UNLV | UNIVERSITY LIBRARIES

Biochemical Characterization of Fsa1572 from Fervidibacter Sacchari, the First Hyperthermophilic GH50 with B-1, 4-Glucanase Activity

Jonathan Covington

Repository Citation

BIOCHEMICAL CHARACTERIZATION OF FSA1572 FROM *FERVIDIBACTER SACCHARI*,

THE FIRST HYPERTHERMOPHILIC GH50 WITH β-1,4-GLUCANASE ACTIVITY

By

Jonathan K. Covington

Bachelor of Science – Biological Sciences
University of Nevada, Las Vegas
2020

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science – Biological Sciences

School of Life Sciences
College of Sciences
The Graduate College

University of Nevada, Las Vegas
May 2023

Thesis Approval

The Graduate College
The University of Nevada, Las Vegas

April 19, 2023

This thesis prepared by

Jonathan Covington

entitled

Biochemical Characterization of Fsa1572 from *Fervidibacter Sacchari*, the First Hyperthermophilic GH50 with B-1, 4-Glucanase Activity

is approved in partial fulfillment of the requirements for the degree of

Master of Science – Biological Sciences
School of Life Sciences

Brian Hedlund, Ph.D.
Examination Committee Chair

Helen Wing, Ph.D.
Examination Committee Member

Kurt Regner, Ph.D.
Examination Committee Member

Ernesto Abel-Santos, Ph.D.
Graduate College Faculty Representative

Alyssa Crittenden, Ph.D.
Vice Provost for Graduate Education &
Dean of the Graduate College

ABSTRACT

The bacterium *Fervidibacter sacchari* is an aerobic hyperthermophile that catabolizes various polysaccharides and is the only cultivated member of the class *Fervidibacteria* within the phylum *Armatimonadota*. Among its glycoside hydrolase (GH) cache is an enzyme from GH family 50 (GH50), an understudied family with only 25 characterized representatives and two known activities from 1,518 predicted members in the Carbohydrate-Active EnZyme (CAZy) Database. Here, we expressed, purified, and functionally characterized an extracellular GH50 from *F.* sacchari called Fsa1572. Using colorimetric assays, we show it has novel β-1,4-glucanase activity and only weak agarose activity that is typical for GH50 enzymes. The purified enzyme has a wide temperature range of 60-95 °C (optimal 80 °C), making it the first characterized hyperthermophilic representative of GH50. The enzyme also has a broad pH range of at least 5.5-11 (optimal 6.5-10). Fsa1572 possesses K_M and k_{cat} parameters of 12.6 mM and 4.62 s^{-1}, respectively. Finally, a phylogenetic analysis of Fsa1572 and other GH50 enzymes revealed a unique phylogenetic position for Fsa1572 that is distant from other characterized enzymes and related to yet-uncharacterized GH50s from genomes of thermophilic archaea. Fsa1572 is the first characterized GH enzyme of *F. sacchari* and is both functionally and phylogenetically novel.

DEDICATION

I lovingly dedicate this work to my grandma, Alice Anne Berly, whose endless fountain of

support has had a massive role in the success of my academic journey. Fly away, butterfly.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Polysaccharides

Polysaccharides are carbohydrates comprising hundreds to thousands of monosaccharide units. These monosaccharide units are bound by glycosidic linkages and are sometimes branched. Polysaccharides are ubiquitous, as they are produced and used by all forms of life. Many organisms use polysaccharides for energy storage, such as glycogen in animals, fungi, protozoa, and bacteria, and starch in plants (Vidal and Vanegas-Calerón, 2019; Wertz and Goffin, 2021). Another source is agar, a gelatinous substance found in the cells walls of some species of *Rhodophyta*, commonly known as red algae, that comprises two polysaccharides: agarose and agaropectin (Zhang et al., 2019a). Polysaccharides excreted from the cells of archaea and bacteria, or exopolysaccharides, are used to protect communities of microbes that make up microbial mats and biofilms from stresses such as desiccation and grazing (Rossi and De Philippis, 2015; Heredia-Ponce et al., 2021). However, the biggest natural source of polysaccharides comes in the form of the building block of all plant fiber: lignocellulose.

Lignocellulose is the form of plant biomass that makes up the most common plants, such as grasses, trees, and the wastes leftover from the agricultural harvest of crops like corn, rice, and sugarcane (Cabrera et al., 2014; Guerriero et al., 2016). Lignocellulose contains two types of polysaccharides: cellulose and hemicellulose. Because of the prominence of lignocellulose in plant cell walls and the plethora of lignocellulosic plant life, lignocellulose is the most abundant form of biomass on Earth (Jha et al., 2020). Cellulose alone comprises up to 50% of material in the plant cell wall and is far and away the most abundant polysaccharide, and indeed the most prominent natural organic substance, on Earth (Aspinall, 1970; Zeng et al., 2017). The value of cellulosic material both to humanity for paper and textiles and plant life as a building block is partly because

of its extreme durability, with a half-life of 5-8 million years at 25 °C (Wolfenden & Snider, 2001). However, we know that cellulosic material degrades far faster, and that is thanks to cellulolytic organisms.

1.2 Cellulolytic Microbes

Cellulose degradation is widespread throughout all forms of life to depolymerize the polysaccharide into its base monomer, glucose. In animals, this process typically relies on a symbiotic relationship between the host and its microbiome. Cellulolytic gut microbiota handle the digestion of lignocellulose in omnivores like humans (Robert et al., 2007). However, cellulose is more vital to the diet of herbivorous ruminants such as cows, who consume up to 2% of their bodyweight (or 14 kg on average) in grass per day. Ruminants sustain this massive volume of grazing thanks to 200 species of cellulolytic bacteria in their foregut alone, amounting to 10^{10}-10^{11} cells/mL (Matthews et al., 2019). Cellulolytic microbes are also prominent in the guts of termites, further establishing the ubiquity of this type of mutualism (Peristiwati et al., 2018). However, guts are far from the only source of cellulolytic microbes.

Beyond the microbiome, cellulolytic microbes are also commonly found as free-living organisms. One example is the soil-dwelling fungal genus *Trichoderma,* which has been shown to possess competitive saccharolytic ability compared to another cellulolytic fungus, *Aspergillus niger* (Tiwari et al., 2013; Do Vale et al., 2014). Cellulolytic bacteria are found in various environments. For example, *Thermobifida cellulolytica,* which thrives in heated manure heaps, and *Geobacillus thermoleovorans*, which was isolated from sugar refinery wastewater (Kukolya et al., 2002; Tai et al., 2004). Cellulolytic bacteria belonging to the genera *Brevibacillus, Paenibacillus, Bacillus,* and *Geobacillus* have even been grown using soil samples collected from the deep sub-surface of a U.S. gold mine (Rastogi et al., 2009). Despite the recalcitrance of

cellulose, cellulolytic microbes dramatically accelerate the process of its degradation by catalyzing its hydrolysis through the use of enzymes called cellulases, a type of glycoside hydrolase.

1.3 Glycoside Hydrolases

Enzymes that specialize in depolymerizing polysaccharides are known as glycoside hydrolases (GHs) (Henrissat, 1991). GHs are a type of carbohydrate-active enzyme (CAZyme), which is a broader category that encompasses enzymes involved in the synthesis and degradation of carbohydrates and glycoconjugates (Bhattacharya et al., 2015). There are four other types of CAZymes (or CAZyme domains): glycosyltransferases, polysaccharide lyases, carbohydrate esterases, and carbohydrate-binding modules. However, none drive the breakdown of cellulose and other polysaccharides as much as GHs (Kumar et al., 2019a). GHs function by catalyzing hydrolysis of the glycosidic linkages in polysaccharide chains, thus degrading them down into their base saccharides. GHs are currently phylogenetically classified into 180 GH families, which often share similar activities (Henrissat, 1991). Because of the large size of polysaccharides, they are typically degraded in the extracellular space. To exit the cell, some GHs possess an N-terminus signal peptide that mediates their secretion into the extracellular space (Zhang et al., 2021). Other GHs may belong to cell-surface enzyme complexes called cellulosomes, which facilitate attachment to insoluble substrates like lignocellulose (Payne et al., 2013). Finally, saccharolytic fungi may attach individual GHs to the plasma membrane through post-translational glycosylphosphatidylinositol anchoring (Nakajima et al., 2012). Overall, GHs are equipped for efficient depolymerization of polysaccharides like lignocellulose. This quality has made them the subject of industrial interest for their potential in developing renewable energy technologies.

One modern interest for GH has surrounded their application in producing biofuels using agricultural waste. Here, applicable GHs are used to degrade lignocellulosic biomass leftover from

common crops, such as corn stover (stalks, leaves, and cobs) and sugarcane bagasse, into simple sugars (Zakir et al., 2016). The monosaccharides are fermented by organisms like *Saccharomyces cerevisiae*, producing bioethanol as a byproduct (Cunha et al., 2020). The engineering of known cellulolytic microbes to degrade lignocellulose more efficiently is one frontier in biofuel research, but another focuses on the characterization of novel microbes and GHs to uncover new, potentially better candidates for lignocellulose conversion. In 2011, Hess et al. (2011) had this goal in mind when they set out to expand the known repertoire of cellulolytic microbes, recovering 27,755 putative CAZyme genes from 268 gigabases of microbial DNA from largely unculturable cow rumen microbes attached to plant fiber. They expressed 90 of these proteins, confirmed enzymatic activity in 57% of them, and assembled 15 metagenome-assembled genomes from uncultured microorganisms for enzymologists to pull CAZymes from. Overall, this study represented a major leap for large-scale characterization of mesophilic CAZymes for biofuel production. Thermophilic GHs, however, have advantages over their mesophilic counterparts that have garnered the interest of the biofuels research community.

1.4 Thermophilic Polysaccharide Catalysis

When compared to mesophilic GHs, the higher reaction temperatures of thermophilic GHs allow for greater reaction velocity, minimized risk of contamination, reduced substrate viscosity, auto-distillation of bioethanol, longer shelf lives, and heat-purification (Peng et al., 2016). There are a few examples of well-studied microbes degrading lignocellulose at high temperatures, ranging from moderate thermophiles to hyperthermophiles. One is the aerobic, moderately thermophilic organism *Thermobifida fusca* (optimal growth temperature = 55 °C)*,* a soil bacterium prominent in compost heaps, rotting hay, and manure piles (Lykidis et al., 2007). Some cases of strictly fermenting cellulolytic bacteria thrive at higher temperatures. These bacteria include

Clostridium thermocellum (60 °C), which ferments microcrystalline cellulose, pretreated switchgrass, and pretreated corn stover; *Caldicellulosiruptor bescii* (75 °C), which ferments crystalline cellulose, xylan, and plant biomass; and *Thermotoga maritima* (80 °C), which ferments starch and glycogen (Yoav et al., 2017; Huber at al., 1986; Yang et al., 2010). Aerobic polysaccharide use under hyperthermophilic conditions is far less common, with only one known exception.

The first traces of a lineage originally referred to as OctSpA1-106 came as DNA extracted and amplified from sediment collected from Octopus Spring at Yellowstone National Park, USA (Blank et al., 2002). OctSpA1-106 was originally thought to belong to a yet-unstudied group that represented the oldest lineage in the bacterial line of descent. A 2013 study of "microbial dark matter" recovered single-amplified genomes representing the same lineage from Great Boiling Spring (GBS) sediments at Gerlach, NV (Rinke et al., 2013). They proposed a candidate phylum, *Fervidibacteria* (later proposed to be a novel class nested within the phylum *Armatimonadota* [Nou, 2022]), and a novel bacterial species, *Fervidibacter facchari*, named for its thermophily, rod morphology, and large cache of genes encoding GH domains.

Nancy O. Nou of the Hedlund Lab isolated *F. facchari* by enriching it in culture to near purity and then cultivating a clonal population derived from a single cell that was isolated manually using optical tweezers (Nou, 2022), resulting in the first axenic culture from class *Fervidibacteria*. Nou successfully predicted possible polysaccharide substrates based on genome annotations and experimentally confirmed the use of 16 polysaccharides by *F. facchari* as sole carbon and energy sources (Nou, 2022). Consistent with its polysaccharide use, the genome of the isolated strain, PD1, encodes 115 genes with GH domains from 46 GH families. Differential transcriptomics and proteomics experiments revealed that 99 of the GHs were expressed, and half were differentially

expressed when grown in media containing five different growth substrates: β-glucan, gellan gum, locust bean gum, starch, and xyloglucan. With these findings, Nou's work paved the way for a deeper study of the GHs from *F. sacchari.*

The Hedlund Lab has since set out to characterize *F. sacchari* GHs and has experimentally confirmed activities of four enzymes: a GH family 3 xylanase, a GH5 xylanase, a GH10 glucanase, and a GH50 called Fsa1572, which is the focus of the research summarized in this thesis. The gene encoding Fsa1572 was codon-optimized for expression in *Escherichia coli*, ligated into a pET21b-GB1 vector, and transformed into *E. coli* T7 Express cells. Fsa1572 was heterologously expressed and purified by heat and subjected to a series of experiments to determine its substrate specificity, bond activity, condition optima, thermostability, kinetic parameters, and phylogenetic placement among family GH50. Fsa1572 was anomalous within GH50, possessing the first known β-1,4-glucanase activity in the family and by far the highest optimal temperature (80 °C). Phylogenetic analysis of Fsa1572 revealed a distant placement relative to other characterized GH50s. With these data as evidence, we propose Fsa1572 represents the first newly described GH50 subfamily, GH_2.

