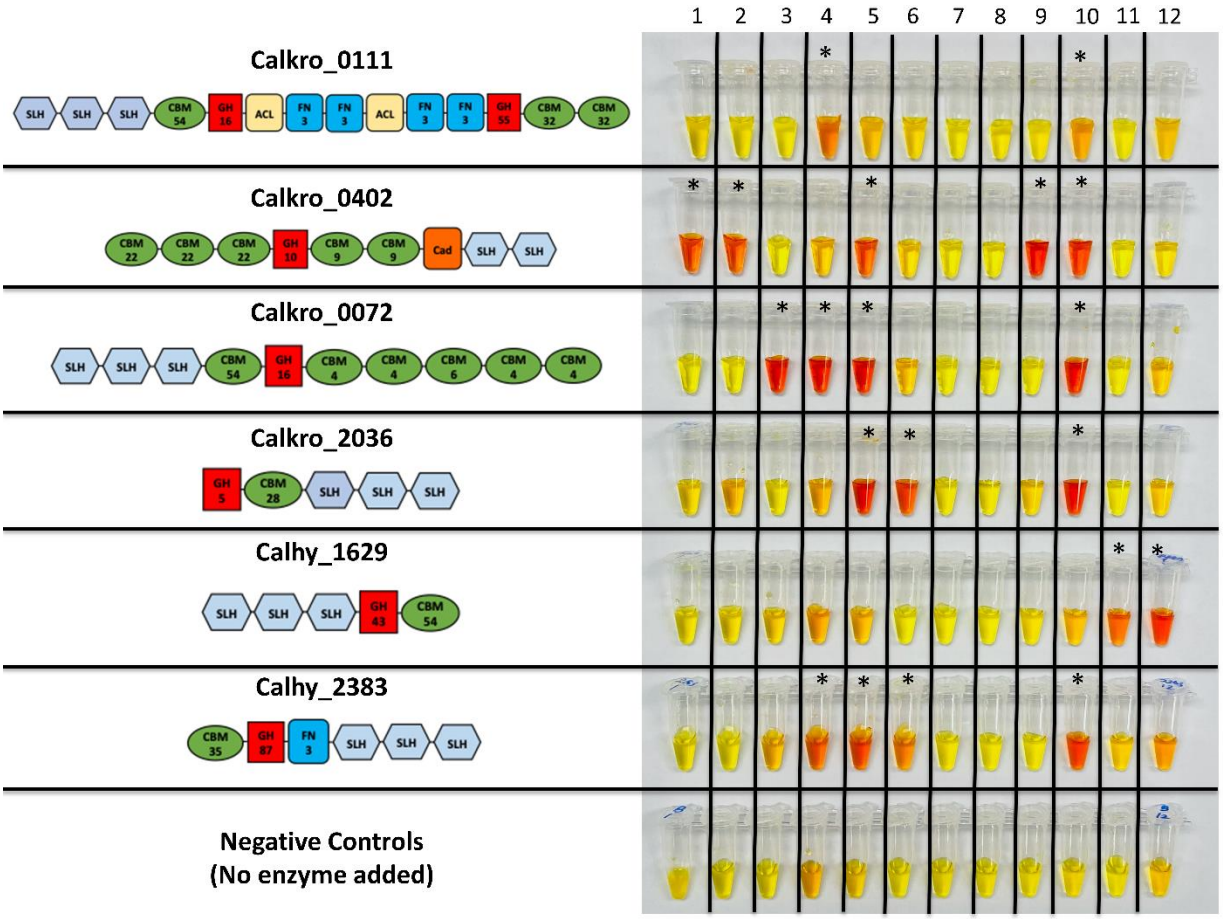


Figure 3.2 - *Caldicellulosiruptor* species and substrate specificity of their SLH-GHs. 1: Birchwood xylan(β-1,4-xylose); 2: Oat Spelt xylan (β-1,4-xylose with arabinose substitution); 3: Curdlan (β-1,3-glucose); 4: Laminarin (β-1,3-glucose with β-1,6 branches); 5: Lichenan (β-1,3 and β-1,4-glucose (ratio 1:2)); 6: Carboxymethyl cellulose (β-1,4-glucose); 7: Pustulan (β-1,6-glucose); 8: Tamarind xyloglucan (β-1,4, β-1,6 linkages, and α-1,6-xylose sidechain); 9: Wheat arabinoxylan (β-1,4-xylose-α-1,2/3-arabinose sidechain); 10: Barley-β-glucan (β-1,3 and β-1,4-glucose); 11: Arabinan (α-1,5-arabinose with α-1,2-L-arabinofuranosyl); 12: Debranched arabinan (α-1,5-arabinose). Surface-Layer Homology domains (SLH), Catalytic domains (Glycoside Hydrolase (GH)), carbohydrate binding domains (Carbohydrate Binding Module (CBM)), and domains with unknown function (e.g., Actin Cross-linking-Like (ACL), Cadherin-like Domain (Cad), Fibronectin type-III (FN3)). *indicates difference between sample and negative control.

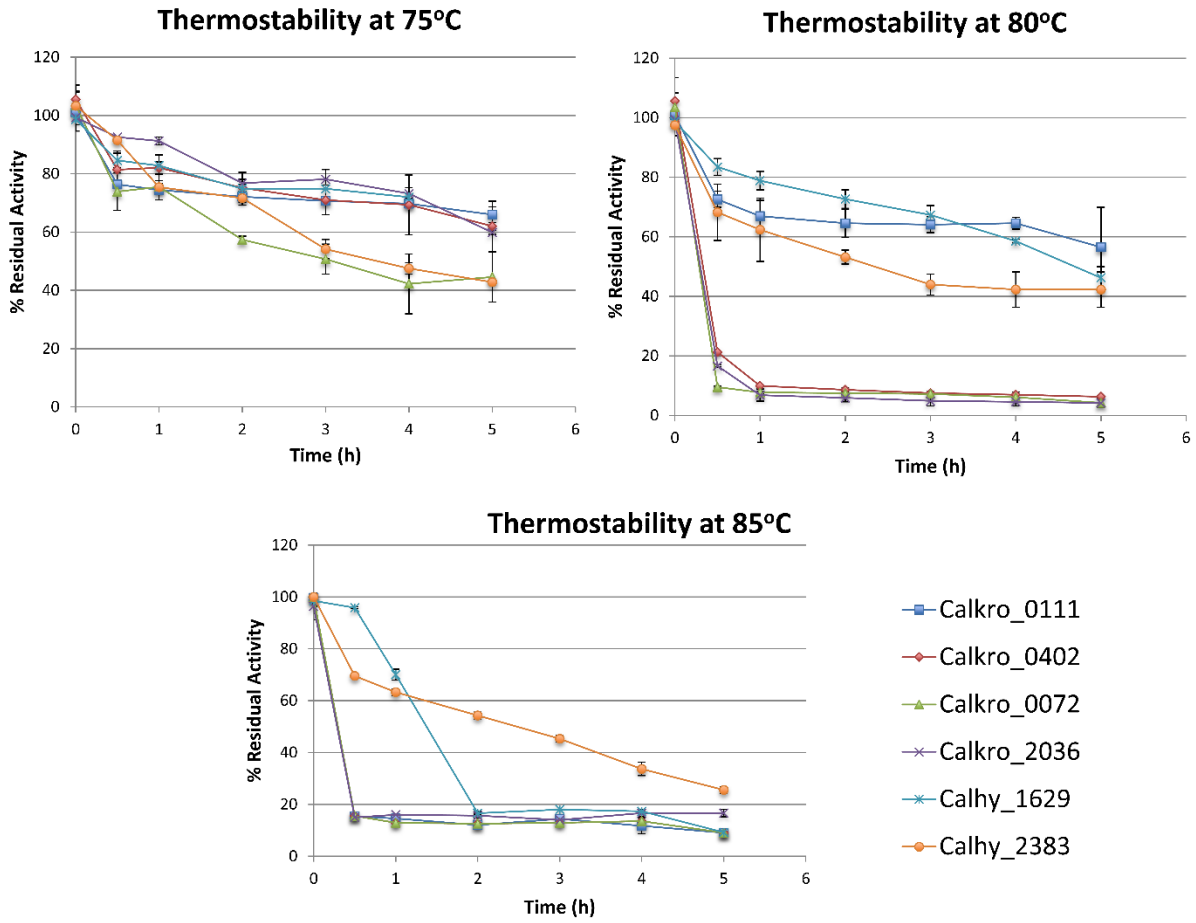


Figure 3.3 - Thermostability analysis of SLH-GHs at different temperature. A, 75ºC; B, 80ºC; C, 85ºC. Enzymatic activity was measured at 0, 0.5, 1, 2, 3, 4, and 5 hours. The residual activities are presented in percentages, based on their initial activities. All experiments were done in tripiclate and the error bars represent the standard deviation of each sample set.

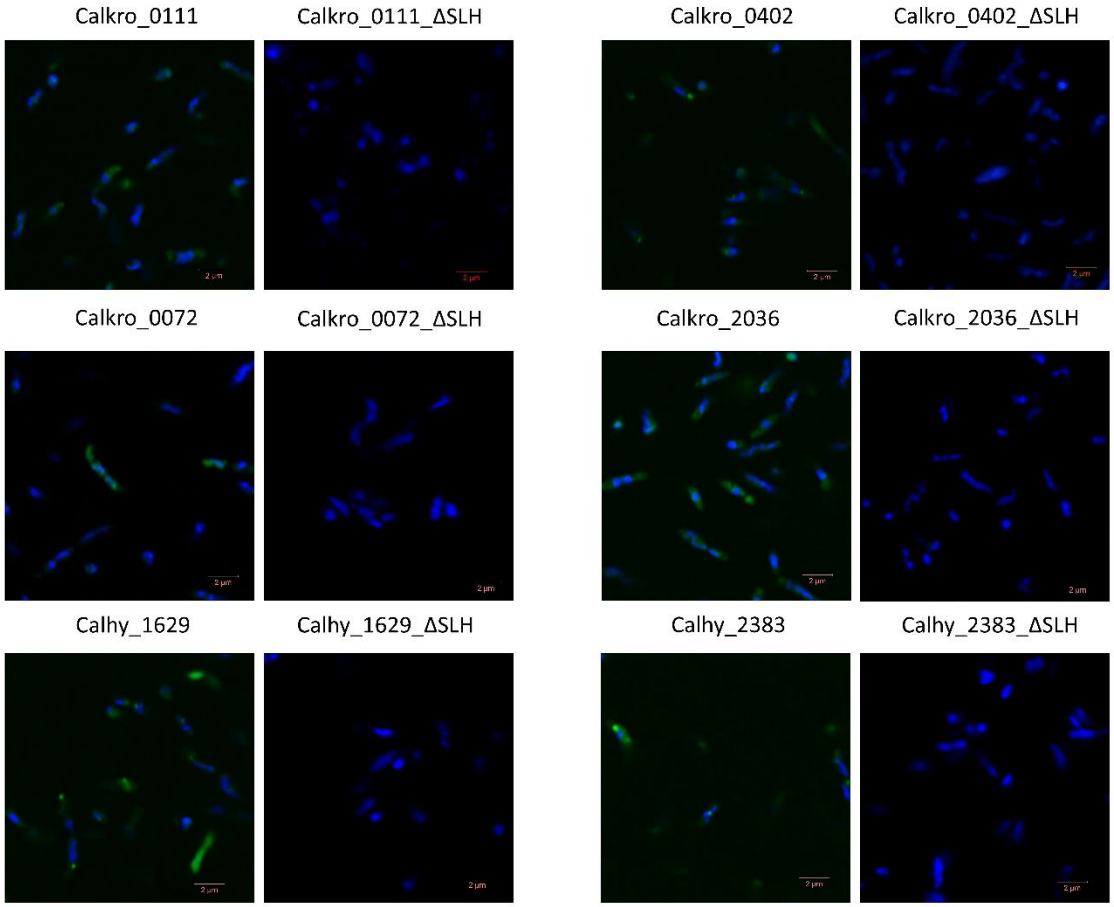


Figure 3.4 - Fluorescence microscopy showing surface bound GHs via SLH domains in different SLH-GH strains revealing that SLH domains in *Caldicellulosiruptor* SLH-GHs function as S-layer binding domains to associate the rest of domains on the cell surface. The expression strains have His-tags on the expressed protein, which can specifically bind to Ni-NTA fluorescent dye. All cells are labeled with Ni-NTA fluorescent dye (green) and are counterstained with DAPI (blue). Scale bars, 2 μm.

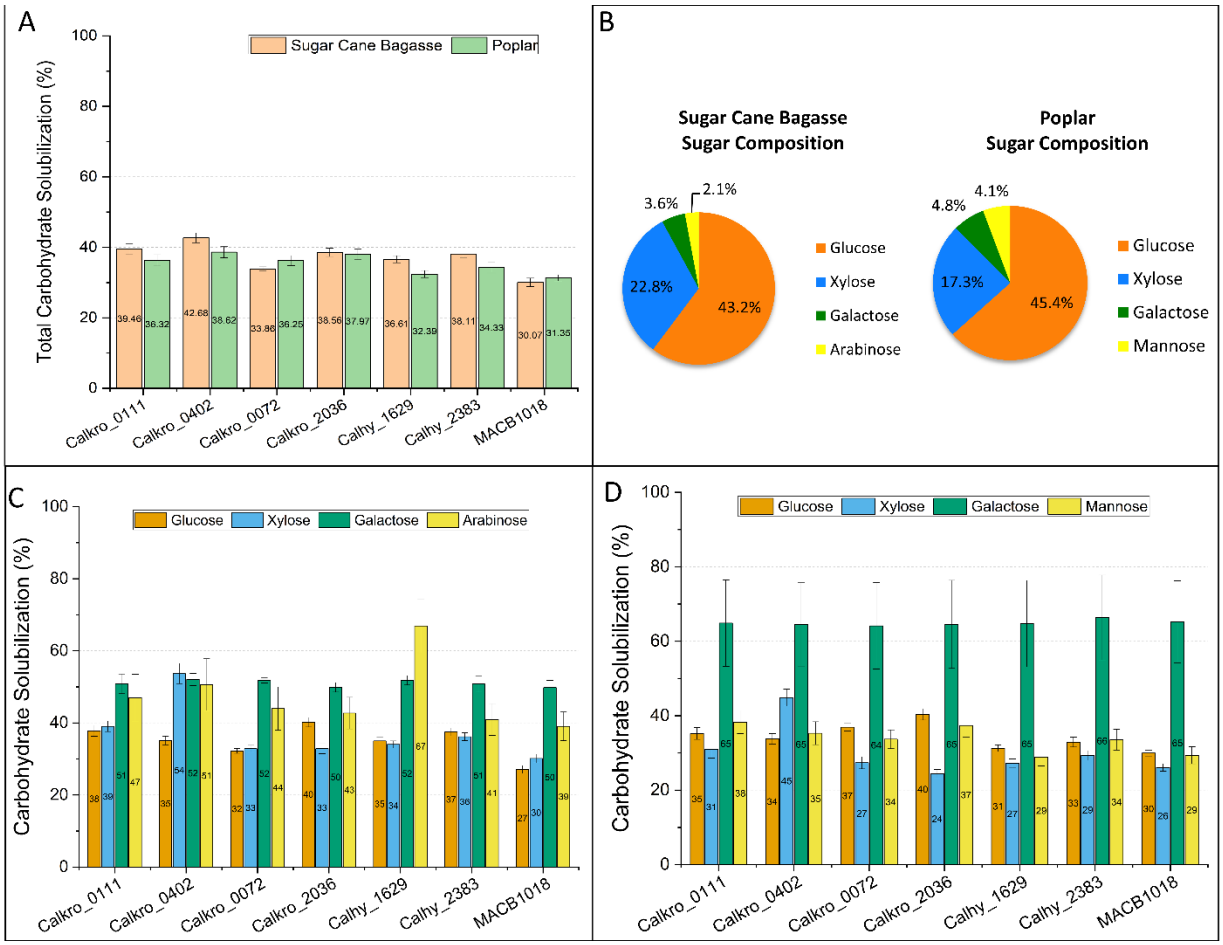


Figure 3.5 - Carbohydrate solubilization of different SLH-GH strains compared with their parent strains. A, total carbohydrate solubilization; B, different sugar composition in sugar cane bagasse and poplar; C, detailed carbohydrate solubilization in sugar cane bagasse; D, detailed carbohydrate solubilization in poplar. Experiments were done in tripiclate and error bars represent the standard deviation of each sample set.

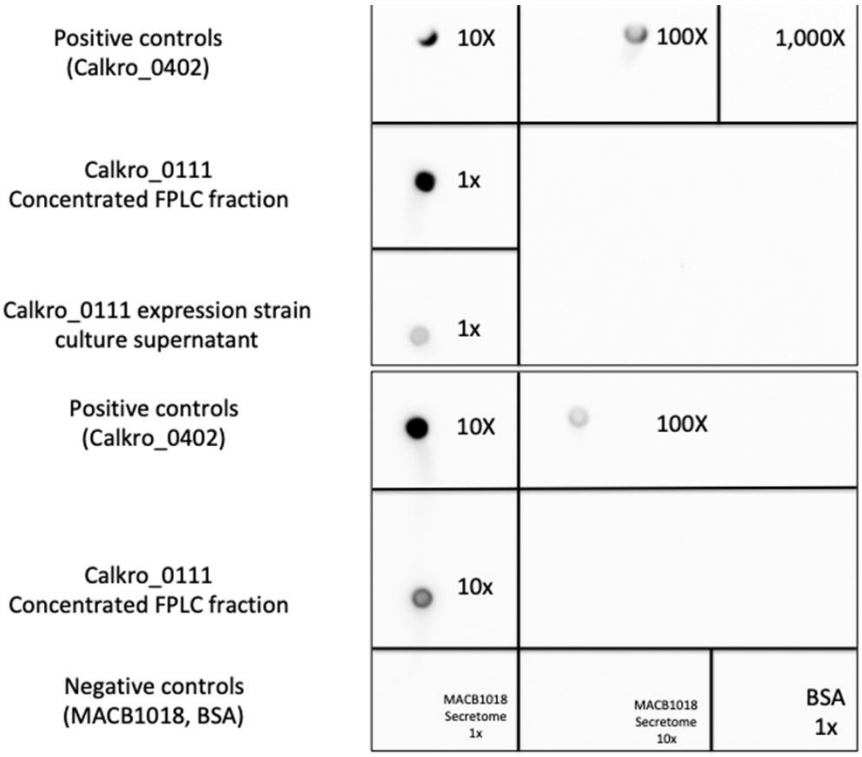


Figure 3.6 - His-tagged expressed protein detection using dot blot analysis confirming the expression of Calkro_0111 in *C. bescii*. The positive control was His-tagged Calkro_0402 expressed in E. coli and the negative controls was MACB1018 secretome and bovine serum albumin (BSA). Purified FPLC fractions indicated the positive results compared with other control.