CHAPTER 2: BIOCHEMICAL CHARACTERIZATION OF FSA1572 FROM

FERVIDIBACTER SACCHARI, THE FIRST HYPERTHERMOPHILIC GH50 WITH β-1,4-

GLUCANASE ACTIVITY

Jonathan K. Covington[1], Nicole Torosian[1], Allison M. Cook[1], Scott G. Bryan[1], Nancy O. Nou[1], Marike Palmer[1], Jan-Fang Cheng[2], Miranda Harmon-Smith[2], Ritesh Mewalal[2], Brian P. Hedlund[1,3*]

[1]School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, NV, USA

[2]Department of Energy Joint Genome Institute, Walnut Creek, CA, USA

[3]Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, Las Vegas, NV, USA

2.1 Abstract

The bacterium *Fervidibacter sacchari* is an aerobic hyperthermophile that catabolizes various polysaccharides and is the only cultivated member of the class *Fervidibacteria* within the phylum *Armatimonadota*. Among its glycoside hydrolase (GH) cache is an enzyme from GH family 50 (GH50), an understudied family with only 25 characterized representatives and two known activities from 1,518 predicted members in the Carbohydrate-Active EnZyme (CAZy) Database. Here, we expressed, purified, and functionally characterized an extracellular GH50 from *F*. sacchari called Fsa1572. Using colorimetric assays, we show it has novel β-1,4-glucanase activity and only weak agarose activity that is typical for GH50 enzymes. The purified enzyme has a wide temperature range of 60-95 °C (optimal 80 °C), making it the first characterized hyperthermophilic representative of GH50. The enzyme also has a broad pH range of at least 5.5-11 (optimal 6.5-10). Fsa1572 possesses K_M and k_{cat} parameters of 12.6 mM and 4.62 s^{-1}, respectively. Finally, a phylogenetic analysis of Fsa1572 and other GH50 enzymes revealed a

unique phylogenetic position for Fsa1572 that is distant from other characterized enzymes and related to yet-uncharacterized GH50s from genomes of thermophilic archaea. Fsa1572 is the first characterized GH enzyme of *F. sacchari* and is both functionally and phylogenetically novel.

2.2 Introduction

Thermoactive enzymes are an essential component of hyperthermophiles, which thrive in geothermal environments at temperatures between 80 °C and 106 °C (Kelly et al., 2001). These environments, which include marine hydrothermal vents, geysers, and hot springs, host bacteria and archaea that use recalcitrant substrates, including polysaccharides, as carbon and energy substrates (Blumer-Schuette et al., 2008; Kelly et al., 2001; Strazzulli et al., 2017). A variety of hyperthermophilic bacteria such as *Thermotoga maritima* and *Caldicellulosiruptor bescii* have been shown to ferment a wide variety of polysaccharides, including lignocellulose, hemicellulose, and glucans, such as β-glucan (Bhalla et al., 2013; Chen et al., 2013; Crosby et al., 2022; Meng et al., 2016).

β-glucans are a type of polysaccharide found in the cell walls of cereal grains and certain bacteria and yeast, and are composed of glucose residues connected by glycosidic linkages (Wang and Fisher, 2015). The structure and water solubility of β-glucan depends on its origin; for example, β-glucans from plants are branched, relatively soluble, and have alternating β-1,3/1,4 linkages, while those from bacteria are unbranched, insoluble, and have β-1,3 linkages (Mudgil, 2017; Paudel et al., 2021). Although the specific sources of β-glucans in hot springs have not been investigated, we can better understand how certain hyperthermophilic organisms use β-glucans by studying the enzymes responsible for breaking them down.

Glycoside hydrolases (GHs) are a class of enzymes that hydrolyze the glycosidic linkages between sugar residues in carbohydrates, breaking them down into oligosaccharides, or even their

base monomers (Kelly et al., 2001). GHs that degrade β-glucans do so by hydrolyzing their β-1,3- or β-1,4 glycosidic bonds, depending on which are present in the given β-glucan. These linkages can be cleaved linearly by either exo- or endo-acting enzymes (Santos et al., 2020). The activity of a GH can be predicted to some degree from its amino acid sequence through conserved active sites and catalytic residues, or by phylogenetic analyses of GH sequences, which enable the categorization of GHs into GH families (Summers et al., 2016). GH families are groups of phylogenetically related GHs that typically share one or more identifying features, such as a conserved catalytic residue or mechanism (Henrissat and Bairoch, 1993). GHs can be classified even more broadly into GH clans, comprising GH families that share structural characteristics, and more finely into subfamilies, where their representatives are clustered within the broader family phylogeny (Henrissat and Bairoch, 1996; Viborg et al., 2019). GH families and subfamilies often share the same catalytic functions and degrade the same substrates.

One example of a catalytically homogenous GH family is GH family 50 (GH50), of which nearly all members to date have been classified as β-agarases that hydrolyze the β-1,4-glycosidic linkages found in agarose to release its degradation products: neoagarohexaose, neoagarotetraose, and neoagarobiose. There are two exceptions to this trend: a β-galactosidase from *Victivallus vadensis*, and a β-1,3-glucanase from *Pseudomonas aeruginosa* (Temuujin et al., 2012a; Yi et al., 2018). The earliest studied GH50 was AgaA from *Vibrio* sp. JT0107, which was described as the first endo-β-agarase capable of hydrolyzing both agarose and neoagarotetraose (Sugano et al., 1993). To date, four structures are known from GH50, one of which is characterized as monomers, two as homodimers, and one as a homotetramer (Giles et al., 2017; Pluvinage et al., 2013; Pluvinage et al., 2020; Zhang et al., 2019b). These structures have C-termini with $(\beta/\alpha)_8$-barrel folds at their C-terminal ends, which are characteristic of the family's clan, GH clan A (GH-A).

Although the catalytic residues of GH50 have not been proven, its designation as belonging to GH-A infers the use of two conserved glutamic acid residues as the acid/base catalyst and nucleophile (Kumar et al., 2019b). Current research surrounding GH50s has focused on their potential in the production of neoagarobiose, which is used as a skin moisturizer and melanoma whitener (Lee at al., 2006). However, GH50 is an under-studied family: of the 1,518 GH50s available from the Carbohydrate-Active EnZyme (CAZy) Database (www.cazy.org), only 25 have been functionally characterized (Drula et al., 2022). The β-galactosidase (VadG925) and β-1,3-glucanase (PaBglu50A) set a precedent that GH50 may be more catalytically diverse than its currently known repertoire suggests. The study of GH50s from novel or understudied microbes can broaden our understanding of the true diversity of the family.

We recently isolated *Fervidibacter sacchari*, an aerobic hyperthermophile (T_{opt} 80 °C) belonging to the phylum *Armatimonadota*. *F. sacchari* uses at least 16 diverse polysaccharides as sole carbon sources and electron donors, including glucans, hemicelluloses, and plant biomass (Nou, 2022). The genome of *F. sacchari* has 115 genes encoding GH domains across 46 GH families, including two that were annotated as members of GH50, designated Fsa1572 and Fsa2534. Since the optimal growth temperature of *F. sacchari* exceeds the optimal temperature of the most thermophilic GH50 enzyme characterized to date, AgrA from *Agarivorans* sp. AG17 (T_{opt} 65 °C) (Nikapitiya et al., 2010), we characterized Fsa1572 to determine whether it is an unusually thermophilic member of GH50 and compare its properties to other GH50 enzymes. To do this, we codon-optimized the gene encoding Fsa1572, ligated it into pET21b-GB1, and expressed it in *Escherichia coli*. The enzyme was then purified and used to determine its substrates, optimal conditions, kinetic parameters, and biochemical characteristics. Finally, we examined the evolutionary context of Fsa1572 within the GH50 family by conducting a comprehensive

phylogenetic analysis of GH50 enzymes. We show that Fsa1572 is indeed hyperthermophilic (T_{opt} 80 °C) and the first GH50 with β-1,4-glucanase activity. Because of this unique activity, and its unique phylogenetic position distant from previously characterized GH50s, Fsa1572 potentially represents a novel GH50 subfamily.

2.3 Materials and Methods

2.3.1 Gene Synthesis, Heterologous Expression, Enzyme Purification, and Structural Prediction

The amino acid sequence of Fsa1572 was initially submitted to the dbCAN2 meta server and dbCAN HMMdb v11.0 to determine the GH family of Fsa1572 (Zhang et al., 2018). The amino acid sequence was submitted for annotation by SignalP (Teufel et al., 2022) to determine the presence of a signal sequence. A MAFFT-DASH (Katoh et al., 2019) multiple sequence alignment of seven GH50 amino acid sequences (Fsa1572, Fsa2534, Aga50D, AgaA, AgrA, PaBglu50A, and VadG925) was generated and visualized in Jalview v2.11.2.6 (Waterhouse et al., 2009) to confirm the presence of two conserved glutamic acid residues. The gene encoding Fsa1572 was codon optimized with a balanced approach, ligated into the pET21b vector (Novagen, Madison, WI) encoding a polyhistidine-tag, a GB1 solubility tag, and a TEV protease cleavage site at the protein N-terminus, and transformed into *E. coli* T7 Express cells (New England Biolabs, Ipswich, MA).

E. coli cells containing the Fsa1572 gene were grown on Luria-Bertani (LB) agar plates (10 g tryptone, 5 g yeast extract, 5 g NaCl per L) containing 100 µg/mL ampicillin. For expression, LB broth with 100 µg/mL ampicillin was incubated at 37 °C with shaking at 225 rpm until reaching an OD_{600} of 0.6-0.8. Isopropylthio-β-galactoside (IPTG) was added to a concentration of 0.5 mM and incubated overnight at 37 °C with shaking at 150 rpm. The cells were centrifuged at 16,100 x g, resuspended in lysis buffer (50 mM Tris-HCl, 100 mM NaCl, pH 7.0), and lysed by sonication

11

for 40 seconds with 40% cycle and level 4 output. The crude lysate was clarified by centrifugation at 16,100 x g to isolate the soluble fraction.

Fsa1572 was purified by heating the soluble fraction to 80 °C in a heat block for 30 minutes, followed by centrifugation at 16,100 x g to remove denatured host proteins in the pellet. Fsa1572 expression and purity were assessed using 7.5% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The molecular weight was estimated using a BriteRuler Pre-stained Protein Ladder (Abcam, Cambridge, UK). The ability of Fsa1572 to oligomerize was assessed using native PAGE alongside a PageRuler Plus Pre-stained Protein Ladder (Thermo Scientific, Vilnius, Lithuania). Protein concentration was assessed by a Pierce™ BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA).

The structure of Fsa1572 was predicted using AlphaFold 2 via the ColabFold platform integrated with ChimeraX v 1.5 using default settings (Pettersen et al., 2021; Mirdita et al., 2022). Four Fsa1572 structure predictions were generated: native, mature Fsa1572 (signal sequence removed); non-native Fsa1572 containing an N-terminal 6x-His tag, GB1 solubility tag, and a TEV protease site; homodimeric, native, mature Fsa1572, and homotrimeric, native, mature Fsa1572. The resultant structures were visualized and edited in ChimeraX v 1.5, and the predicted aligned error (PAE) plots of the multimeric predictions were analyzed to gain insight into the potential of Fsa1572 to homooligomerize.

2.3.2 Enzyme Activity Assays

To determine the enzyme activity of Fsa1572, the following potential substrates were tested initially: chondroitin sulfate (Alfa Aesar), colloidal chitin as prepared according to Hsu and Lockwood (1975) (Beantown Chemicals), brown algae fucoidan (BestVite), lupin galactan (Megazyme), gellan gum (Serva), karaya gum (Sigma), locust bean gum (Spectrum), xanthan gum

(Sigma), birch wood xylan (Sigma), tamarind xyloglucan (CarboMer), oat β-glucan (Megazyme), oyster glycogen (TCI), potato starch (J.T. Baker), and ammonia fiber expansion (AFEX)-pretreated corn stover, *Miscanthus*, and sugarcane bagasse (DuPont). These 16 substrates were chosen based on their use by *F. sacchari* as sole carbon sources and electron donors. Agarose was included in the screen to test for β-agarase activity that would be consistent with GH50. The substrates (0.5% w/v) and enzyme (67.2 µg/mL) or empty vector control were mixed 1:1 in microplate wells (final volume 40 µL) then sealed and incubated overnight at 80 °C with a beaker of water to prevent evaporation. To measure the quantity (µmol) of reducing sugars released by the reactions, 3,5-dinitrosalicylic acid (DNS) solution (0.25 g DNS, 75 g sodium potassium tartrate, 50 mL 2 M sodium hydroxide solution, brought to 250 mL using ultrapure water) was added 4:1 to each reaction mixture (final volume 200 µL), and the plate was wrapped in foil and incubated at 100 °C for 20 minutes (Kim et al., 2014). Absorbance was measured using a SpectraMax Plus Platereader at 570 nm. A second screen with robust quantitation was conducted with Fsa1572 (67.2 µg/mL) in triplicate and compared to the empty vector control using an unpaired t-test. A DNS standard curve was generated using glucose at 0, 5, 10, 15, 20, and 25 mM.