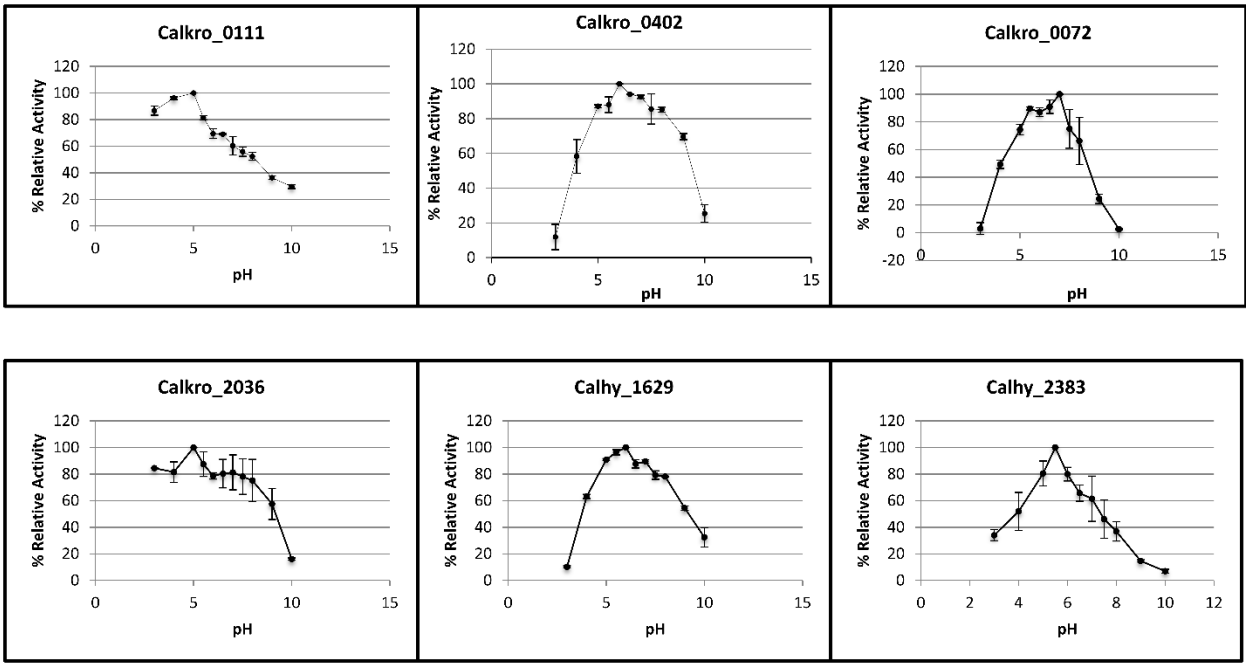


Figure 3.7 - pH optima of SLH-GHs

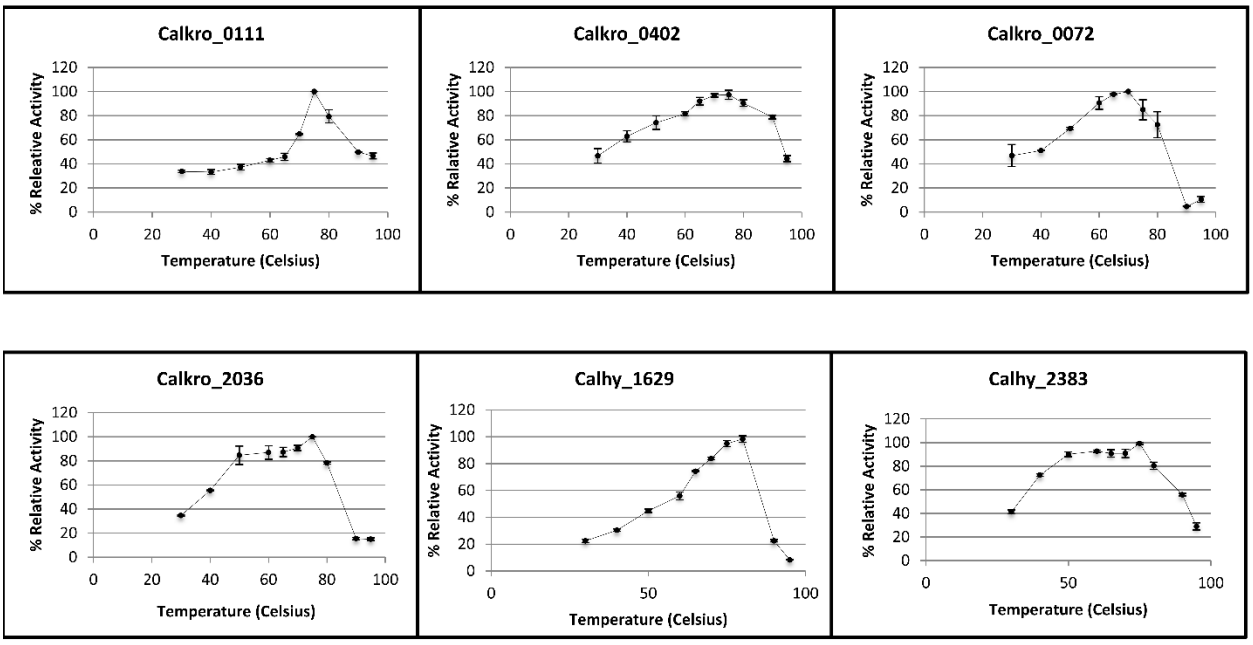


Figure 3.8 - Temperature optima of SLH-GHs

Table 3.1 - *Caldicellulosiruptor* sp. surface layer homology domain glycoside hydrolases

Domains (*)	Locus	M_r (truncated)	pH_{opt}	T_{opt} (°C)	Homolog
SLH-SLH-SLH-CBM54-FN3-**GH16**-ACL-FN3-FN3-ACL-FN3-FN3-**GH55**-CBM32-CBM32	Calkro_0111	240 kDa	5	75	-
CBM22-CBM22-CBM22-**GH10**-CBM9-CBM9-Cad-SLH-SLH	Calkro_0402	160 kDa	6	75	Calow_1924, Calla_0206, Calkr_2245
SLH-SLH-SLH-CBM54-**GH16**-CBM4-CBM4-CBM6-CBM4-CBM4	Calkro_0072	170 kDa	7	70	Calhy_0060
GH5-CBM28-SLH-SLH-SLH	Calkro_2036	60 kDa	5	75	Athe_0594, Calhy_2064, Csac_0678, Calkr_2007, Calow_0352
SLH-SLH-SLH-**GH43**-CBM54	Calhy_1629	110 kDa	6	80	-
CBM35-**GH87**-FN3-SLH-SLH-SLH	Calhy_2383	200 kDa	5.5	75	-

*Surface Layer Homology (SLH), Carbohydrate Binding Module (CBM), Glycoside Hydrolase (GH), Actin Cross-linking-Like (ACL), Cadherin-like Domain (Cad), Fibronectin type-III (FN3)

Table 3.2 - Specific activity of SLH-GHs on selected substrates (µmol reducing sugar per min per µmol protein)

Enzyme	Substrate 1	Substrate 2	Substrate 3	Substrate 4	Substrate 5
Calkro_0111	Laminarin	Barley-β-glucan	-	-	-
	2,324 ± 187	878 ± 35	-	-	-
Calkro_0402	Barley-β-glucan	Wheat arabinoxylan	Birchwood xylan	Oat spelt xylan	Lichenan
	4,048 ± 769	3,920 ± 305	3,846 ± 254	3,810 ± 455	2,399 ± 187
Calkro_0072	Laminarin	Lichenan	Curdlan	Barley-β-glucan	-
	4,291 ± 221	4,241 ± 260	3,494 ± 260	2,857 ± 187	-
Calkro_2036	Lichenan	CMC	Barley-β-glucan	-	-
	4,310 ± 218	3,231 ± 195	2,317 ± 293	-	-
Calhy_1629	Arabinan	Debranched arabinan	-	-	-
	1,585 ± 38	3,214 ± 158	-	-	-
Calhy_2383	Barley-β-glucan	Lichenan	Laminarin	CMC	-
	1,482 ± 48	1,470 ± 53	1,219 ± 48	1,078 ± 20	-

Table 3.3 - *In vitro* biomass hydrolysis after 24 hours of incubation

SLH-GH cocktail	Normalized sugar release (µg sugar/100 µg enzyme)	% improvement vs. Secretome only
Sugar cane bagasse		
Secretome only	15.3	-
25% secretome 75% SLH-GHs	17.1	12.1
50% secretome 50% SLH-GHs	20.2	32.4
75% secretome 25% SLH-GHs	27.3	78.8
SLH-GHs only	12.1	-20.9
Poplar		
Secretome only	4.2	-
25% secretome 75% SLH-GHs	6.1	47.6
50% secretome 50% SLH-GHs	8.6	107
75% secretome 25% SLH-GHs	17.5	321
SLH-GHs only	2.2	-47.6

Table 3.4 - Improvement in sugar cane bagasse carbohydrate solubilization of SLH-GH strains

Strain	carbohydrate			glucose			xylose			arabinose		
	sugar solub (mg)	%		sugar solub (mg)	%		sugar solub (mg)	%		sugar solub (mg)	%	
		solubilized	improved		solubilized	improved		solubilized	improved		solubilized	improved
MACB1018	76.2	29.5	-	42.1	27.1	-	24.7	30.1	-	2.955	39.1	-
Calkro_0111	101	39.5	33.6	58.8	37.8	39.6	32.1	39.1	29.6	3.55	47.0	20.4
Calkro_0402	109	42.7	44.5	54.6	35.1	29.7	44.1	53.7	78.1	3.82	50.6	29.6
Calkro_0072	87.1	33.9	14.7	50.1	32.2	19.0	26.9	32.8	8.9	3.32	44.0	12.6
Calkro_2036	99.2	38.6	30.6	62.5	40.2	48.5	27.0	32.9	9.0	3.23	42.7	9.4
Calhy_1629	94.1	36.6	24.0	54.4	35.0	29.3	27.9	34.0	12.9	5.06	66.9	71.3
Calhy_2383	97.6	38.1	29.1	58.3	37.5	38.4	29.7	36.2	20.0	3.09	40.9	4.8

Table 3.5 - Improvement in poplar carbohydrate solubilization of SLH-GH strains

Strain	carbohydrate			glucose			xylose			mannose		
	Sugar solub (mg)	%		Sugar solub (mg)	%		Sugar solub (mg)	%		sugar solub (mg)	%	
		solubilized	improved		solubilized	improved		solubilized	improved		solubilized	improved
MACB1018	78.6	31.4	-	47.6	30.0	-	15.8	26.1	-	4.2	29.4	-
Calkro_0111	91.0	32.4	15.6	55.9	35.2	17.5	18.7	30.9	18.4	5.5	38.3	30.3
Calkro_0402	96.8	34.3	23.0	53.7	33.8	12.8	27.2	44.9	71.9	5.1	35.3	20.2
Calkro_0072	90.8	36.3	15.4	58.7	36.9	23.3	16.6	27.4	4.7	4.8	33.7	14.7
Calkro_2036	95.2	38.0	20.9	64.2	40.4	34.8	14.8	24.4	-6.4	5.4	37.4	27.3
Calhy_1629	81.2	32.4	3.1	49.7	31.3	4.4	16.5	27.2	4.1	4.2	28.9	-1.5
Calhy_2383	86.0	34.3	9.3	52.3	32.9	9.9	17.8	29.3	12.3	4.8	33.6	14.3

Table 3.6 - Primers used in this study. Underlined portions identify Gibson Assembly overlap regions