To determine how much of the β-glucan substrate was depolymerized by Fsa1572, an acid-hydrolyzed β-glucan control was prepared by heating a 1% solution of β-glucan in 0.5 M H_2SO_4 sulfuric acid to 121 °C for 30 minutes followed by neutralization with 1 M NaOH. The acid-hydrolyzed control was incubated with water in place of the enzyme alongside Fsa1572 mixed with non-hydrolyzed β-glucan as before. The amount of reducing sugars released from exhaustive overnight degradation by Fsa1572 was compared to the reducing sugar content of the acid-hydrolyzed control.

Specific bond cleavage by Fsa1572 was assessed using the *para*-nitrophenyl- (pNP) linked substrate pNP-β-D-glucopyranoside (pNPG) (Megazyme). pNPG (20 µg/mL) was mixed with Fsa1572 (67.2 µg/mL), incubated for 30 minutes at 80 °C, and the reaction was terminated with 2% w/v disodium phosphate solution (pH 12.0). Bond cleavage was assessed as an increase in absorbance at 400 nm, indicating substrate hydrolysis. A standard curve was generated using a range of *para*-nitrophenol concentrations (0, 6, 12, 18, 24, and 30 µg/mL).

2.3.3 Biochemical Characterization and Kinetic Parameters

To determine the temperature range and optimum of Fsa1572, the enzyme was incubated with 0.5% w/v β-glucan at a range of temperatures (4, 20, 30, 50, 60, 70, 80, 90, and 95 °C). The resulting solution was analyzed as described before using DNS solution. To determine the thermostability of the enzyme, Fsa1572 was incubated for one hour at either 70, 80, 90, or 100 °C, then incubated at 80 °C with β-glucan and tested using the DNS assay. To determine the pH range and optimum of Fsa1572, the enzyme was expressed and purified in 50 mM buffers at different pHs: 2-morpholin-4-ylethanesulfonic acid (MES) (5.5-6.5), 2-amino-2-(hydroxymethyl)propane-1,3-diol (Tris) (7.0-9.0), 2-(cyclohexylamino)ethane-1-sulfonic acid (CHES) (9.5-10.0), and 3-(cyclohexylamino)propane-1-sulfonic acid (CAPS) (10.5-11.0), all with 100 mM NaCl. β-glucan was dissolved in the same buffers to capture the whole pH range. Assays were conducted with 0.5% w/v β-glucan in triplicate, then analyzed as previously described using DNS method. The temperature and pH ranges and optima and thermostability were determined using one-way Analysis of Variance (ANOVA) with post-hoc Tukey's Honestly Significant Difference (HSD) tests ($P \leq 0.05$).

To determine the Michaelis-Menten kinetic parameters and k_{cat} of Fsa1572, pNPG was mixed with purified Fsa1572 (67.2 µg/mL) at a range of concentrations (0.625, 1.25, 2.5, 5, 10,

and 20 µg/mL) and incubated for 30 minutes. The reactions were terminated as before and the absorbance at 400 nm was measured using a plate reader. A Lineweaver-Burk plot was generated from these data to calculate V_{max}, K_M, and k_{cat} (Ochs, 2022).

2.3.4 Phylogenetic Analysis of GH50 Within Family GH50

To perform a phylogenetic analysis of all GH50s, all sequences belonging to GH50 in the CAZy Database (Drula et al., 2022) were extracted from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) using the Batch Entrez site (https://www.ncbi.nlm.nih.gov/sites/batchentrez). BioEdit v. 7.2.5 (Hall, 1999) was used for visualizing and editing all data matrices. After the removal of all duplicate sequences, protein sequences of less than 50 amino acids were also removed and Fsa1572 along with all other GH50 homologs from *Fervidibacteria* (133 in total) were added to the dataset, resulting in 1,638 protein sequences. A multiple sequence alignment was generated using the MAFFT online server v. 7 (Katoh et al., 2019) with the MAFFT FFT-NS-2 progressive alignment algorithm (Rozewicki et al., 2019). A smaller alignment was similarly generated to more easily visualize the relationship between Fsa1572 and its closest relatives. For this smaller phylogeny, the MAFFT-DASH structural alignment algorithm (Katoh et al., 2019) was used instead to bring structural information into account. The larger alignment was then subjected to an approximate Maximum-Likelihood analysis with FastTree v. 2.1.11 (Price et al., 2010), using Shimodaira-Hasegawa approximate Likelihood Ratio Test (SH-aLRT) branch support values (Guindon et al., 2010) from 1,000 replicates, while the smaller dataset was also subjected to a true Maximum-Likelihood analysis with IQTree v. 2.2.0 (Minh et al., 2020) using SH-aLRT branch support values from 1,000 replicates and Ultrafast bootstrapping (UFBoot) branch support from 1,000 replicates. The

obtained phylogenies were visualized using the Interactive Tree of Life v. 6.7 website (https://itol.embl.de) and edited in Inkscape v. 0.92.4 (https://www.inkscape.org).

2.4 Results

2.4.1 Structural Prediction, Heterologous Expression, and Substrate Screening

Following annotation by HMMER (E value = 3.90×10^{-102}; cutoff = 1×10^{-15}) and dbCAN-sub (E value = 5.00×10^{-134}; cutoff = 1×10^{-15}), Fsa1572 was determined to belong to GH50 with no carbohydrate-binding modules or other annotated domains. Fsa1572 had an N-terminal signal sequence (MSWTRREFVKLVGFATTFAGASCRSEG) that was predicted by SignalP to belong to the twin-arginine translocation (TAT) pathway, which suggests that Fsa1572 is secreted to the extracellular space as a folded protein (Ren et al., 2016). A multiple sequence alignment of seven GH50 amino acid sequences (Fsa1572, Fsa2534, Aga50D, AgaA, AgrA, PaBglu50A, and VadG925) revealed that Fsa1572 has two conserved GH-A residues at Glu192 and Glu354, and Fsa2534 has the conserved residues at Glu409 and Glu575 (Figure S1). After expression and purification by heat alone, the purity of Fsa1572 was estimated by SDS-PAGE to be >90% and the molecular weight was estimated at 60 kDa (Figure 1). The size of the recombinant protein was affected by the removal of the TAT signal sequence (3 kDa), and the addition of a GB1 solubility tag (6 kDa), a polyhistidine tag (1 kDa), and a TEV protease site (1 kDa) (Figure S2). However, the recombinant Fsa1572 was still slightly larger than the predicted native size of 52 kDa. The resolved native PAGE gel revealed three bands at greater than 130 kDa, suggesting the formation of multimeric structures.
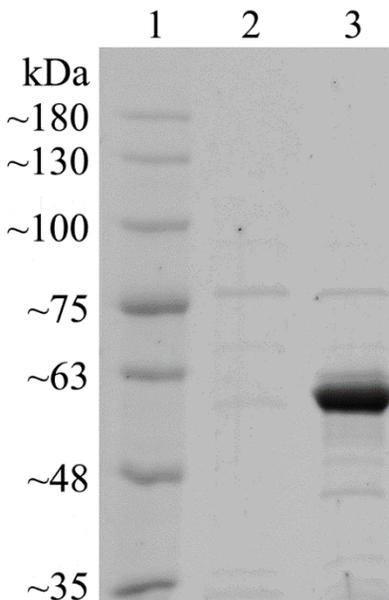
Figure 1: SDS-PAGE of purified Fsa1572. Proteins were resolved on a 7.5% polyacrylamide SDS-PAGE gel and subsequently stained using Coomassie Brilliant Blue. Lane 1: BriteRuler Pre-stained Protein Ladder (Abcam, Cambridge, UK); 2: lysate from empty vector control following heat-purification procedure; 3: heat-purified non-native Fsa1572 containing N-terminal 6x-His tag, GB1 solubility tag, and TEV protease site.

Structural prediction of native, mature Fsa1572 produced a structure with low PAE scores (0-5) indicative of high confidence in the accuracy of the prediction (Figure S3A, S3B). Two conserved glutamic acid residues that are proposed as catalytic residues of GH-A enzymes are found in a tunnel through the structure's center, marking this as the candidate active site (Figure S3C) (Kumar et al., 2019b). Furthermore, the residues surrounding the active site tunnel had relatively higher PAE scores (2.5-5) suggesting lesser rigidity, supporting this tunnel as the active site. The predicted structure of non-native Fsa1572 containing an N-terminal 6x-His tag, GB1 solubility tag, and a TEV protease site (Figure S3D) had low PAE scores (0-5) in the catalytic domain, indicative of normal predicted folding, and was the form of Fsa1572 used for biochemical characterization. Predictions of the homodimeric and homotrimeric forms of Fsa1572 revealed

asymmetric multimers (Figure S3E). However, the PAE plot of the homodimer shows low PAE

scores within each monomer (0-5), and low-to-moderate PAE scores between monomers (5-15),

while the PAE plot of the homotrimer shows high PAE scores between monomers (25-30),

suggesting low confidence in a homotrimer structure and moderate confidence in a homodimer

structure of Fsa1572 (Figure S3F). The predicted structure of Fsa1572 was a close match to three

uncharacterized enzymes in the AlphaFold Protein Structure Database (Jumper et al., 2021; Varadi

et al., 2022): two predicted β-agarases from the bacterial phylum *Armatimonadota* (A0A7C3N829

[UniProt accession number], E-value = 0, percent identity = 92.68% via the National Center for

Biotechnology Information Basic Local Alignment Search Tool [NCBI BLAST]; and

A0A7C2VJI7, E-value = 0, percent identity = 78.17%) (Zhou et al., 2020), and a predicted β-

agarase from bacterium HR17 (A0A2H5XB18, E-value = 0, percent identity = 77.90%) (Kato et

al., 2018). All three enzymes possess very high homology with at least one uncharacterized GH50

from *Fervidibacteria* (E-values = 0, percent identities >95%). The experimentally validated GH50

structures that were most similar to Fsa1572 were Aga50D, a homotetrameric β-agarase from

Saccharophagus degradans 2-40 (E value = 6 x 10^{-54}, percent identity = 33.42%) (Kim et al., 2010;

Pluvinage et al., 2013), and AgWH50C, a homodimeric β-agarase from *Agarivorans gilvus*

WH0801 (E value = 3 x 10^{-53}, percent identity = 33.24%) (Liu et al., 2014b; Zhang et al., 2019b).

Aga50D homotetramers possess similar tunnel-shaped active sites on each molecule that contain

conserved GH-A catalytic residues Glu695 and Glu534, which align with Glu354 and Glu192 from

Fsa1572, respectively. Additionally, Fsa1572 possesses other GH50 active site residues present in

both Aga50D and AgWH50C: Glu417, Arg413, Asn191, and Phe403 in Fsa1572; Glu757,

Arg752, Asn533, and Phe742 in Aga50D; Glu705, Arg700, Asn483, and Phe690 in AgWH50C.

However, Fsa1572 does not have an active site tryptophan residue aligning with Trp199 and

Trip150 in the CBM domains of Aga50D and AgWH50C, respectively, which is consistent with the absence of a CBM domain in Fsa1572. Overall, these shared active site residues provide additional support for the active site tunnel of Fsa1572. Although structural prediction supported Fsa1572 asymmetric homodimerization, its greater similarity to the asymmetric homotetramer Aga50D compared to the asymmetric homodimer AgWH50C and a cyclically symmetric GH50 homodimer called PfGH50B (E-value = 3 x 10^{-51}, percent identity = 28.94%) (Giles et al., 2017) suggests Fsa1572 may be more likely to tetramerize. The catalytic mechanisms of Aga50D and AgWH50C were described as retaining and putatively retaining, respectively, based upon the role of the conserved Aga50D active site glutamic acid residues as a nucleophile (Glu695) and proton donor (Glu534), which is consistent with the retaining mechanism of GH-A enzymes (Kumar et al., 2019b). Thus, we putatively identify the catalytic mechanism of Fsa1572 to be retaining, wherein Glu354 acts as a nucleophile and Glu192 serves as a proton donor.

After screening with 17 potential substrates using the DNS method, Fsa1572 was active on agarose and β-glucan. These activities were confirmed by repeating the experiment in triplicate, which revealed much stronger activity on β-glucan (Figure 2). Complete acid hydrolysis of the same amount of β-glucan yielded 1.303 µmol of reducing sugars compared to 0.767 µmol produced by Fsa1572. Consequently, Fsa1572 hydrolyzed 58.9% of the β-glucan polysaccharide overnight at 80 °C. Fsa1572 released 0.3680 µmol of pNP from pNPG after 30 minutes, confirming that it had β-1,4-glucanase activity (Figure S4). Fsa1572 was not active on chondroitin sulfate, colloidal chitin, brown algae fucoidan, lupin galactan, gellan gum, karaya gum, locust bean gum, xanthan gum, birch wood xylan, tamarind xyloglucan, oyster glycogen, potato starch, or the AFEX-pretreated substrates corn stover, *Miscanthus*, or sugarcane bagasse.
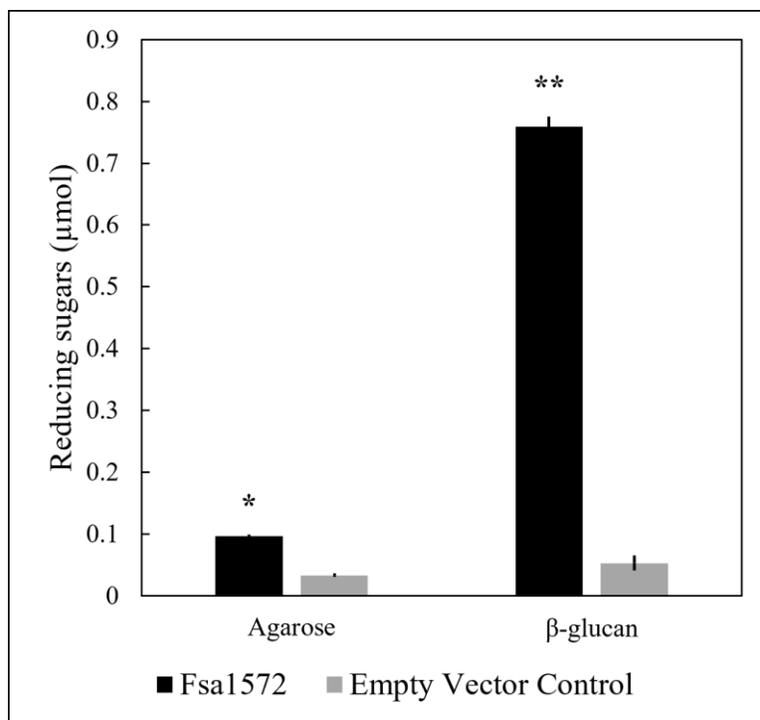
Figure 2: Substrate specificity. Fsa1572 is active on agarose and β-glucan compared to an empty vector control (* P ≤ 0.00005 and ** P ≤ 0.000005 via an unpaired t-test.) Error bars are based on standard error.

2.4.2 Biochemical Characterization and Kinetic Parameters

When tested with β-glucan at a range of temperature and pH values, Fsa1572 was a hyperthermophilic enzyme with a broad pH range. The temperature range of Fsa1572 activity was 50-95 °C with an optimum of ~80 °C (Figure 3). Fsa1572 was thermostable up to 90 °C, with similar amounts of reducing sugars released following a one-hour treatment at 70 °C (1.015 µmol), 80 °C (0.978 µmol), and 90 °C (0.922 µmol), while pre-incubation at 100 °C resulted in the complete loss of activity (Figure 4). Fsa1572 was active across the entire tested pH range of 5.5-11.0, and optimally active between 6.5-10.0 (Figure 5). Using a Lineweaver-Burk plot (R^2 = 0.9907) and assuming 100% enzyme activity, V_{max}, K_M, and k_{cat} were calculated as 357.1 µM/min,

12.6 mM, and 4.62 s^{-1}, respectively, when Fsa1572 was used at a concentration of 67.2 µg/mL (Figure S5).
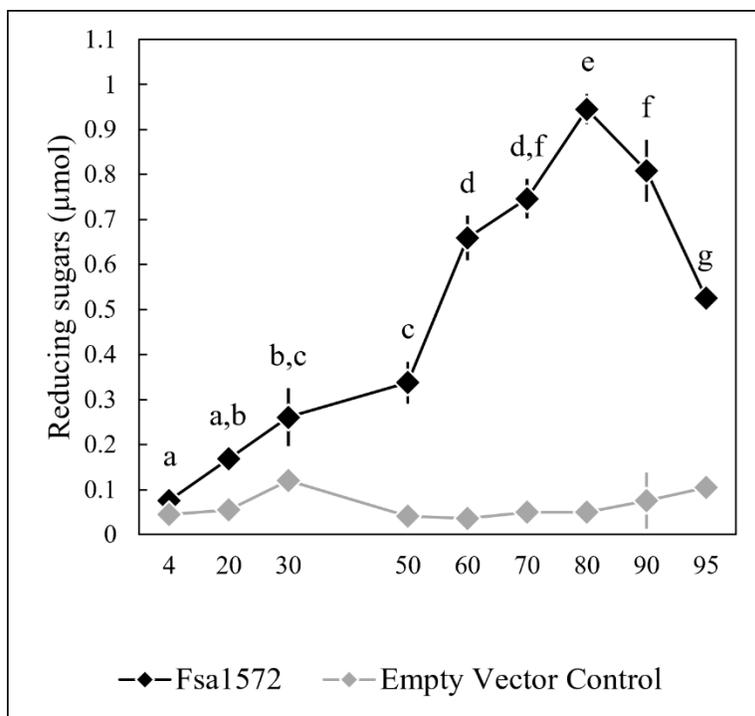
Figure 3: Temperature range and optimum of Fsa1572. Fsa1572 was most active on β-glucan between 60-95 °C, and optimally active at 80 °C. Temperatures with a shared letter are not significantly different (P ≤ 0.05 via a one-way ANOVA with post-hoc Tukey's HSD.) Error bars are based on standard error.
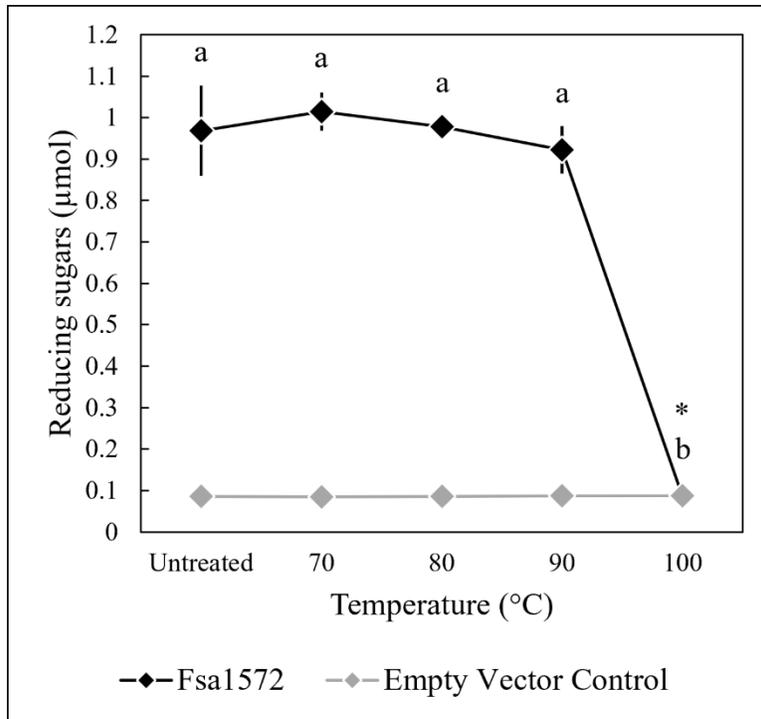
Figure 4: Thermostability of Fsa1572. Fsa1572 was stable up to 90 °C for one hour, while incubation at 100 °C for one hour resulted in the complete loss of activity on β-glucan. Temperatures with a shared letter are not significantly different (P ≤ 0.05 via a one-way ANOVA with post-hoc Tukey's HSD) and Fsa1572 treated at 100 °C for one hour was not significantly different from an empty vector control (P > 0.05 via an unpaired t-test). Error bars are based on standard error.
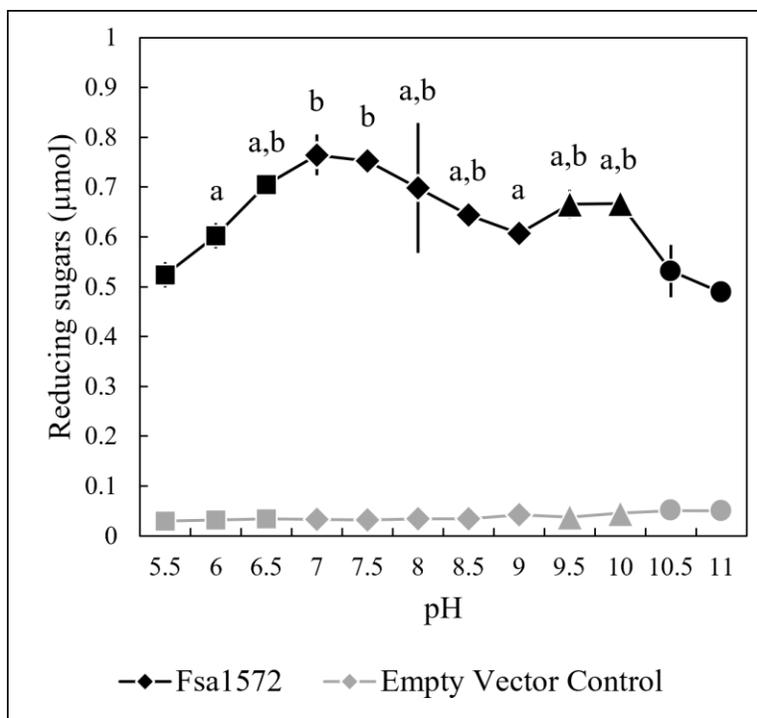
Figure 5: pH range and optimum of Fsa1572. Fsa1572 was active on β-glucan at all tested pH values, and optimally active between pH 6.5 and 10.0. pH values with a shared letter are not significantly different ($P \leq 0.05$ via a one-way ANOVA with post-hoc Tukey's HSD.) The buffers used were MES (■, 5.5-6.5), Tris (♦, 7.0-9.0), CHES (▲, 9.5-10.0), and CAPS (●, 10.5-11.0). Error bars are based on standard error.

2.4.3 Phylogenetic Analysis of Fsa1572 Within Family GH50

Based on approximate maximum-likelihood phylogenetic analyses of Fsa1572, other GH50 enzymes from class *Fervidibacteria*, and all members of family GH50 within the CAZy database, Fsa1572 was found to have a unique phylogenetic placement compared to other characterized GH50s (Figure 6A). *Fervidibacteria* GH50s were placed into five monophyletic clades throughout the phylogeny, containing 18, 19, 22, 30, and 45 sequences, with Fsa1572 belonging to the largest clade. The 25 previously characterized GH50s included in this analysis belonged to *Proteobacteria* (22 from *Agarivorans gilvus* WH0801, *Agarivorans* sp. AG17, *Agarivorans* sp. HZ105, *Agarivorans* sp. JA-1, *Agarivorans* sp. JAMB-A11, *Agarivorans* sp.

QM38, *Alteromonas* sp. E-1, *Paraglaciecola hydrolytica* S66, *Pseudoalteromonas* sp. NJ21, *P. aeruginosa*, *Saccharophagus degradans* 2-40, *Thalassotalea agarivorans* BCRC 17492, *Vibrio* sp. CN41, *Vibrio* sp. PO-303); *Verrucomicrobiota* (one from *V. vadensis* ATCC BAA-548); *Firmicutes* (one from *Paenibacillus agarexedens* BCRC 16000); and *Actinobacteriota* (one from *Streptomyces coelicolor* A3(2)). Most *Proteobacteria* sequences clustered into two groups of enzymes according to their lowest common taxonomic rank: β-agarases from *Enterobacterales* (from the genera *Agarivorans* and *Vibrio*) and β-agarases from a broader group of *Gammaproteobacteria* (from the orders *Enterobacterales* and *Pseudomonadales*; and genera *Alteromonas*, *Paraglaciecola*, *Pseudoalteromonas*, *Pseudomonas*, *Saccharophagus*, and *Thalassotalea*.) Consistent with their distinct activities from the other characterized enzymes, VadG925 (homolog 1 in Figure 6A and Table S1) and PaBglu50A (homolog 25 in Figure 6A and Table S1) were distant from the major characterized GH50 groups. The maximum-likelihood analysis of Fsa1572 and its closest relatives showed that the *Fervidibacteria* group that Fsa1572 belongs to was most closely related to six archaeal GH50s from two unrelated archaea, *Thermosphaera aggregans* in the Crenarchaeota/Thermoproteota and several species of *Thermococcus* in the *Euryarchaeota*/Methanobacteriota_B (Figure 6B) (phylum taxonomies are listed as in NCBI/GTDB.) Along with sequences from *Fervidibacteria*, these were the only GH50s in this analysis from thermophiles. The monophyly of *Fervidibacteria* homologs in this lineage suggests vertical evolution of the enzymes related to Fsa1572 in the class *Fervidibacteria*, with horizontal transfers to different thermophilic archaea, either independently, or horizontal transfer to *Thermosphaera aggregans*, and then from *Thermosphaera aggregans* to *Thermococcus*. The second *F. sacchari* GH50 enzyme, Fsa2534, belonged to a second major *Fervidibacteria* clade that was closely related to that containing Fsa1572, implying that the two paralogous GH50s of *F.*

sacchari represent distinct enzymes with potentially differing activities or regulation. Organisms encoding GH50s that were related to Fsa1572 included members of the *Proteobacteria* (75), *Cyanobacteria* (5), *Firmicutes* (4), and *Verrucomicrobiota* (3).

2.5 Discussion

Prior to this study, characterized GH50s have been almost exclusively classified as β-agarases, with the notable exceptions being a β-galactosidase and a β-1,3-glucanase (Temuujin et al., 2012a; Yi et al., 2018). Our work, however, reveals a GH50 with novel β-1,4-glucanase activity. Although Fsa1572 could degrade agarose, its hydrolysis of β-glucan released considerably more reducing sugars. Given that galactose and glucose have similar structures—both monosaccharides with six-membered rings that differ only in the orientation of the 4th carbon hydroxyl group—and the β-1,4 linkages present in both agarose and β-glucan, these two functions within a single GH family is feasible. So Fsa1572 is the first known GH50 with β-1,4-glucanase activity and the first known to degrade substrates with both galactose and glucose units. Fsa1572 furthermore deviates from other GH50s through its high temperature optimum of 80 °C, a marked increase over the second-highest temperature optimum in GH50: 65 °C of AgrA from *Agarivorans* sp. AG17 (Nikapitiya et al., 2010). Overall, Fsa1572 is a functionally distinct GH50 as the first known GH50 with β-1,4-glucanase activity and the first known hyperthermophilic enzyme within family GH50.