Primer	Sequence	Use
ECalkro_0402 F	GACGACGACAAGATTCAAAGCAGCCAAACTAA	pET46 cloning
ECalkro_0402 R	GAGGAGAAGCCCGGTTACTTTTTGAGATTGTTAAGTGCTCTG	pET46 cloning
ECalkro_0072 F	GACGACGACAAGATAAATTCAGTTCAAATCTTCAAGGACGTT	pET46 cloning
ECalkro_0072 R	GAGGAGAAGCCCGGTTACTGGTCACTCACAACAGTCA	pET46 cloning
ECalkro_2036 F	GACGACGACAAGATAGGAAAATTATTTTAAAGTTTTGTGCAC	pET46 cloning
ECalkro_2036 R	GAGGAGAAGCCCGGTTATAGCTTGCCTGCTAAGTTCAAGAG	pET46 cloning
ECalhy_1629 F	GACGACGACAAGATGTATGTAGATAATGAAAG	pET46 cloning
ECalhy_1629 R	GAGGAGAAGCCCGGTTATTTTGTTTGAAACAATTTT	pET46 cloning
ECalhy_2383 F	GACGACGACAAGATGGAAGAACCCAATAGCAA	pET46 cloning
ECalhy_2383 R	GAGGAGAAGCCCGGTTAATTAGCTAAATTAGCA	pET46 cloning
Calkro_0111 F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGAAACGAGGGGGAGCA	Cloning into pJMC046
Calkro_0111 R	CCTCCTTTAATGGTGGTGGTGGTGATGGTTCTTGCTCCCATACACTTCAAACTC	Cloning into pJMC046
Calkro_0402 F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGAAAAAGCGACTTATAGCATTTCTTG	Cloning into pJMC046
Calkro_0402 R	CCTCCTTTAATGGTGGTGGTGGTGATGGTTCTTTTTGAGATTGTTAAGTGCTCTGAATAA	Cloning into pJMC046
Calkro_0072 F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGCGTGTAAAAAATCTTGTTGC	Cloning into pJMC046
Calkro_0072 R	CCTCCTTTAATGGTGGTGGTGGTGATGGTTCTGGTCACTCACAACAGTC	Cloning into pJMC046
Calkro_2036 F	CAAGGAGGTTTGGTGAGTAGTTATGAGGAAAATTATTTTAAAGTTTTGTGC	Cloning into pJMC046
Calkro_2036 R	GTCCTCCTTTAATGGTGGTGGTGGTGATGGTTTAGCTTGCCTGCTAAGTTCAAG	Cloning into pJMC046
Calhy_1629 F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGACCGAGGAAAGTGTTTATTTG	Cloning into pJMC046
Calhy_1629 R	CCTCCTTTAATGGTGGTGGTGGTGATGGTTTGCCAAAACTTTAATTTCATC	Cloning into pJMC046
Calhy_2383 F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGGTCAGAAGAATTTATAAAAAG	Cloning into pJMC046
Calhy_2383 R	CCTCCTTTAATGGTGGTGGTGGTGATGGTTATTAGCTAAATTAGCAATAG	Cloning into pJMC046
Calkro_0111 F	AGTGGTTTGGATACAAGTAAGTGG	SLH domain deletion
SCalkro_0111 R	TCCTGGTGTGCCGTTAAAAAC	SLH domain deletion

Table 3.6 (continued)

SCalkro_0402 F	CAGAGCACTTAACAATCTCAAAAAG	SLH domain deletion
SCalkro_0402 R	AATGTTAACAGCTACAGGTGTTTT	SLH domain deletion
SCalkro_0072 F	GGCGGTGGAGAAAATTCTGTC	SLH domain deletion
SCalkro_0072 R	CATAACTACTCACCAAACCTCCTTG	SLH domain deletion
SCalkro_2036 F	CTAAACCATCACCACCACCAC	SLH domain deletion
SCalkro_2036 R	CATCAGCAATTGGAACATCGTC	SLH domain deletion
SCalhy_1629 F	GATTCTGGGGTTCGTAACGATAAT	SLH domain deletion
SCalhy_1629 R	CATAACTACTCACCAAACCTCCTTG	SLH domain deletion
SCalhy_2383 F	GAGAATTTGGCAAGGTTTGCT	SLH domain deletion
SCalhy_2383 R	TGCTCTTGTTATGTTCCTTTGAGG	SLH domain deletion

CHAPTER 4

Improving biomass solubilization ability of *Caldicellulosiruptor bescii* using unique non-catalytic carbohydrate binding proteins (tāpirins)

Tunyaboon Laemthong[1], James R. Crosby[1], Ryan G. Bing[1], Mohamad M. J. Manesh[1], Michael W. W. Adams[2], and Robert M. Kelly[1,*]

[1]*Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695*

[2]*Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602*

Abstract

Plant biomass deconstruction is the first step for cellulose-degrading bacteria to potentially convert substrate into chemicals and fuels. Members in the genus *Caldicellulosiurptor* are cellulolytic and deploy unique carbohydrate-binding protein, called “tāpirins”. These proteins, located nearby type IV pili loci, are involved in biomass deconstruction and can be a targeted to help cells overcome plant biomass recalcitrance. Here, we proposed a novel technique to make *Caldicellulosiruptor bescii*, already a strong cellulose degrader, adhere to plant biomass substrate better by using non-native tāpirins from other species in the same genus. The accessibility of cellulosic enzymes to plant biomass could be pivotal for improved degradation. Tāpirins (Calhy_0908 from *Caldicellulosiruptor hydrothermalis*, Calkr_0826 from *Caldicellulosiruptor kristjanssonii, and* Calkro_0844 from *Caldicellulosiruptor kronotskyensis*) were compared *in vitro* and some *in vivo*. Calkr_0826 bound better to cellulosic substrates *in vitro* and *C. bescii* strain with overexpressed Calkr_0826 exhibited approximately 15% higher poplar biomass solubilization and 40% more acetate production compared with its parent strain. However, further studies on other lignocellulosic plant biomasses needed to be done to prove the improved ability in different plant biomass deconstruction.

Introduction

Cellulolytic microbes deploy multidomain secreted and/or surface-associated enzymes to deconstruct plant biomass (1, 2). These enzymes are complex and contain catalytic domains (Glycoside Hydrolase (GH)) and non-catalytic domains (e.g., Carbohydrate Biding Domain (CBM), Surface-Layer Homology (SLH), Fibronectin type III) (3). Generally, the CBMs are responsible for enhancing enzymatic activity towards plant biomass recalcitrance (4, 5). CBMs act specifically on plant biomass carbohydrates. Increases in CBM concentration reportedly improved overall enzymatic activity by providing extra accessibility of the enzyme to the carbohydrate substrate (6). For example, the addition of an X2-CBM3 polypeptide increased GH74 activity ~4-fold in *Paenibacillus polymyxa* A18 towards the insoluble substrate, enhanced overall plant biomass hydrolysis (7). Fusing *Thermotoga maritima* CBM9-2 into *Aspergillus niger* XynAm1, a thermophilic xylanse (GH11), at the N-terminus with an additional linker improved simple sugar release from sugarcane xylan hydrolysis 5-fold compared to only XynAm1 (8). Additionally, fusing CBM6 into *Bacillus subtilis* endo-xylanase (GH11), which lacks CBMs, increased of reducing sugar release from sugarcane bagasse 17% compared with the wild-type form (9). Adding different CBMs into a multifunctional *Acetivibrio thermocellus* GH5 increased its catalytic reaction towards several polysaccharides (10). These examples underscore the importance of non-catalytic domains present in multi-domain enzymes.

Caldicellulosiruptor species are promising candidates for use in Consolidated BioProcessing (CBP) due to their ability to grow on variety of complex carbohydrates (11-13). In particular, *Caldicellulosiruptor bescii* has been developed as a chemical production platform due to the established genetic tools (14-18). *Caldicellulosiruptor* species have surface-associated glycoside hydrolases (19) and novel carbohydrate binding proteins, known as tāpirins (20), to facilitate plant biomass deconstruction. Engineering the surface of *C. bescii* by using surface-associated enzyme improved

lignocellulosic plant biomass deconstruction (Laemthong et al, 2022 (in press)). The synergism between catalytic domains and non-catalytic domains enhances overall enzymatic activity. A kinetic model of *Trichoderma* cellulases indicated cellulose hydrolysis was not necessarily affected by the improved substrate binding ability of cellulases, but there are other intrinsic factors that impacted overall enzymatic activity (21). Are there any other independent non-catalytic proteins that can help enzymes overcome recalcitrance in plant biomass?

Tāpirins are proteins that bind specifically to crystalline cellulosic substrates and are found only in the genus *Caldicellulosiruptor* (20, 22). Tāpirins (average M_r ~70 kDa) are one of the mechanisms that *Caldicellulosiruptor* sp. use for insoluble cellulose attachment. From their crystallization and *in vitro* characterization, tāpirins from less cellulolytic species bound better to microcrystalline substrates (22). Moreover, deletion of tāpirin encoding gene in the genome potentially resulted in the lack of ability of *C. bescii* to bind to Avicel (22). Less cellulolytic species, like *Caldicellulosiruptor hydrothermalis*, may bind more tightly to growth substrates for their survival in the same habitat of other species like *C. bescii*. In nature, scavenging is a vital skill for species like *C. hydrothermalis* or *C. kristjanssonii* that have less efficient biomass degrading enzymes. Moreover, one of *Caldicellulosiruptor* sp. studies proposed mechanism indicated that tāpirins coordinate with type IV pili to adhere to cellulose (23). This leads to an intriguing question for strong plant biomass degrader like *C. bescii*. Could enhanced binding to cellulose improve its ability to utilize carbohydrate contents in plant biomass? Here, we propose a different strategy to improve plant biomass deconstruction by overexpressing non-native tāpirins on the cell surface of *C. bescii* to make them bind more tightly to crystalline cellulose substrate in order to provide more accessibility to plant biomass degrading enzymes. Site-directed mutational study was also done for *in vitro* studies to investigate which protein

residues are the most important for cellulose binding. Both *in vitro* and *in vivo* results were investigated to determine the relative merits of using various tāpirins.

Results and Discussion

Possible binding site of tāpirins. The fact that tāpirins bind to cellulose has been clearly established (20), although the specific details of their binding site were not determined. In order to understand the substrate binding mechanism of tāpirins, site-directed mutational studies were done to test the hypothesis that the aromatic plane of each tāpirin is involved in cellulosic substrate binding (**Figure 1**). The proposed strategy was to replace specific residues with cysteines to create disulfide bonds using information from the solved crystal structure of wild-type tāpirins. Under anaerobic conditions during protein production, disulfide bonds would then be created. Once the disulfide bonds were created, the repositioning loop in the structure would seal the aromatic plane, resulting in reduced binding ability. After production and purification, *in vitro* binding assays were performed to detect any changes in substrate binding. However, densitometry analysis indicated that there were no changes in substrate binding between mutated Calhy_0908 and Calkro_0844 and their wild-type forms. Unexpectedly, the mutated Calkro_0826 bound better to cellulose than its wild-type form (see **Table 4.1**). This was opposite to our initial hypothesis that the mutated tāpirins would bind less to substrate. One possibility was that either disulfide bonds, if they formed, did not entirely cover the plane leading to exposing aromatic rings to the environment or that the disulfide bonds were not actually present in the structure. However, crystal structure studies of these tāpirins are needed to draw a solid conclusion. Crystallization trials were conducted but no crystals of a quality needed for structural analysis were obtained.

Improving cellulosic substrate binding of *C. bescii* by overexpressing non-native tāpirins.

C. bescii, a strong cellulose degrader, encodes a tāpirins (Athe_1870) in its genome. However, Athe_1870 bound less tightly to Avicel compared to Calhy_0989 and Calkr_0826 from less cellulolytic species, *C. hydrothermalis* and *C. kronotskyensis*, respectively (22). Calhy_0908, Calkr_0826, and Calkro_0844 were expressed in *C. bescii* under the S-layer promoter (P_{slp}). Fluorescence microscopic images suggest that *C. bescii* strains expressing Calhy_0908, Calkr_0826, Calkr_0826_2SDM were found in closer proximity to Avicel, compared with the Calkro_0844 and the parent strains (**Figure 4.2**). Fewer cells of Calkro_0844 and MACB1018 strains were found mainly in the supernatant.

Calkr_0826 strain and MACB1018 strains were then grown on printer paper. The post-fermentation surface of substrate was then visualized using Scanning Electron Microscopy (SEM) (see **Figure 4.3**). SEM images showed that these strains were able to bind to cellulosic substrate. No cells were observed in an abiotic control and the surface of substrate stayed intact with some debris next to fibers. Interestingly, the post-fermentation substrate surface of MACB1018 and Calkr_0826 strain was quite different. The surface of substrate that Calkr_0826 grew on was more porous and deformed compared with MACB1018. Fibers were relatively smaller than on the post-fermentation substrate surface of the MACB1018 strain. The post-fermentation substrate structure of MACB1018 strain was less degraded than for the Calkr_0826 strain, and the surface seemed to stay more intact similar to the abiotic sample.

Quantitative analysis of the live cell binding assay was done to validate the qualitative observation above. After cells were incubated with substrate, the mixture was then separated into bound and unbound fractions. The bound fractions were plated and grown anaerobically. As described in **Figure 4.4 and Table 4.2**, the number of cells in the bound fraction of Calhy_0908, Calkr_0826, and Calkr_0826_2SDM strains was higher than Calkro_0844 and MACB1018 strains. There were no

differences between Calkro_0844 and MACB1018 strains. The only major difference observed was between the Calkr_0826 strain when compared with the parent strain. The quantitative data here support the conclusion from fluorescence the microscopic images above - cells were closer in proximity with the substrate surface when overexpressing nonnative tāpirins from less cellulolytic species, such as Calkr_0826 (*C. kristjanssonii*), on the cell surface of the parent strain.

Lignocellulosic plant biomass solubilization. Changes in lignocellulosic plant biomass solubilization were determined for expression strains with different tāpirins. Recall the Calkr_0826 strain bound more tightly to cellulosic substrate. The strain was used to compare its binding ability to lignocellulosic plant biomass compared with the parent strain. As shown previously in **Figure 4.4** and **Table 4.3**, Calkr_0826 strains (with and without the residue mutations) bound better to Avicel than the Calkro_0844 and the parent strain. Calhy_0908 bound slightly better to Avicel, but not as well as Calkr_0826 strains; the Calhy_0908 data also suggest better binding although with less statistical significance. On the other hand, Calkro_0844 bound to Avicel similarly to its parent strain. Calkro_0844 from *C. kronotskyensis*, one of strongest cellulose degrader in *Caldicellulosiruptor* sp., may bind less tightly to cellulose due to its more efficient plant biomass degrading enzymes (22). The poplar biomass solubilization results (see **Figure 4.5** and **Table 4.4**) indicated that Calkr_0826_2SDM, Calkr_0826, and, Calhy_0908 strains were able to solubilize poplar biomass better when compared to their parent strain 15%, 13%, and 10%, respectively; this also seems to be the case for Calkro_0826_2SDM but with less statistical significance. No significant difference was observed between the Calkro_0844 strains and the parent strain for biomass solubilization. Mass loss during fermentation is not be the only basis for evaluating these strains; the amount of acetate production, indicative of fermentation intensity, was determined for each strain as well. For the Calhy_0908,

Calkr_0826, Calkr_0826_2SDM strains increased acetate production 21%, 39%, and 41%, respectively compared with their parent strain.

To further test ability to solubilize plant biomass, the Calkr_0826 strain was grown on different lignocellulosic plant biomasses such as wheat straw and sugar cane bagasse. In grass cell wall, ferulic acid (3-(3-methoxy-4-hydroxymethyl)-2-propenoic acid) is often found attached primarily to arabinoxylan and cross-linked between polysaccharide-lignin complex (24). The major difference between these lignocellulosic plant biomasses is that wheat straw lacks ferulic acid that is linked in the lignin structure resulting in less recalcitrant cell wall, whereas sugar cane bagasse has more complexed lignin structure with linked ferulic acid and arabinoxylan (25). Increasing ability to attach to substrate more tightly to these substrates might improve chances of biomass degrading enzymes from *C. bescii* to deconstruct plant biomass better.

Conclusions and Future Directions

While *Cadicellulosiruptor bescii* is a prolific cellulose degrader, there are still some missing elements needed to fully improve *C. bescii* strains. Improving plant biomass deconstruction remains a challenge. We provide another path for developing strains further by increasing cell binding to the substrate to enhance cellulose degrading enzymes targeting. *C. bescii* was engineered with nonnative tāpirins from other members in the same genus. Expression of non-native tāpirins from less cellulolytic species, such as Calkr_0826 from C. *kristjanssonii* and Calhy_0908 from *C. hydrothermalis*, improved lignocellulosic plant biomass solubilization and hence acetate production. Since Calkr_0826 and Calhy_0908 bound better to cellulose *in vitro*, their ability to deconstruct lignocellulosic plant biomass (poplar) improved.

Even though tāpirins are known to bind to cellulose, comprehensive structural studies of their binding sites are still needed to further improve the substrate binding. Truncated versions of tāpirins could help identify binding site of tāpirin binding sites. Furthermore, tāpirin expression strains should be tested with other lignocellulosic substrates that have different carbohydrate composition and recalcitrance, such as sugar cane bagasse and wheat straw, in order to examine whether improved binding ability could overcome recalcitrant plant biomass.

Materials and methods

Bacterial Strains, plasmids, and reagents. Plasmid construction and recombinant protein expression were done using the following *Escherichia coli* strains: NEB 5-alpha (New England Biolabs, Ipswich, MA), Rosetta 2(DE3) pLysS (Fisher Scientific). Overexpression of tāpirins were performed in C. bescii MACB1018 (17). Genes of interest were amplified from *Calicellulosiruptor* sp. genomic DNA using NEB Monarch kit and polymerase chain reaction (PCR) (NEB Phusion High-Fidelity DNA polymerase or NEB Q5 High-Fidelity DNA Polymerase). *E. coli* expression plasmids were done using pET-45 plasmids provided by the DOE Joint Genome Institute (JGI), as described in (22). pJMC046 replicating vector containing 6 x His tag on the C-terminus for *C. bescii* expression (26) using NEB HiFi Builder (New England Biolabs, Ipswisch, MA). Plasmid sequences were confirmed by Sanger sequencing (Azenta Life Sciences, Morrisville, NC). Plasmid were maintained at -80oC in LB medium + 15% glycerol and 50 μg/mL kanamycin, 50 μg ampicillin, and/or 34 μg/mL chloramphenicol, as appropriate.

Site-directed mutation study. Protein production *E. coli* strains were obtained from (22). Two cysteine residues were replaced in the specific sites of each original genes, as shown in **Table 4.2**. The

nucleotide sequence was changed to TGT or TGC by PCR reaction using primers containing mutated nucleotide sequence and followed by kinase, ligase and *Dpn*I reaction in order to introduce cysteine residues to the protein structure.

Protein expression and purification. Protein expression from *E. coli* strains was performed in 2x YT medium (10 g/L yeast extract, 16 g/L tryptone, 5 g/L sodium chloride) with appropriate antibiotics (50 μg/mL kanamycin or 50 μg/mL plus 34 μg/mL chloramphenicol) at 37°C with 250-rpm agitation. Once the OD600 reached 0.8, the culture was cooled and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C for 18 h. Cells were harvested at 6,000 x g for 10 min. Cell pellets were kept at -20°C prior to lysis and purification. Cell pellets were thawed at room temperature and then resuspended in 20 mM sodium phosphate, 500 mM sodium chloride, pH 8.1. Cell suspensions were lysed twice in a French press (Sim-Aminco) at ~15,000 psi. Cell lysate was heat-treated at 60°C for 20 min to remove heat labile protein from *E. coli*. The cell lysate was centrifuged at 25,000 x g for 30 min at 4°C, and the supernatant was passed through a 0.22 μm filter to obtain cell-free extract. Protein was purified using 5 mL HisTrap Ni-Sepharose (Cytiva) immobilized metal affinity chromatography (IMAC) columns, operated according to the manufacturer's instructions. The chromatography steps were performed using a Biologic DuoFlow FPLC (Bio-Rad). The purity of purified proteins was assessed via SDS-PAGE using 4-15% Mini-PROTEAN TGX stain-free gels (Bio-Rad) using Laemilli's method with a Benchmark protein ladder (Life Technologies). The gel images were analyzed using densitometry (ImageJ). The concentration of purified protein was determined using Pierce BCA protein assay (Thermo Fisher) with Bovine Serum Albumin (BSA) as a protein standard.

***In vitro* tāpirin binding assay.** The protocol was modified from (20, 22). Full-length of purified recombinant proteins were tested on various substrates. All substrates were prepared as following: soaked in 100 mL of binding buffer (50 mM MES and 3.9 mM sodium chloride at pH 7.2) overnight at 70^{o}C and then dried at 70^{o}C overnight. One hundred μg of protein were mixed with 10 mg of prepared substrates in thermomixer (Eppendorf) at 70^{o}C with 500 rpm for 1 hr. Washed sand was used as a non-binding control. Mixed samples were then centrifuged at 13,000 x g to separate supernatant (unbound fraction) and pellet (bound fraction). The bound fraction was then washed three times by resuspending in binding buffer and centrifuging at 13,000 x g. The washed bound fractions were finally resuspended in 250 μL of binding buffer. Equal volumes of both bound and unbound fractions were then mixed with Laemilli loading buffer containing 5% of 2-mercaptoethanol and boiled at 95^{o}C for 30 min prior to SDS-PAGE gel (4-15% Mini-PROTEAN TGX stain-free gels (Bio-Rad)). The percentage binding of each sample was quantified using ImageJ software by comparing to the same amount of initial protein loading.