The kinetics of Fsa1572 fall short of other GH50s and thermophilic β-1,4-glucanases, with its relatively low k_{cat} of 4.62 s^{-1}. For example, GH50 β-agarase k_{cat} values range from 0.27 s^{-1} to 2200 s^{-1} (median = 298 s^{-1}) (Fu et al., 2008; Chen et al., 2018; Temuujin et al., 2012b) and thermophilic β-1,4-glucanase k_{cat} values have been reported between 104 s^{-1} and 1326.7 s^{-1} (median = 174.7 s^{-1}) (Akram and Haq, 2020; McCarthy et al., 2003). The low k_{cat} of Fsa1572 may be

explained in two ways. First, because *F. sacchari* occupies an unusual niche as an aerobic and hyperthermophilic polysaccharide degrader, it may have less competition for polysaccharides in oxic, high-temperature environments and thus relaxed selection for high velocity enzymes. High temperatures also contribute to substrate denaturation and degradation, further lessening the need for high velocity enzymes. Alternatively, the low velocity may be explained by pNPG being a poor substrate for Fsa1572. In this study, we tested Fsa1572 with 17 substrates that *F. sacchari* can use for growth and pNPG, an artificial substrate that contains a β-1,4-glycosidic linkage. Hydrolysis of pNPG by Fsa1572 released a lower quantity of *para*-nitrophenol (0.3680 µmol) than the quantity of reducing sugars (0.9689 µmol) produced by Fsa1572 following the same 30-minute incubation with β-glucan. At a concentration of 20 mM pNPG with 67.2 µg/mL Fsa1572, the enzyme was adequately saturated with 1.79×10^4 moles of pNPG for every mole of Fsa1572, indicating that the lower reaction rate with pNPG was not due to substrate limitation. Additionally, Fsa1572 does not completely hydrolyze oat β-glucan, evidenced by the release of only 58.9% of reducing sugars after overnight incubation compared to complete acid hydrolysis of β-glucan. As oat β-glucan possesses alternating β-1,3 and β-1,4 glycosidic linkages, each bond accounts for approximately half of one polymer. Therefore, incomplete hydrolysis of oat β-glucan following an exhaustive reaction suggests that only one type of glycosidic linkage within oat β-glucan is the substrate for Fsa1572. Overall, these patterns suggest Fsa1572 may also possess β-1,3-glucanase activity, and that hydrolysis of the β-1,3-glycosidic linkage in β-glucan may be more kinetically favored than hydrolysis of the β-1,4-glycosidic linkage. Alternatively, we speculate that unknown but related polysaccharides that are more relevant to the geothermal spring environment of *F. sacchari* may be better substrates for Fsa1572. Microbial mats are one source of exopolysaccharides, and extracts from photosynthetic mats from the fringes of Great Boiling

Spring were originally used to enrich for *F. sacchari* (Nou 2022). Further experiments are needed to determine if other substrates, such as microbially derived polysaccharides, are the preferred substrate of Fsa1572.

Fsa1572 has a unique phylogenetic position relative to other characterized GH50s. Aside from other GH50s from *Fervidibacteria*, Fsa1572 is most closely related to a subset of six GH50s encoded by several *Thermococcus* species and *Thermosphaera aggregans*. Given the presence of homologs of Fsa1572 in most available *Fervidibacteria* genomes, and the distant relationship between *Fervidibacteria*, *Thermococcus*, and *Thermosphaera* we infer that this enzyme passed through horizontal gene transfer from *Fervidibacteria* to these archaea, which are common co-inhabitants in geothermal environments. The lower activity of Fsa1572 on agarose coincides with substrate availability, as all agarolytic GH50s are from marine environments, where agarose is a major constituent of red seaweeds (Yun et al., 2021). The presence of weak agarase activity in Fsa1572 suggests that β-agarase and β-1,4-glucanase activities may have been ancestral in GH50, but divergence brought on by adaptation to polysaccharides in different environments selected for divergent activities. Further characterization of GH50s more closely related to Fsa1572 would offer insight into the validity of this hypothesis.

GH family annotation from amino acid sequences is limited by the breadth of known GH families. However, new GH families are often discovered, as recently as the discovery of GH174 by Liu et al (2023). Because of this, the functional and phylogenetic novelty of Fsa1572 prompted consideration that it could have been incorrectly annotated as a GH50 when it truly belonged to an as-yet discovered GH family. However, the presence of two conserved GH-A glutamic acid residues in Fsa1572 leads us to believe that Fsa1572 correctly falls within GH50 but should be distinguished from other characterized GH50s. Viborg et al. (2019) previously proposed a

roadmap for identifying GH subfamilies based on combinations of phylogenetic closeness and functional similarity. Because of its unprecedented β-1,4-glucanase activity, 80 °C optimum, and distant relationship to other characterized GH50s, we propose Fsa1572 belongs to the first distinct subfamily of GH50, for which we propose the name GH50_2. Likewise, we propose the GH50 clade containing the characterized *Enterobacterales* and *Gammaproteobacteria* β-agarases be designated as GH50_1.

2.6 Funding

2.7 Acknowledgments

2.8 Supplemental Figures and Tables

A

N-terminus [28] [Glu192] [Glu354] [454] C-terminus

Signal peptide   Glycoside Hydrolase 50 domain

B

N-terminus [5 12] [74] [81] [Glu245] [Glu407] [507] C-terminus

6x-His tag   TEV protease site   Glycoside Hydrolase 50 domain

GB1 solubility tag
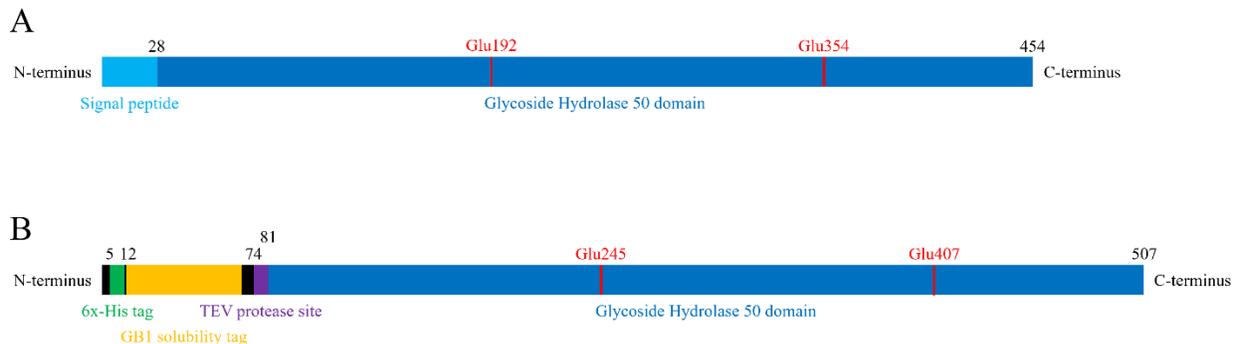
Figure S2: Native and non-native Fsa1572 schematic. (A) The native Fsa1572 protein containing an N-terminal signal peptide (light blue) and a glycoside hydrolase domain (blue) with conserved residues Glu192 and Glu354 (red). (B) The non-native Fsa1572 protein containing an N-terminal 6x-His tag (green), a GB1 solubility tag (yellow), a TEV protease site (purple), and short linkers between motifs (black), as well as a GH50 domain (blue) with conserved residues Glu245 and Glu407 (red).
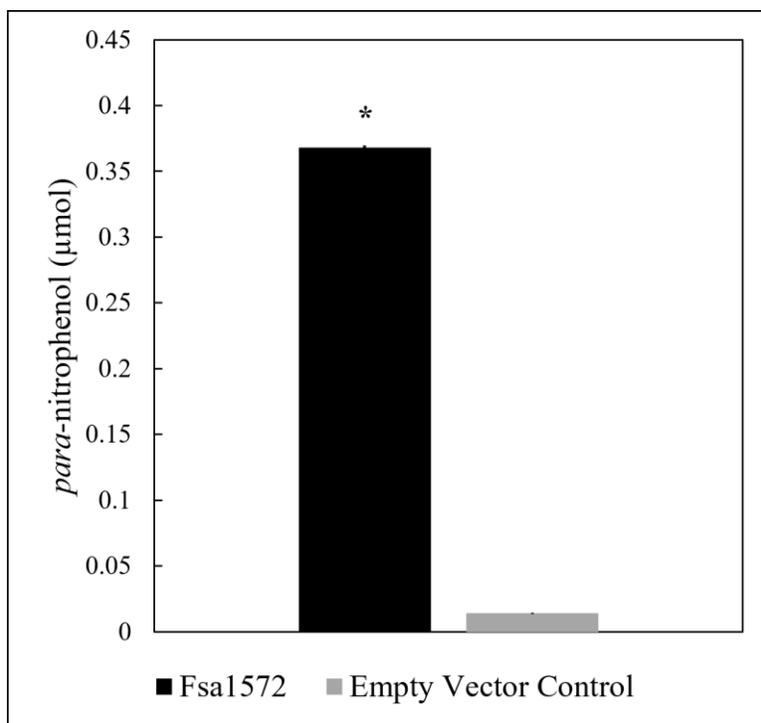
Figure S4: pNPG activity assay. Fsa1572 is active on pNPG compared to an empty vector control (* $P \leq 0.000005$ via an unpaired t-test.) Errors bars are based on standard error.

29
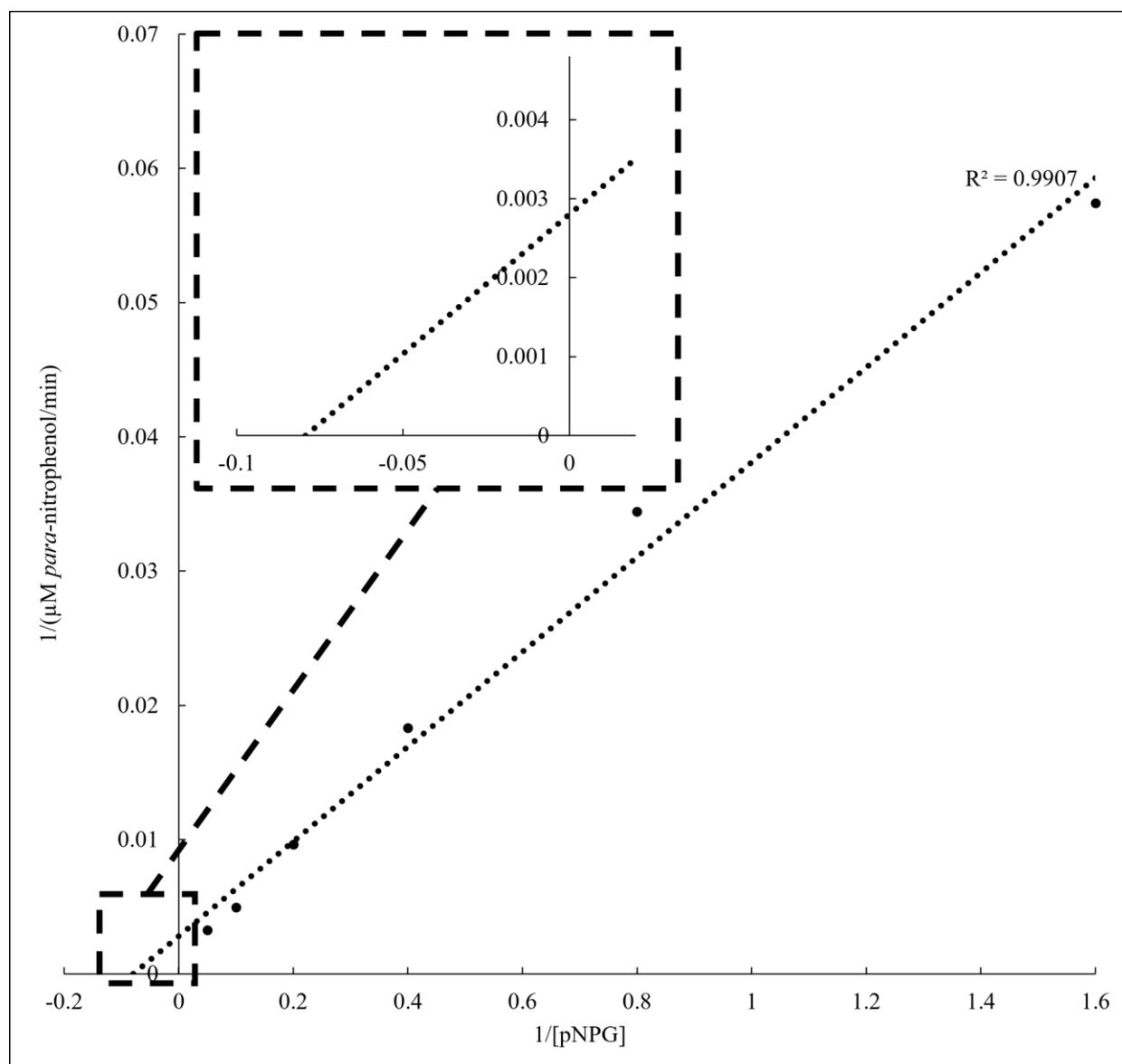
Figure S5: Lineweaver-Burk plot. Incubation of Fsa1572 with varying concentrations of pNPG for 30 minutes resulted in a highly linear Lineweaver-Burk plot ($R^2 = 0.9907$). The x- and y-intercepts were used to calculate the K_M and V_{max} values of 12.6 mM and 357.1 μM/min, respectively. The x- and y-intercept region is magnified within a dashed box.