Genetic manipulation of *C. bescii*. Competent cells of MACB1018 strain were prepared as described in (22, 26). The electroporation was performed as described in (Laemthong et al, 2022 (in press)), with the exception that the electroporation was initiated at 2.0 kV, 400 Ω. The voltage and resistance were adjusted depending on the size of plasmids. Electroporated cells were immediately transferred to preheated low osmolality complex (LOC) medium (27) and incubated at 70^{o}C for 1 h for recovery. After recovery, transformants were passaged into defined modified medium DSM 516 (28) with 50 μg/mL kanamycin selection and 24 mM NH_4Cl. After 2-3 days, the grown culture was plated and embedded in DSM516 medium with 1.5% agar and grown anaerobically at 65^{o}C. Single colonies were selected and screened in liquid medium. The transformed DNA was isolated from liquid culture

and sequenced. Plate purifications were done at least twice for each strain. Primers used are listed in **Table 4.5**.

Fluorescence microscopy. To visualize how cells bound to substrate, 1 mL of overnight culture grown by inoculating $5x10^{10}$ cells 75ºC without agitation on Avicel was drawn after the culture was grown overnight. The drawn sample was mixed in the final concentration of 1.5% v/v glutaraldehyde. Cells were stained with acridine orange, as described in (29). Cell imaging was performed using Nikon eclipse 50i microscope with a Plan Flour x 100 (numerical aperture 1.3) oil emersion objective and Excelis microscope camera. Images were processed using Captavision+^{TM} software.

Scanning Electron Microscope (SEM). Cultures were grown by inoculating total $5x10^{10}$ cells overnight at 75ºC without agitation. The cultures were then cooled to room temperature. Cells were fixed in the final concentration of 2.0% V/V paraformaldehyde and 2.5% V/V of glutaraldehyde in culture medium for 1 h. After cell fixation, samples were removed and washed three times with 1 x Phosphate-Buffered Saline (PBS). Samples were then dehydrated in 70%, 95%, and 100% ethanol, respectively. The substrates were processed for critical point drying. Prior to taking SEM images, the samples were sputter coated with Au/Pd.

Live-cell binding quantification. Cultures were grown on Avicel by cells overnight at 75ºC with 150 rpm. The cultures were then cooled to room temperature. Cells were separated from substrate by centrifuging at 400 x g for 1 min. Planktonic cells were collected from the supernatant and counted. The cells were resuspended with 100 mg Avicel to the final of $5x10^{10}$ cells/mL. The mixture was then incubated anaerobically at 75ºC with 150 rpm for 1 h. The bound and unbound fractions were obtained

by centrifuging the mixture at 400 x g. Both bound and unbound fractions were then plated and embedded in DSM516 medium with 1.5% agar and grown anaerobically for 3-4 days. CFU was then calculated based on the numbers of colonies seen on the serial diluted plate.

Biomass solubilization. Protocol was derived from (Laemthong et al, 2022 (in press)). Briefly, plant biomass substrates were milled and sieved. The 40/80 sieve fraction was washed 3 times with 75ºC distilled water to remove any soluble sugars and subsequently dried at 75ºC overnight. *Caldicellulosiruptor* strains were passaged three times on modified 671d medium on each biomass substrate, supplemented 250 mM uracil and 50 μg/mL kanamycin. Solubilization cultures were prepared by inoculating $5x10^8$ cells/mL, and grown at 75ºC with 250 rpm for 7 days. All experiments were done in triplicate. After the fermentation, cultures were cooled to room temperature and harvested by centrifuging at 5,000 rpm for 10 min at room temperature. Soluble fermentation products were measure using Waters Arc3 HPLC, as described in (14) using a refractive index detector and Rezex-ROA column with a 5 mM sulfuric acid mobile phase flowing at 0.6 mL/min at 50ºC (Model 2414, Phenomenex, Waters Corp.). Residual substrate biomass was washed three times and dried at 75ºC for 24 h. The final dried plant biomass weight was recorded to calculate biomass solubilization.

Acknowledgments

This work was supported by the US Department of Agricultural (2018-67021-27716), the US National Science Foundation (CBET-1802939), and the US Department of Energy BER Awards DE-SC0019391 and DE-SC0022192. T Laemthong acknowledges support from the Government of Thailand. JR Crosby acknowledges support from a US DoEd GAANN Fellowship (P200A160061). RG Bing acknowledges support from an NIH Biotechnology Traineeship (NIH T32 GM008776-16). We acknowledge the North Carolina State University Analytical Instrumentation Facility, Dr. Aaron Bell, and Chuck Mooney for their assistance with variable pressure scanning electron microscopy (Hitachi SU3900).

References

1. Lochner A, Giannone RJ, Rodriguez Jr M, Shah MB, Mielenz JR, Keller M, Antranikian G, Graham DE, Hettich RL. 2011. Use of label-free quantitative proteomics to distinguish the secreted cellulolytic systems of *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis*. Appl Environ Microbiol 77:4042-4054.
2. Blumer-Schuette SE, Lewis DL, Kelly RM. 2010. Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. Appl Environ Microbiol 76:8084-8092.
3. Fagan RP, Fairweather NF. 2014. Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol 12:211-222.
4. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769-781.
5. Abbott DW, van Bueren AL. 2014. Using structure to inform carbohydrate binding module function. Curr Opin Struct Biol 28:32-40.
6. Bernardes A, Pellegrini V, Curtolo F, Camilo C, Mello B, Johns M, Scott J, Guimaraes F, Polikarpov I. 2019. Carbohydrate binding modules enhance cellulose enzymatic hydrolysis by increasing access of cellulases to the substrate. Carbohydr Polym 211:57-68.
7. Pasari N, Adlakha N, Gupta M, Bashir Z, Rajacharya GH, Verma G, Munde M, Bhatnagar R, Yazdani SS. 2017. Impact of Module-X2 and Carbohydrate Binding Module-3 on the catalytic activity of associated glycoside hydrolases towards plant biomass. Sci Rep 7:1-15.

8. Li Y, Song W, Yin X, Rao S, Zhang Q, Zhou J, Li J, Du G, Liu S. 2022. Enhanced catalytic performance of thermophilic GH11 xylanase by fusing carbohydrate-binding module 9-2 and linker for better synergistic degradation of wheat bran. Process Biochem 121:349-359.

9. Hoffmam ZB, Zanphorlin LM, Cota J, Diogo JA, Almeida GB, Damásio AR, Squina F, Murakami MT, Ruller R. 2016. Xylan-specific carbohydrate-binding module belonging to family 6 enhances the catalytic performance of a GH11 endo-xylanase. New Biotechnol 33:467-472.

10. Walker JA, Takasuka TE, Deng K, Bianchetti CM, Udell HS, Prom BM, Kim H, Adams PD, Northen TR, Fox BG. 2015. Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. Biotechnol Biofuels 8:1-20.

11. Straub CT, Bing RG, Wang JP, Chiang VL, Adams MW, Kelly RM. 2020. Use of the lignocellulose-degrading bacterium *Caldicellulosiruptor bescii* to assess recalcitrance and conversion of wild-type and transgenic poplar. Biotechnol Biofuels 13:1-10.

12. VanFossen AL, Verhaart MR, Kengen SM, Kelly RM. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl Environ Microbiol 75:7718-7724.

13. Lee LL, Crosby JR, Rubinstein GM, Laemthong T, Bing RG, Straub CT, Adams MWW, Kelly RM. 2020. The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in 'Caldi World'. Extremophiles 24:1-15.

14. Bing RG, Straub CT, Sulis DB, Wang JP, Adams MW, Kelly RM. 2022. Plant biomass fermentation by the extreme thermophile *Caldicellulosiruptor bescii* for co-production

of green hydrogen and acetone: Technoeconomic analysis. Bioresour Technol 348:126780.

15. Lipscomb GL, Conway JM, Blumer-Schuette SE, Kelly RM, Adams MW. 2016. A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii*. Appl Environ Microbiol 82:4421-4428.

16. Rodionov DA, Rodionova IA, Rodionov VA, Arzamasov AA, Zhang K, Rubinstein GM, Tanwee TN, Bing RG, Crosby JR, Nookaew I. 2021. Transcriptional regulation of plant biomass degradation and carbohydrate utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. mSystems 6:e01345-20.

17. Williams-Rhaesa AM, Poole FL, Dinsmore JT, Lipscomb GL, Rubinstein GM, Scott IM, Conway JM, Lee LL, Khatibi PA, Kelly RM. 2017. Genome stability in engineered strains of the extremely thermophilic lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol 83:e00444-17.

18. Zhang K, Zhao W, Rodionov DA, Rubinstein GM, Nguyen DN, Tanwee TN, Crosby J, Bing RG, Kelly RM, Adams MW. 2021. Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for bio-based chemical production. mSystems 6:e01351-20.

19. Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, Zurawski JV, Lee LL, Blumer-Schuette SE, Kelly RM. 2016. Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. J Biol Chem 291:6732-6747.

20. Blumer-Schuette SE, Alahuhta M, Conway JM, Lee LL, Zurawski JV, Giannone RJ, Hettich RL, Lunin VV, Himmel ME, Kelly RM. 2015. Discrete and structurally unique

proteins (tāpirins) mediate attachment of extremely thermophilic *Caldicellulosiruptor* species to cellulose. J Biol Chem 290:10645-10656.

21. Gao D, Chundawat SP, Sethi A, Balan V, Gnanakaran S, Dale BE. 2013. Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. Proc Natl Acad Sci 110:10922-10927.

22. Lee LL, Hart WS, Lunin VV, Alahuhta M, Bomble YJ, Himmel ME, Blumer-Schuette SE, Adams MWW, Kelly RM. 2019. Comparative biochemical and structural analysis of novel cellulose binding proteins (tapirins) from extremely thermophilic *Caldicellulosiruptor* species. Appl Environ Microbiol 85.

23. Khan AM, Hauk VJ, Ibrahim M, Raffel TR, Blumer-Schuette SE. 2020. *Caldicellulosiruptor bescii* adheres to polysaccharides via a type IV pilin-dependent mechanism. Appl Environ Microbiol 86:e00200-20.