Table S1: Characterized GH50 enzymes.

#	Enzyme name	GenBank accession number	Organism	Activity	Reference
1	VadG925	EFB02686.1	*Victivallis vadensis* ATCC BAA-548	β-galactosidase	Temuujin et al., 2012a
2	Ph1636	QEP52093.1	*Paraglaciecola hydrolytica* S66	β-agarase	Schultz-Johansen et al., 2018
3	AgaB-4	AVV48109.1	*Paenibacillus agarexedens* BCRC 16000	β-agarase	Chen et al., 2018
4	Hz1	ADY17918.1	*Agarivorans* sp. HZ105	β-agarase	Hu et al., 2009
5	agarase-B	BAG71427.1	*Vibrio* sp. PO-303	β-agarase	Araki et al., 1998
6	N/A	ABK97391.1	*Agarivorans* sp. JA-1	β-agarase	Lee et al., 2008
7	AgaA	BAA03541.1	*Vibrio* sp. JT0107	β-agarase	Sugano et al., 1993
8	AgrA	ACU52709.1	*Agarivorans* sp. AG17	β-agarase	Nikapitiya et al., 2010
9	AgaA11	BAD99519.1	*Agarivorans* sp. JAMB-A11	β-agarase	Ohta et al., 2005
10	Aga41A	ADM25828.1	*Vibrio* sp. CN41	β-agarase	Lao et al., 2011
11	AgWH50A	AFP32918.1	*Agarivorans gilvus* WH0801	β-agarase	Liu et al., 2014a
12	Wh50B	AQT38174.1	*Agarivorans gilvus* WH0801	β-agarase	Liang et al., 2017
13	Hz2	ADY17919.1	*Agarivorans* sp. HZ105	β-agarase	Lin et al., 2012
14	AgaD02	ABM90422.1	*Agarivorans* sp. QM38	β-agarase	Du et al., 2011
15	AgaB	BAA04744.1	*Vibrio* sp. JT0107	β-agarase	Sugano et al., 1993
16	DagB	CAB61811.1	*Streptomyces coelicolor* A32	β-agarase	Temuujin et al., 2012b
17	AgWH50C	AHM94172.1	*Agarivorans gilvus* WH0801	β-agarase	Liu et al., 2014b
18	N/A	BAE97587.1	*Alteromonas* sp. E-1	β-agarase	Kirimura et al., 1999
19	Aga50D	ABD81904.1	*Saccharophagus degradans* 2-40	β-agarase	Kim et al., 2010

20	Ph1609	QEP52089.1	*Paraglaciecola hydrolytica* S66	β-agarase	Schultz-Johansen et al., 2018
21	AgaB1	AGT98631.1	*Thalassotalea agarivorans* BCRC 17492	β-agarase	Liang et al., 2014
22	Aga21	AHC72907.1	*Pseudoalteromonas* sp. NJ21	β-agarase	Li et al., 2015
23	AgaA	ABD80438.1	*Saccharophagus degradans* 2-40	β-agarase	Ko et al., 2012
24	Ph1624	QEP52091.1	*Paraglaciecola hydrolytica* S66	β-agarase	Schultz-Johansen et al., 2018
25	PaBglu50A	AST24418.1	*Pseudomonas aeruginosa* CAU 342A	β-1,3-glucanase	Yi et al., 2018

CHAPTER 3: SUMMARY AND FUTURE DIRECTIONS

3.1 Summary

After purification, rigorous biochemical experimentation, and phylogenetic analysis, Fsa1572 was shown to possess several characteristics that set it apart from other enzymes from glycoside hydrolase family 50 (GH50). First, testing the active and optimal temperature range of Fsa1572 on β-glucan revealed high activity between 60 °C and 95 °C, with an optimum of 80 °C. Prior to this study, the highest optimal active temperature from GH50 was AgrA (65 °C) from *Agarivorans* sp. AG17 (Nikapitiya et al., 2010), making Fsa1572 by far the most thermophilic GH50. Second, an activity assay using the artificial β-1,4-glycosidically-linked substrate *para*-nitrophenyl-β-d-glucopyranoside (pNPG) confirmed that Fsa1572 had β-1,4-glucanase activity. 24 of the previously characterized GH50 enzymes are β-agarases, with only two exceptions: VadG925 (β-galactosidase) from *Victivallus vadensis* (Temuujin et al., 2012a) and PaBglu50A (β-1,3-glucanase) from *Pseudomonas aeruginosa* (Yi et al., 2018); thus, Fsa1572 is the first known GH50 enzyme with β-1,4-glucanase activity. Finally, an approximate maximum-likelihood phylogenetic analysis of all known GH50 family enzymes placed Fsa1572 outside of the two well-characterized GH50 clades, and distantly related to the two outliers, VadG925 and PaBglu50A. The characterization of Fsa1572 pushes the boundaries of characterized GH50s into a new phylum. Thus, Fsa1572 is distinct from other characterized GH50 enzymes.

GHs are categorized in the Carbohydrate-Active Enzyme (CAZy) Database (www.cazy.org) into families according to sequence similarity and phylogenetic relationships, and subfamilies according to phylogenetic relationships and functional conservation (Drula et al., 2022). However, both categories become less informative for GH groups that are poorly characterized. GH50 is an understudied family with only 25 characterized members out of 1,518

sequences in the CAZy database. Across the characterized enzymes, almost all are β-agarases, with few phylogenetic outliers. Because of this, GH50 enzymes have not been assigned to subfamilies. Viborg et al. (2019) delineated a well-characterized GH family, GH16, into 23 subfamilies based on functional and phylogenetic differences, and established a road map for other researchers to do the same. Because Fsa1572 is functionally unique as a hyperthermophilic GH50 with β-1,4-glucanase activity and phylogenetically distinct given its distance from other characterized GH50s, we proposed Fsa1572 represents the first characterized member of a novel GH50 subfamily, designated GH50_2, and proposed that the *Enterobacterales* and *Gammaproteobacteria* β-agarase clade should be designated GH50_1.

3.2 Future Directions

The characterization and phylogenetic analysis of Fsa1572 leaves three clear avenues to explore in future studies, and the first begins where this thesis has left off: with the description of GH50 subfamilies. Although this study has focused on Fsa1572 and its relationship to other GH50s, the results of our phylogenetic analysis, combined with functional data on other GH50s, have clearly illustrated a GH family that could be further delineated beyond the subfamilies GH_1 and GH_2 proposed by this work. Between the previously characterized GH50s, there are two functionally and phylogenetically distinct groups. First is the β-agarase group for which we have already proposed the name GH50_1. The second is the group containing *Pseudomonas* GH50s. These constitute 870 out of 1,518 GH50 sequences in the CAZy database (57%), from which the only characterized enzyme to date (PaBglu50A) possesses another unique activity for the family, β-1,3-glucanase (Yi et al., 2013). Furthermore, our phylogenetic analysis revealed several large clades from which no GH50s have been characterized to date. Further efforts to characterize

GH50s from understudied clades would expand our understanding of the family and help answer questions about the divergence of their activities.

To support the delineation of GH50 subfamilies, characterization of GHs in lesser-studied clades like those from *Fervidibacteria, Pseudomonas*, and those with no characterized members, should be conducted. The ability to culture a microorganism has been a necessary step and a barrier to the study of novel enzymes in decades past. However, metagenomics and single-cell genomics provide novel gene sequences from uncultured organisms, while recombinant DNA technology and DNA synthesis have allowed us to express and study their proteins. We only focused on one GH50 enzyme from *Fervidibacter sacchari* in this work, but another, Fsa2534, has yet to be studied. Characterization of Fsa2534 and its confirmation as either having β-1,4-glucanase activity or active on another bond altogether, as suggested by it belonging to a major *Fervidibacteria* clade closely related to the clade with Fsa1572, would provide valuable insight into the evolution of GH50. Similar characterization efforts for enzymes related to VadG925 and PaBglu50A would be valuable to bridge the evolutionary and ecological gap between GH50 β-agarases and those with other activities. An experiment examining the differential expression of *F. sacchari* GHs when grown with five different substrates (β-glucan, gellan gum, locust bean gum, starch, and xyloglucan) showed that Fsa1572 had greater expression when the organism was grown with xyloglucan compared to β-glucan (Nou, 2022). This pattern hints at complex and potentially consortial polysaccharide depolymerization and prompts further study of other GHs from *F. sacchari*.

Although our work revealed one highly unique enzyme, there is a whole cache of *F. sacchari* GHs that remains to be explored in depth. To date, the Hedlund Lab has attempted to determine the activities for 50 of 115 GHs from *F. sacchari*, of which four have been confirmed

as active. This low return raised the question of whether the wrong substrates were being tested, despite their variety. This speculation was supported by the greater degradation of oat β-glucan (alternating β-1,3/1,4-glycosidic linkages) than a saturating concentration of pNPG (β-1,4-glycosidic linkage), the incomplete hydrolysis of oat β-glucan following an exhaustive reaction compared to an acid-hydrolyzed control, as well as the low velocity of Fsa1572 on pNPG compared to other GH50s and thermophilic β-1,4-glucanases. When looking at the environment that *F. sacchari* was isolated from, Great Boiling Spring (GBS), there are several polysaccharide sources, including photosynthetic microbial mats (a rich source of exopolysaccharides) and *Distichlis spicata*, a grass surrounding the geothermal spring. A fourfold road map for characterizing Fsa1572 and other *F. sacchari* GHs is proposed.

First, the k_{cat} of Fsa1572 should be assessed and compared across pure glucans containing only β-1,3 or β-1,4 linkages. To ensure that the reactions are limited only by the rate of hydrolysis, rather than substrate availability, only water-soluble glucans will be tested. Carboxymethyl curdlan (CM-curdlan) is a water-soluble curdlan (β-1,3-glucan) that will be used to determine the β-1,3-glucanase rate of Fsa1572. Similarly, carboxymethyl cellulose (CMC) is a water-soluble cellulose with β-1,4-glycosidic linkages and thus will be used to assess the β-1,4-glucanase rate of Fsa1572. The kinetic parameters of these reactions will be determined by incubating 67.2 µg/mL of Fsa1572 (67.2 micrograms/mL) with various concentrations of CM-curdlan or CMC for 30 minutes and quantification of reducing sugars by the 3,5-dinitrosalicylic acid assay, followed by use of a Lineweaver-Burk plot to calculate V_{max}, K_M, and k_{cat}. To assess reaction completeness, the reducing sugars released by each reaction will be compared to the total reducing sugar content of CM-curdlan and CMC, as determined by acid hydrolysis. Faster and more complete degradation of either β-1,3-glucan or β-1,4-glucan would support the favored hydrolysis of that bond by

Fsa1572. We hypothesize that Fsa1572 cleaves the β-1,3-glycosidic linkage of β-glucans more effectively than their β-1,4-glycosidic linkages. Thus, we expect that this experiment will show that Fsa1572 has a high k_{cat} with CM-curdlan and a low k_{cat} with CMC. Furthermore, no substrate tested in this study contained β-1,6 bonds, which are found in the branches of fungal β-glucans. An additional test using an enzymatic yeast β-glucan assay kit (Megazyme) containing β-1,3 and β-1,6 linkages should be conducted and compared to activity from CM-curdlan to determine if Fsa1572 can hydrolyze β-1,6 linkages. In addition to determining the preferred substrate of Fsa1572, a more thorough multimeric analysis should be conducted. For this work, a tetrameric Fsa1572 structure prediction was attempted but failed due to limiting GPU from the free ColabFold platform (Mirdita et al., 2022). Due to the moderately confident homodimer prediction and the structural similarity of Fsa1572 to Aga50D, a tetramer comprising two identical dimers (Pluvinage et al., 2013), we hypothesize that Fsa1572 also tetramerizes. Thus, an AlphaFold 2 structural prediction for the tetramerization of Fsa1572 should be conducted using a purchased GPU. Second, microbial mat homogenate and *D. spicata* biomass should be tested as growth substrates for *F. sacchari* and for *F. sacchari* GH activity screens. Third, the remaining 65 GHs should be expressed and investigated for activity. Finally, the other three active *F. sacchari* GHs: a GH3 xylanase, a GH5 xylanase, and a GH10 glucanase should be fully characterized. *F. sacchari* occupies a fascinating niche as an aerobic and hyperthermophilic polysaccharide-degrader and developing a solid understanding of its GH cache would supply valuable knowledge about the life of such a unique microbe and others like it that have yet to be discovered.

A

β-agarases:

2 Ph1636	10 Aga41A	18 BAE97587.1
3 AgaB-4	11 AgWH50A	19 Aga50D
4 Hz1	12 Wh50B	20 Ph1609
5 agarase-B	13 Hz2	21 AgaB1
6 ABK97391.1	14 AgaD02	22 Aga21
7 AgaA	15 AgaB	23 Aga50A
8 AgrA	16 DagB	24 Ph1624
9 AgaA11	17 AgWH50C	

β-galactosidase: **β-1,4-glucanase:**
1 VadG925   25 PaBglu50A

B

Diverse bacteria (137 sequences)

Diverse bacteria (14 sequences)

Firmicutes (4 sequences)

Diverse bacteria (11 sequences)

'QNN20901.1 hypothetical protein HED60 00995 *Planctomycetales* bacterium ZRK34'

Class *Fervidibacteria* with Fsa2534 (30 sequences)

Proteobacteria (75 sequences)

Verrucomicrobiota (3 sequences)

Cyanobacteria (5 sequences)

Class *Fervidibacteria* with Fsa1572 (45 sequences)

'ADG91381.1 agarase *Thermosphaera aggregans* DSM 11486'
'AJC71238.1 beta-agarase *Thermococcus guaymasensis* DSM 11113'
'ACS89391.1 beta-agarase extracellular *Thermococcus sibiricus* MM 739'
'ASI98670.1 beta-agarase *Thermococcus celer*'
'ASJ07082.1 beta-agarase *Thermococcus pacificus*'
'ASJ00655.1 beta-agarase *Thermococcus gorgonarius*'

Figure 6: Phylogenetic analysis of GH50. (A) An approximate maximum-likelihood analysis of GH50 using SH-aLRT branch support. Sequences from *Fervidibacteria* are denoted by red branches. Fsa1572 and Fsa2534 are indicated by red stars, while other characterized GH50s are indicated by black branches, black stars, and an identifying number that corresponds to their protein names or GenBank accession numbers if no name was listed. Clusters from the same genera are marked with colored highlights, and characterized GH50s are clustered by their lowest common taxonomic rank. The sequences included in panel B are circled by a dashed gray line. Well-supported branches (SH-aLRT ≥ 0.9) are indicated by a black dot. The scale bar shows the number of amino acid changes per site. (B) A maximum-likelihood analysis of the Fsa1572 lineage and its closest relatives (circled in dashed grey outline) using SH-aLRT and UFBoot branch support. Sequences are clustered into five *Fervidibacteria* clades (red; the number of sequences is numbered at the tips), the archaeal GH50s (blue), and related GH50s by taxonomic rank. Well-supported branches (SH-aLRT ≥ 0.9 and UFBoot ≥ 95%) are indicated by a black dot. The scale bar shows the number of amino acid changes per site.

```
Fsa 1572/1-454     1  M...........................................                1
Fsa 2534/1-1154    1  MR-KL.......................................                4
Aga 50D/1-793      1  MG..........................................                2
Aga A/1-995        1  MKIKFLSAAIAASLALPLSAATLVTSFEEADYSSSENNAEFLEVSGDAT          49
Agr A/1-995        1  MKIKFLSAAIAASLALPLSAATLVTSFEEADYSSSENNTEFLEVSGDAT          49
PaBglu 50A/1-749   1  MI..........................................                2
Vad G925/1-1425    1  MS-KLLHLYV...........ILGTIF.................               15

Fsa 1572/1-454     2  ...................................SW.........              3
Fsa 2534/1-1154    5  ...............................AWVVGAI.........            11
Aga 50D/1-793      3  .......AIGGLVKINISFIPLFVI.....SASIFIGAC........            29
Aga A/1-995       50  SEVSTEQATDGNQSIKASFDAAFKPMVVWNWASWNWGAEDVMSVDVVNP          98
Agr A/1-995       50  SEVSTEQATDGNQSIKASFDAAFKPMVVWNWGSWNWGAEDVMSVDVVNP          98
PaBglu 50A/1-749   3  ...............................RSRWHLPLLL.......           12
Vad G925/1-1425   16  ............SLEAAPIPANPP.....VTAW..............            31

Fsa 1572/1-454        ............................................
Fsa 2534/1-1154   12  ....AAFAL..........WL.........ASANAHPFRFS.....            30
Aga 50D/1-793     30  ...................NSSKLESGVDSNNISPVMLF........            49
Aga A/1-995       99  NDTDVTFAIKLIDSDILPDWVDESQTSLDYFTVSANTTQTFSFNLNGGN         147
Agr A/1-995       99  NDTDVTFAIKLIDSDILPDWVDESQTSLDYFTVSANTTQTFSFNLNGGN         147
PaBglu 50A/1-749  13  ..GLLAVATPLAASDI...................QQVLF.......            31
Vad G925/1-1425       ............................................

Fsa 1572/1-454        ............................................
Fsa 2534/1-1154   31  ......DDFSK...................YPAGSIGE                     43
Aga 50D/1-793     50  .......DFENDQV.........PSNIHFLNAR.............             66
Aga A/1-995      148  EFQTHGENFSKDKVIGVQFMLSENDPQVLYFDNIMVDGETVTPPPSDGA         196
Agr A/1-995      148  EFQTHGENFSKDKVIGVQFMLSENDPQVLYFDNIMVDGETVTPPPSDGA         196
PaBglu 50A/1-749  32  ......NFVK.............PMAVVGITLE...DADLP......            50
Vad G925/1-1425   32  ...............PQLLRQGKLQ.............                     41

Fsa 1572/1-454        ............................................
Fsa 2534/1-1154   44  PNWEVTDI...................GFEIRNGTMHAEVVAGRGYAI          72
Aga 50D/1-793     67  ...........ASIETYTGING-EPSKGLKL...AMQSKQHSYTGLAI          99
Aga A/1-995      197  VNTQTAPVATLAQIEDFETIPDYLRPDGGVNV...STTTEIVTKGAAAM        242
Agr A/1-995      197  VNTQTAPVATLAQIEDFETIPDYLRPDGGVNV...STTTEIVTKGAAAM        242
PaBglu 50A/1-749  51  ...................SATAEATPEGDILR                         64
Vad G925/1-1425   42  ...................GSPESSFACDGEML                         55

Fsa 1572/1-454        ............................................
Fsa 2534/1-1154   73  LKVAPMGRSVTVEATLTVHRAISTGWKIAGVG....IFVD.....ERNY        112
Aga 50D/1-793    100  V.....................PEQPWDWSEFT-SASLYFDIVSVGDHST       127
Aga A/1-995      243  AAEFTAGWNGLV.......FAGTWNWAELGEHTAVAVDVSNNSDSNI          282
Agr A/1-995      243  AAEFTAGWNGLV.......FAGTWNWAELGEHTAVAVDVSNTSDSNI          282
PaBglu 50A/1-749  65  RVTFSPAQRPTLRMSP.....ALGRWDWSAAD-YVSLRIQNAMSWDMTL        107
Vad G925/1-1425   56  VFSVPAGRGGSM.........TFPVTEPSAGESFTRLELELENRGDTGT         95

Fsa 1572/1-454        ............................................
Fsa 2534/1-1154  113  WHIALVESPDNQGK....RHFAELHQMLDGVWLSDVQDPTRLTTEEETG        157
Aga 50D/1-793    128  QFYLDVTD--QNGAV-FTRSIDIPVGKMQSYYAKLSGHDLEVPDSGDVN        173
Aga A/1-995      283  WLYSRIEDVNSQGET-ATRGVLVKAGESKTIYTSLNDNPSLLTQDERVS        330
Agr A/1-995      283  WLYSRIEDVNSQGET-ATRGVLVKAGESKTIYTSLNDNPSLLTQDERVS        330
PaBglu 50A/1-749 108  EVAIE....GEQGAPGLQASIELPAGPPQTLLVPLRAVSPE........        144
Vad G925/1-1425   96  GFTIRLRS--ADGRA-VTVHCRIEAGQCKQAVAPMPRDRKTLETR....        137

Fsa 1572/1-454     4  ..............TR............................                5
Fsa 2534/1-1154  158  GFDWEYGRPYRLR...............IALSKERV......I              179
Aga 50D/1-793    174  DLNLASG....LRSNPPTWTSDDRQFVWMWGV-KNLDLSGIAKISL--         214
Aga A/1-995      331  ....ALG....LRDIPADPMSAQN....GWGD-FVALDKSQITAIRYFI        366
Agr A/1-995      331  ....ALG....LRDIPADPMSAQN....GWGD-FVALDKSQITAIRYFI        366
PaBglu 50A/1-749 145  ....ALG....MRAGPPMPQMVEGQRVLLAPRVEGSLDRARVGALSL--        183
Vad G925/1-1425  138  ...YAAQ....LRDFPKMAGLPGG....LFEN-WESVDAGDLEAVTL--        172
```

```
Fsa 1572/1-454       6  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - R E F V K L V G F A T T   17
Fsa 2534/1-1154    180  G E V F D A D - - - - - - - G K L R Y R C V R R L D K R A V T F G - - - - R P I L T C G G F V A T   217
Aga 50D/1-793      215  - S V Q S A M H D K T V I I D N I R I Q P N P P Q D E N F L V - - - - - - - - - - - - - G L V D E   249
Aga A/1-995        367  G E L A S G E T S Q T L V F D N M R V I K D L N H E S A Y - - - - - - - - - - - - - A E M T D A   401
Agr A/1-995        367  G E L A S G E T S Q T L V F D N M R V I K D L N H E S A Y - - - - - - - - - - - - - A E M T D A   401
PaBglu 50A/1-749   184  - S L R S P Q A P Q S I L L G R F G I R T G R A V E R S I L T - - - - - - - - - - - - - G L I D R   218
Vad G925/1-1425    173  - E F P E A A A P L T L R I G Q V R F - - S H P A A P S L Y A A A P E R F - - - - - - F P F I D R   212

Fsa 1572/1-454      18  F A G - - - - - - - - - - - - - - - - - - - - - A S C R S E G Q E E A K F C R Y G G W L A M   42
Fsa 2534/1-1154    218  F D D - - - - - - - - - - - - - - - - - - - V K A E V N E I V - - S E P K R E R K T Y P P F V S R   245
Aga 50D/1-793      250  F G Q N A K V D Y K G K I H S L E E L H A A R D V E L A E L D - G K P M P S R S K F G G W L A G   296
Aga A/1-995        402  M G Q N N L V T Y A G K V A S K E E L A K L S D P E M A A L G - - - - E L T N R N M Y G G N P D S   446
Agr A/1-995        402  M G Q N N L V T Y A G K V A S K E E L A K L S D P E M A V L G - - - - E L T N R N M Y G G N P D S   446
PaBglu 50A/1-749   219  Y G Q Y S R A D W P E K I R S D E Q L R S A Y A A E A A Q L R D W E R Q T P A R D R F G G L L G G   267
Vad G925/1-1425    213  Y G Q Y I H E G W P G K V K S D Q E L K L A W E A E K N D L A A - H R R P A S F S K F G G W R E G   260

Fsa 1572/1-454      43  K R - - - - - - - K A T G F F R T E Q V N G V W W F V D P D G H L F I S K G V N H V S F G G D Y C   84
Fsa 2534/1-1154    246  P S P H A P R P V K G T G F F R T Q Q I N G I W W L I D P N G H P T L S I G T D H V N Y F V H W C   294
Aga 50D/1-793      297  P K - - - - - - L K A T G Y F R T E K I N G K W M L V D P E G Y P Y F A T G L D I I R L S N S S T   339
Aga A/1-995        447  S P A T D C V L A T P A S F N A C K D A D G N W Q L V D P A G N A F F S T G V D N I R L Q D T Y T   495
Agr A/1-995        447  S P T T D C V L A T P A S F N A C K D A D G N W Q L V D P A G N A F F S T G V D N I R L Q D T Y T   495
PaBglu 50A/1-749   268  P V - - - - - - F E A S G F F R T E K R G G R W W L V T P E G H P F W S L G V N A V T A D G S R T   310
Vad G925/1-1425    261  P R - - - - - - F K A T G S F Y P K K Y D G K W Y L A D P E G Y L F W S N G I N C V R A E T Y G T   303

Fsa 1572/1-454          . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Fsa 2534/1-1154         . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Aga 50D/1-793      340  M T G Y D Y D Q A T V A Q R S A D D V T P E D S K G L M A V S E K S F A T R H L A S P T R A A M F   388
Aga A/1-995        496  M T G V S S D A E S - - - - - - - - - - - - - - - - - - - - - - - - - - E S A L R Q S M F   514
Agr A/1-995        496  M T G V S S D A E S - - - - - - - - - - - - - - - - - - - - - - - - - - E S A L R Q S M F   514
PaBglu 50A/1-749   311  Y V - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - E G R E P M F   319
Vad G925/1-1425    304  L V - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - E G R E R F F   312

Fsa 1572/1-454      85  - - - - - - - - - - - - - - - - - - - - - - - - P A L G Y S P Y N R N V Q A K Y   100
Fsa 2534/1-1154    295  - - - - - - - - - - - - - - - - - - - - - - - - E K L G Y A P Y H E N V K R K Y   310
Aga 50D/1-793      389  N W L P D Y D H P L A N H Y N Y R R S A H S G P L K - - - - - - - R G E A Y S F Y S A N L E R K Y   430
Aga A/1-995        515  T E I P S - - D Y V N E N Y G - - - P V H S G P V S - - - - - Q G G A V S F Y A N N L I T R H   551
Agr A/1-995        515  T E I P S - - D Y V N E N Y G - - - P V H S G P V S - - - - - Q G G A V S F Y A N N L I T R H   551
PaBglu 50A/1-749   320  A E L P A E G E P L A A F F G E G - D D R R G V A A Q A G R R F G H G R W F D F L G A N R Q R I A   367
Vad G925/1-1425    313  A G L P E A G T P C S A F F H N T - E R Y G G - - - - - - - - - K H L A Y N F A A A N L L R K Y   350

Fsa 1572/1-454     101  G S E - - - - - - E K W A E - - - - A T A K R L R E W G F N T I G A W S - S R S L F K L - - - - M   134
Fsa 2534/1-1154    311  G S E - - - - - - E A W A K - - - - E A V K R L L S W N F N V L G A N N S V K A R Y Q G - - - - L   345
Aga 50D/1-793      431  G E T Y P G S Y L D K W R E - - - - V T V D R M L N W G F T S L G N W T - D P A Y Y D N - - N R I   472
Aga A/1-995        552  A S E - - - - - - D V W R D - - - - I T V K R M K D W G F N T L G N W T - D P A L Y A N - - G D V   587
Agr A/1-995        552  A S E - - - - - - D V W R D - - - - I T V K R M K D W G F N T L G N W T - D P A L Y A N - - G D V   587
PaBglu 50A/1-749   368  P Q A S A D Q L A G E W R R - - - - R T L E R L S A W G F N S L G N W S - D P A L A A Q - - A R M   409
Vad G925/1-1425    351  G P D - - - - - - - - W R D L H L G L T G K R L A S W G F N T V G N W S - D - - - Y A G K I A D F   387

Fsa 1572/1-454     135  P Y T V I L N M G A - - - - - - - - R V G A D W - - - - L K G S F P D V F S P K F R Q V L D E I A   171
Fsa 2534/1-1154    346  A H T E F L A F G S - - - D F - - - A S T A D I V P K V H W T G F P D V F D P R F E R F C D L R A   388
Aga 50D/1-793      473  P F F A N G W V I G - - - D F K T V S S G A D - - - - - F W G A M P D V F D P E F K V R A M E T A   513
Aga A/1-995        588  P Y V A N G W S T S G A D R L P V K Q I G S G - - - - - Y W G P L P D P W D A N F A T N A A T M A   631
Agr A/1-995        588  P Y V A N G W S T S G A D R L P V K Q I G S G - - - - - Y W G P L P D P W D A N F A T N A A T M A   631
PaBglu 50A/1-749   410  P Y S L P L S I A G - - - D Y A T V S S G F D - - - - - W W G A M P D P F D P R F A M A A E R V I   450
Vad G925/1-1425    388  P Y V L T L G H A C R - - - - - - - S S G - - - - - - - - - G W R D D P F D P A H R E N I E Q A L   420

Fsa 1572/1-454     172  A R - E C A P R K N D P L L V G Y F T D N E L R W - G P D W R S P - - - - - R H L L D D Y L L L L   213
Fsa 2534/1-1154    389  K R - R C A P N K D D P W L L G Y F L D N E L E W W G K S G R - - - - - - P W G M A E E A W K K   429
Aga 50D/1-793      514  R V - V S E E I K N S P W C V G V F I D N E K S F - G R P D S D K A - - - Q Y G I P I H T L G R   556
Aga A/1-995        632  A E I K A Q V E G N E E Y L V G I F V D N E M S W - G N V T D V E G S R Y A Q T L A V F N T D G T   679
Agr A/1-995        632  A E I K A Q V E G N E E Y L V G I F V D N E M S W - G N V T D V E G S R Y A Q T L A V F N T D G T   679
PaBglu 50A/1-749   451  A I - A A R D H R D D P W L L G Y Y A D N E L A W A G R D G S A Q A - - - - R Y G L A F G A L T L   494
Vad G925/1-1425    421  Q S G R Y D A A I R D P Y C I G F F S G N E L P W - S D P V T Y A G - - - - - - - - - - N L L R   457
```