24. Xu F, Sun R-C, Sun J-X, Liu C-F, He B-H, Fan J-S. 2005. Determination of cell wall ferulic and p-coumaric acids in sugarcane bagasse. Anal Chim Acta 552:207-217.

25. Buranov AU, Mazza G. 2008. Lignin in straw of herbaceous crops. Ind Crops Prod 28:237-259.

26. Conway JM, Crosby JR, McKinley BS, Seals NL, Adams MW, Kelly RM. 2018. Parsing *in vivo* and *in vitro* contributions to microcrystalline cellulose hydrolysis by multidomain glycoside hydrolases in the *Caldicellulosiruptor bescii* secretome. Biotechnol Bioeng 115:2426-2440.

27. Farkas J, Chung D, Cha M, Copeland J, Grayeski P, Westpheling J. 2013. Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by *Caldicellulosiruptor bescii*. J Ind Microbiol Biotechnol 40:41-9.

28. Conway JM, McKinley BS, Seals NL, Hernandez D, Khatibi PA, Poudel S, Giannone RJ, Hettich RL, Williams-Rhaesa AM, Lipscomb GL. 2017. Functional analysis of the glucan degradation locus in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. Appl Environ Microbiol 83:e01828-17.

29. VanFossen AL, Ozdemir I, Zelin SL, Kelly RM. 2011. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Biotechnol Bioeng 108:1559-1569.

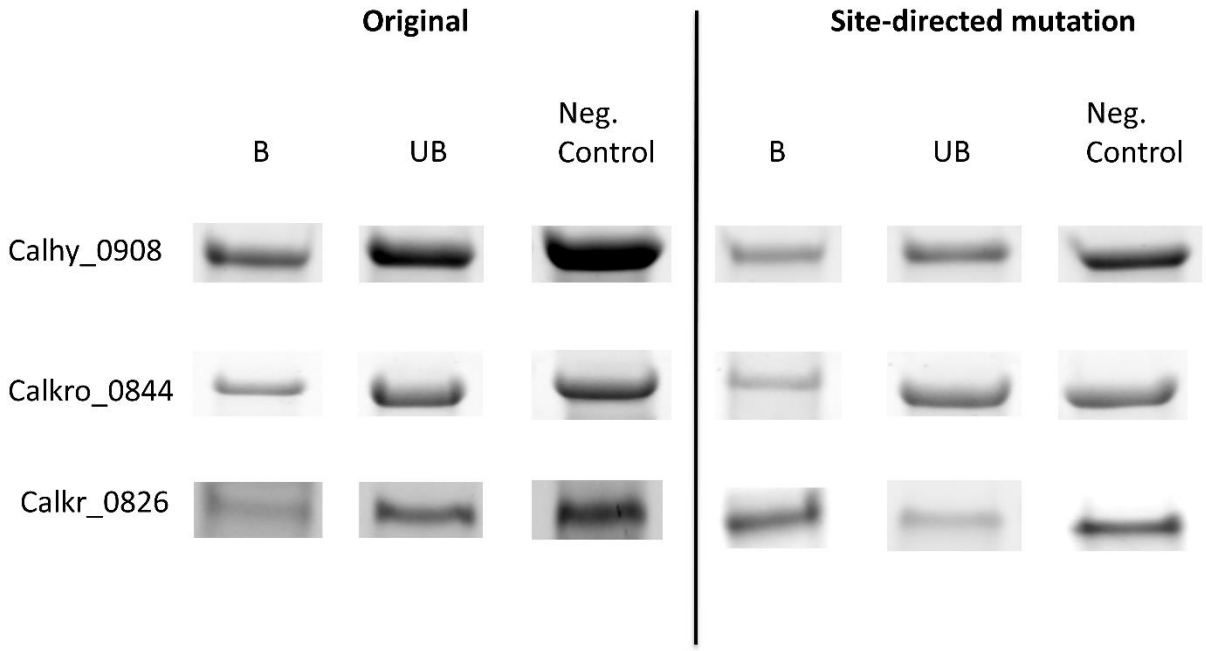


Figure 4.1 - *in vitro* binding assay comparing differences in substrate binding between original and site-directed mutational tāpirins. B refers to bound fraction, UB refers to unbound fraction, Neg. Control refers to negative control (no substrate added).

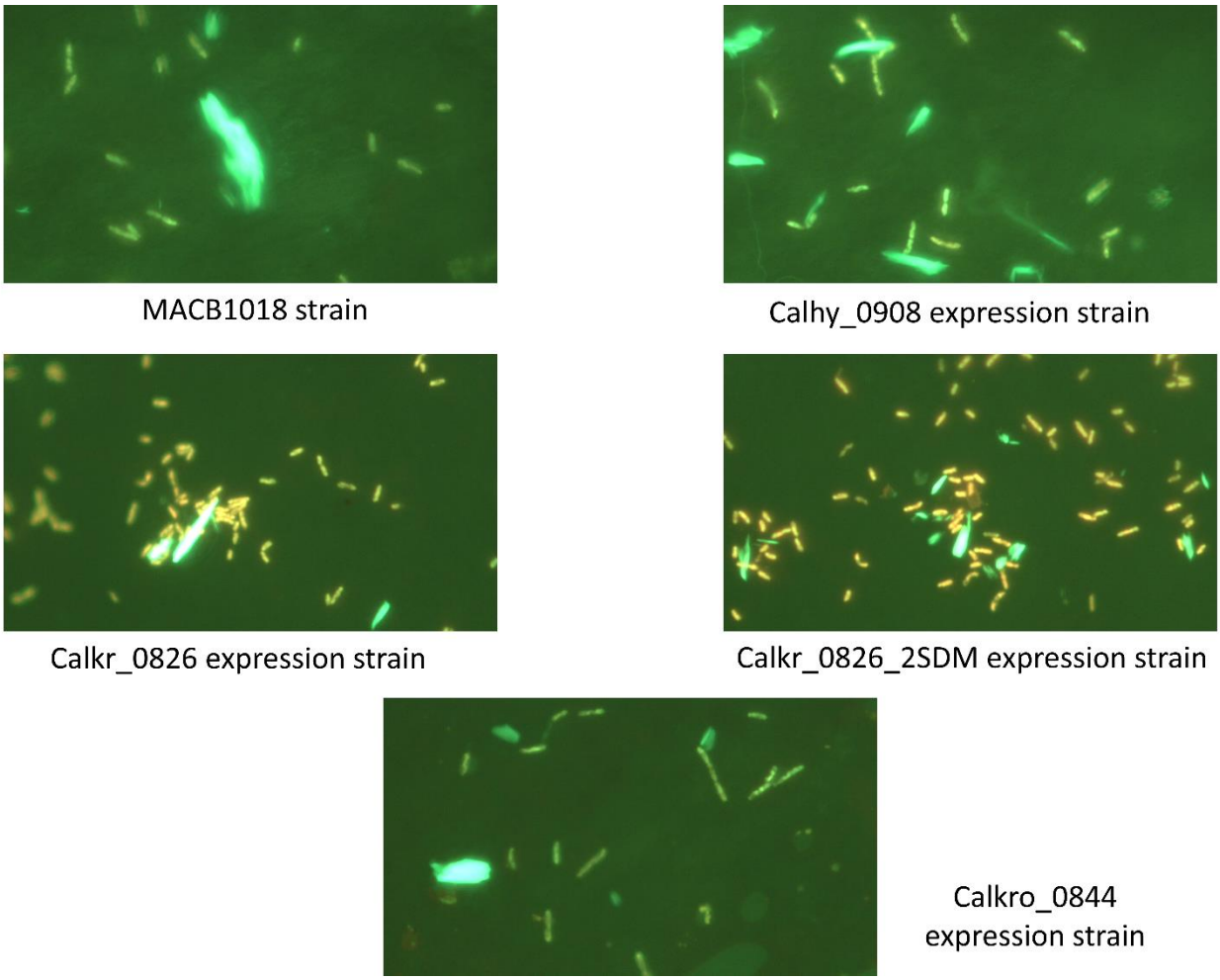


Figure 4.2 - Florescence microscopy showing bound cells on substrate surface. Strain with overexpressed Calhy_0908, Calkr_0826, and Calkr_0826 with site-directed mutations, the cells seemed to bind to substrate better since more cells were seen on the substrate surface. Green florescence indicates substrate (Avicel), orange fluorescence indicates cells.

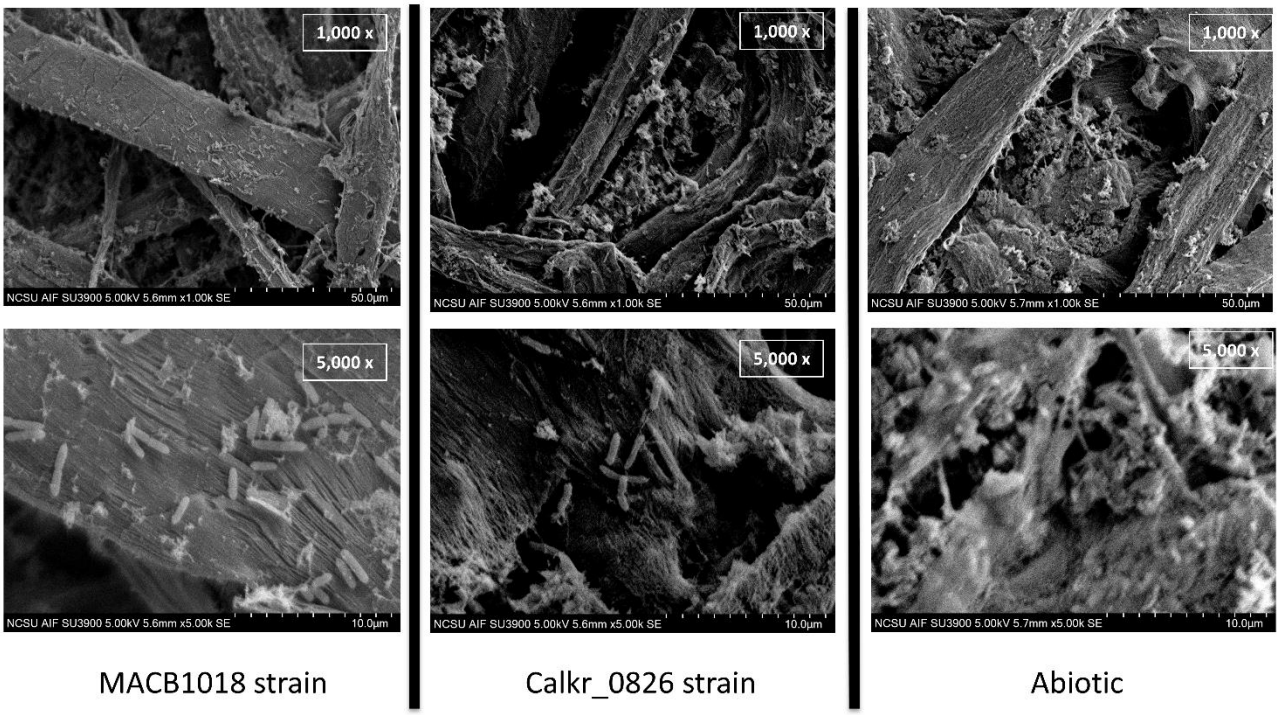


Figure 4.3 - Scanning Electron Microscopic (SEM) images showing deformation of post-fermentation substrate surface of Calkr_0826 strains compared with its parent strain (MACB1018). Substrate surface that Calkr_0826 strain grew on appeared to be more porous and deformed than its parent strain. The substrate surface hardly stayed intact to each other compared with parent strain and abiotic samples.

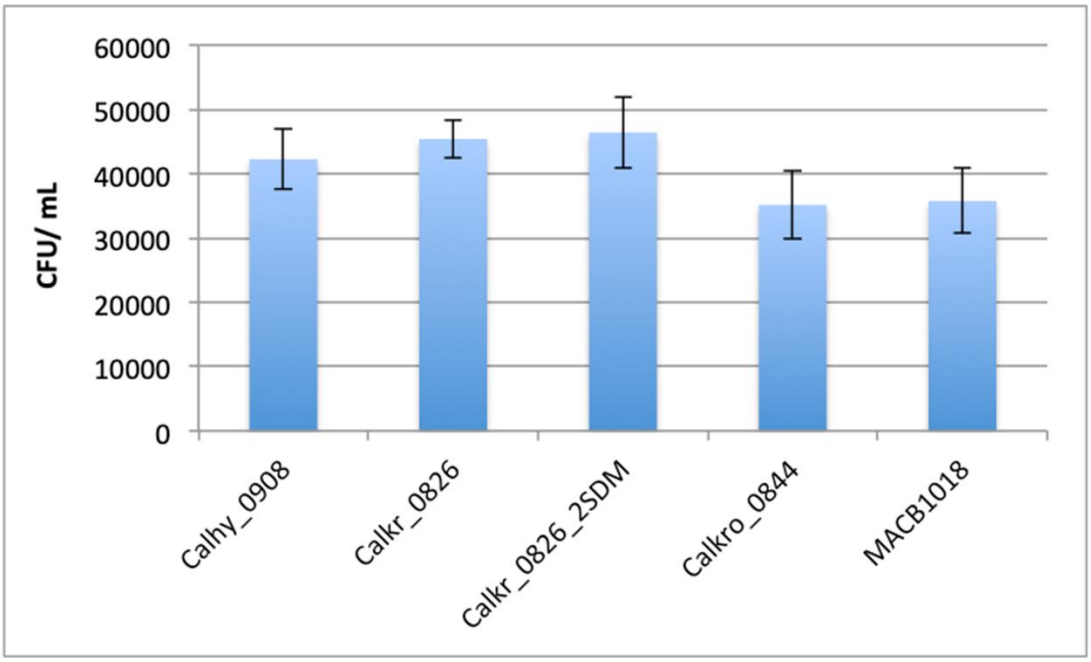


Figure 4.4 - Quantitative analysis of live cell binding to substrate (Avicel). After incubation with Avicel, bound fraction of each strain was separated, plated and grown anaerobically. Colony Forming Units (CFU) were calculated based on numbers of colony found on serial diluted plates. Calhy_0908, Calkr_0826, and Calkr_0826_2SDM bound higher to substrate compared with their parent strains. No changes were observed for Calkro_0844 strain. Experiments were done in tripiclate and error bars represent the standard deviation of each sample set.

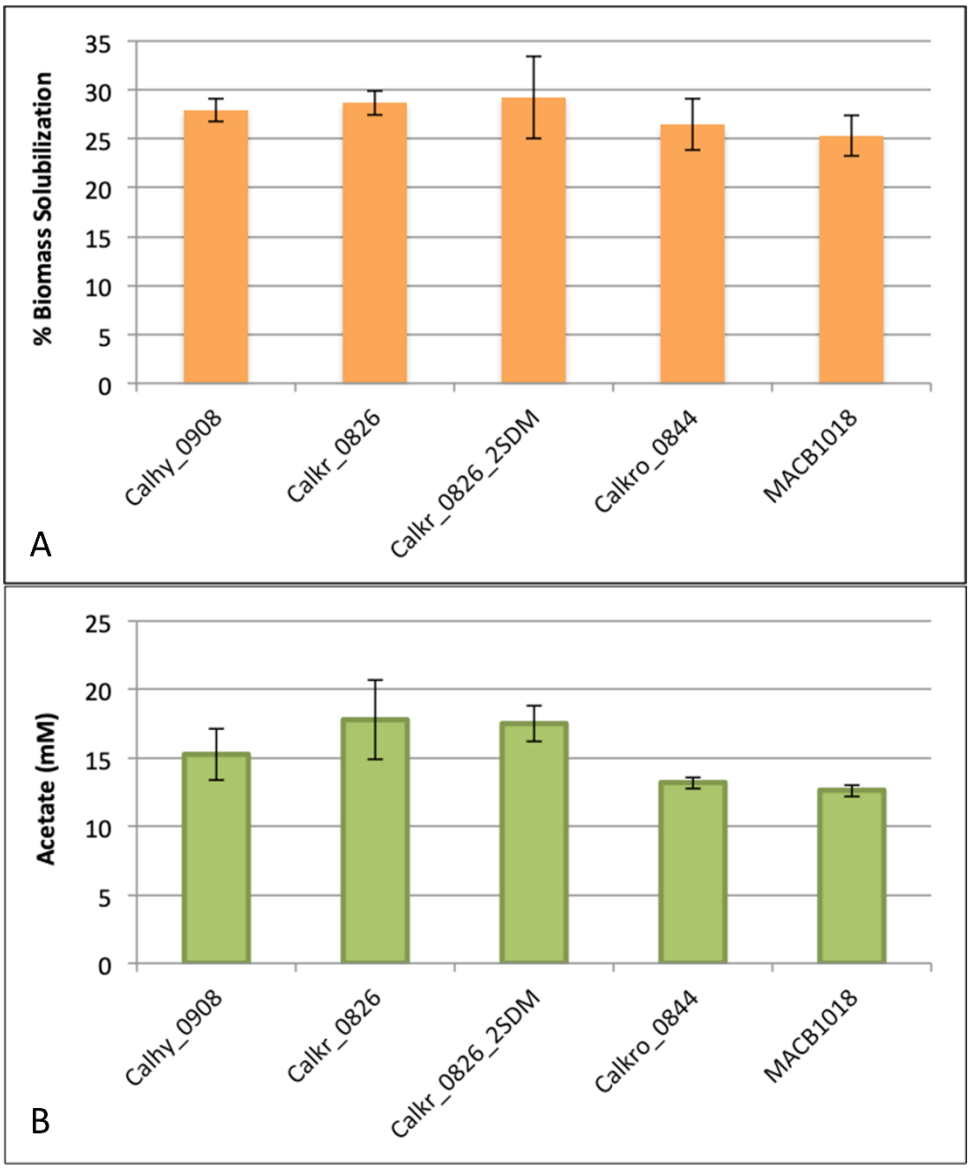


Figure 4.5 - Poplar biomass solubilization of all strains after 7-day fermentation. A, poplar biomass solubilization. B, the amount of acetate production from each strain. MACB1018, a parent strain, was used as a control for comparison.

Table 4.1- Protein residues that were replaced with cysteine residues to create double site-directed mutations.

Tāpirin	Protein residue#1	Protein residue#2
Calhy_0908	Phenylalanine#335	Tyrosine#446
Calkr_0826	Tyrosine#356	Phenylalanine#402
Calkro_0844	Tyrosine#397	Tyrosine#446

Table 4.2- Densitometry analysis showing changes in binding between original and site-directed mutated tāpirins from *in vitro* binding assay.

Sample	Original	Site-directed mutations
Calhy_0908	38.93±1.42	38.56±1.09
Calkro_0844	32.36±0.01	29.57±1.03
Calkr_0826	44.77±2.10	53.27±3.73

Table 4.3- Live cell binding assay result of bound fraction from tāpirin expression strain after incubated with Avicel

Sample	Colony Forming Units/mL (CFU/mL)
Calhy_0908	42,250±4630
Calkr_0826	45,416±2982
Calkr_0826_2SDM	46,416±5558
Calkro_0844	35,166±5305
MACB1018	35,750±5055

Table 4.4- Post-fermentation analysis of poplar biomass solubilization and acetate production of tāpirins expression strains. Percentage improvement was calculated based on their parent strain (MACB1018).

Strain	% Biomass solubilization	% improvement	Acetate production (mM)	% improvement
Calhy_0908	27.97±1.19	10.27	15.27±1.34	21.16
Calkr_0826	28.70±1.25	13.16	17.83±2.92	41.42
Calkr_0826_2SDM	29.34±4.20	15.26	17.57±0.93	39.40
Calkro_0844	26.50±2.65	4.47	13.16±0.40	4.44
MACB1018	25.37±2.10	-	12.60±0.41	-

Table 4.5- Primers used in this study. Underlined portions identify Gibson assembly overlap regions.

Primer	Sequence	Use
Calhy_0908_F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGGCCTGGGACTCTAACATC	Cloning into pJMC046
Calhy_0908_R	TTAATGGTGGTGGTGGTGATGGTTTTTCACGACAAAATGGCGGGC	Cloning into pJMC046
Calkr_0826_F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGTCTAGCACGAGTAGCTGG	Cloning into pJMC046
Calkr_0826_R	TTAATGGTGGTGGTGGTGATGGTTTTTGATAATCATCCCGCGGCG	Cloning into pJMC046
Calkro_0844_F	AATCATACAAGGAGGTTTGGTGAGTAGTTATGTCACTGAACCAGTCCAC	Cloning into pJMC046
Calkro_0844_R	CCTCCTTTAATGGTGGTGGTGGTGATGGTTACGCGTTACCATATTCCGA	Cloning into pJMC046