41

```
Fsa 1572/1-454      214  PADA-PGKKVLLEFFRKRYQT--VEAFNKAWGLNVPNWDELANLKDLPP         259
Fsa 2534/1-1154     430  PKDR-ACKQALVRILSEFYRD-DIKAFNADFGTNFARFDDLLT-SQTPS         475
Aga 50D/1-793       557  PSEGVPTRQAFSKLLKAKYKT--IAALNNAWGLKLSSWAEFDLGVDVKA         603
AgaA/1-995          680  DATTSPAKNSFIWFLENQRYTGGIADLNAAWGTDYASWDAT---SPAQE         725
AgrA/1-995          680  DATTSPAKNSFIWFLENQRYTGGIADLNAAWGTDYASWDAM---RPAQE         725
PaBglu 50A/1-749    495  SMDS-PAKRAFVKQLKAKYLG--HEALAEAWGIELAAWEALEAPGYAAP         540
Vad G925/1-1425     458  EKKS-PARSALTDALKRKFGK--IEQLNTVLDTAFDSWRELEE-NPVKA         502


Fsa 1572/1-454      260  APSGEIEKQRLADRLDFLRIIAREYFRACYEAIRNHDSNHLILGVRFAG         308
Fsa 2534/1-1154     476  QPLNE---RGQKALMAFVREAAERYFRITAQAIRKYDPNHLNLGCRFAW         521
Aga 50D/1-793       604  LPVTD---TLRADYSMLLSAYADQYFKVVHGAVEHYMPNHLYLGARFPD         649
AgaA/1-995          726  LAYVA---GMEADMQFLAWQFAFQYFNTVNTALKAELPNHLYLGSRFAD         771
AgrA/1-995          726  LAYVA---GMEADMQFLAWQFAFQYFNTVNTALKAELPNHLYLGSRFAD         771
PaBglu 50A/1-749    541  LPGEG-HPAIAEDYSAFLRLYADAYFKTLRDALQWHAPNHLLLGGRFA-         587
Vad G925/1-1425     503  LPDLT---PIEAECRNFYRTALEHYFRPIRDMIRRYAPGKLYLGVRFMH         548


Fsa 1572/1-454      309  YAPRP-VVEAMGEFVDVVSFNWYG-FEPPIK---TLE----GLHQITGK         348
Fsa 2534/1-1154     522  DAPEP-AWEMAGKYCDIVTVNLYPRIDLERGVVLGIEEHLRKRYELCRK         569
Aga 50D/1-793       650  WGMPMEVVKAAAKYADVVSYNSYK-EGLPKQ---KW-----AFLAELDK         689
AgaA/1-995          772  WGRTPDVVSAAAAVVDVMSYNIYK-DSIAAA---DWDADALSQIEAIDK         816
AgrA/1-995          772  WGRTPDVVSAAAAVVDVMSYNIYK-DSIAAA---DWDADALNQIEAIDK         816
PaBglu 50A/1-749    588  -VSTPEAITSCARYCDLLSFNLYT--PLPGQ---GLDD---SLLARLDK         627
Vad G925/1-1425     549  YPDIT-LGEVAADFCDVVSMNRYSYDESE-----------MRIPGRDV         584


Fsa 1572/1-454      349  PVMITEFSFKAMDSG------LPNTRGAGKPVPTQKDRADRFERYVTAL         391
Fsa 2534/1-1154     570  PIIVTEWSFPALDAKDSQGRPLPCKHGAGMRVDNQTQKARCYAIMQRTL         618
Aga 50D/1-793       690  PSIIGEFHIGAMDHG--------SYHPGLIHAASQADRGEMYKDYMQSV         730
AgaA/1-995          817  PVIIGEFHFGALDSG--------SFAEGVVNATSQQDRADKMVSFYESV         857
AgrA/1-995          817  PVIIGEFHFGALDSG--------SFAEGVVNATSQQDRADKMVSFYESV         857
PaBglu 50A/1-749    628  PVLISEFHFGSRDRG-------PFWGGVSEAANERARGDSYRTFLEAA         668
Vad G925/1-1425     585  PILVTEFHFGALDRG-------MFHPGLRWGGDQQDVAWCYTEYMKAL         625


Fsa 1572/1-454      392  MKLPYCVGYHWFQWSDQPAQG-R-FDGENSNYGLVKETDEAWELLTQRM         438
Fsa 2534/1-1154     619  FSLPFIVGSHYFMWVDEPALGISSTFPEDSNYGLVNEADEPYPELTEMA         667
Aga 50D/1-793       731  IDNPYFVGAHWFQYMDSPLTG-RAYDGENYNVGFVDVTDTPYQEMVDAA         778
AgaA/1-995          858  NAHKNFVGAHWFQYIDSPLTG-RAWDGENYNVGFVSNTDTPYTLMTDAA         905
AgrA/1-995          858  NAHKNFVGAHWFQYIDSPLTG-RAWDGENYNVGFVSNTDTPYTLMTDAA         905
PaBglu 50A/1-749    669  LKSPYIVGAHWFQYLDQPASG-RLLDGENGHIGLVGITGLPFAGFVDTV         716
Vad G925/1-1425     626  LRNPACVGGHWFQYASQPFTG-RCGDGENAQIGAVDITDNPRPEFRAAL         673


Fsa 1572/1-454      439  KEVNGRIEEVHA----------------------------------          450
Fsa 2534/1-1154     668  TKVNAQMFALHSGMTAEL---SVIVVPHQKVMRVDNMGKVAATFTLALW         713
Aga 50D/1-793       779  KEVNAKIYTERLGSK-------------------------------          793
AgaA/1-995          906  REFNCGMYGTDCSSLSNA---TEAASRAGELYTGTNIGVSHSG------         945
AgrA/1-995          906  REFNCGMYGTDCSSLSNA---TEAASRAGELYTGTNIGVSHSG------         945
PaBglu 50A/1-749    717  RRSNLAALSRLSAMARSM----------------------------         734
Vad G925/1-1425     674  REVGEHLYRWRAGAQHSGDGGPVIVLPPDAIPAERTAAQELQAFLEQIS         722


Fsa 1572/1-454           ------------------------------------------------
Fsa 2534/1-1154     714  VNGKRTDRRIELKPKLSLVDTLPIDLPENEAVYIRAVCDPEDEVPERNE         762
Aga 50D/1-793            ------------------------------------------------
AgaA/1-995          946  -----------------------P--EA------------PDPGE         953
AgrA/1-995          946  -----------------------P--EA------------PDPGE         953
PaBglu 50A/1-749         ------------------------------------------------
Vad G925/1-1425     723  GEKPPIAERADAGPAIHIGQS-----P--EAARALGVSDWKQLKPD---         761


Fsa 1572/1-454           ------------------------------------------------
Fsa 2534/1-1154     763  TDNVAEAVLPPKGQ-----GTRGKGQVKQICAVA--------WNPTE--         796
Aga 50D/1-793            ------------------------------------------------
AgaA/1-995          954  ---------------------------------------PVD--         956
AgrA/1-995          954  ---------------------------------------PVD--         956
PaBglu 50A/1-749         ------------------------------------------------
Vad G925/1-1425     762  -----EIILKTVGNRLYLAGDRPRGSLYAVYELLERAYGVRFWSPAATR         805
```

42

```
Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   797 - - - Q T L R N V P V S V P - - L P P A L - S S A E D I F V S D - - - - - - A K G N L L P S Q V   832
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995        957 - - - - - - - - - P P I D P P - T P P T - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   966
AgrA/1-995        957 - - - - - - - - - P P I D P P - T P P T - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   966
PaBglu 50A/1-749  735 - - - - - - - - - - P A V E P - - L P P R E - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   744
Vad G925/1-1425   806 V P R A S L Q Q L P R I D L R Y A P P F E V R S V G S I L T R N D F R Y A V R L R H N G Q S A F V   854

Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   833 S D K L G S V T V L V K E L K P Y S A V T L W L S T T K G G R G S S R A E I P F A V H H A A K G E   881
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995        967 - - - - G G V T - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   970
AgrA/1-995        967 - - - - G G V T - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   970
PaBglu 50A/1-749      - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Vad G925/1-1425   855 P P E W G G L V T L L G N V H T F S E - - - Y P D N A L I P R D S G F R E H P - - - E W F A E R D   897

Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   882 G Y A I E T P K L L L I K D E P D G D A F D R I - - I L R D G - - - - - - - G R G A R D E E I E   920
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgrA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
PaBglu 50A/1-749      - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Vad G925/1-1425   898 G R R V P N G Q L C L T N P E L H R E L V R R V R E L L R A A P E S R Y I S V S Q N D N D D F C Q   946

Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   921 L G S F T P L I W Q V V A G Q N L W V R P D R V E R I E V V D V G P A I L V V D I V F V K G R G T   969
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgrA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
PaBglu 50A/1-749      - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Vad G925/1-1425   947 C R S C A A F V E K H G N Q S D L L L D T V N A V A A A V A E E F P A T L V E T L A Y R Y T R T P   995

Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   970 G D E G R V I T E V - - - - - - G A G G K F E P L K A E P Q P F R C A Y R F T F F P D Q P F F L S   1012
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgrA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
PaBglu 50A/1-749      - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Vad G925/1-1425   996 P A T V K A A P N V L I R Y C T F E A D S F R P L T A K - - - - - - - - - - - - - Q N R Q F F R D   1031

Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   1013 Q C L W V E N T G K Q A W Q W R G Y Y H Y A L S R I G G N S A D D E V G G V N V P N Y W L Q F A S   1061
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgrA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
PaBglu 50A/1-749      - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Vad G925/1-1425   1032 L A A W - S R T A K Q L M I W N Y I - - - - - - - - - - - - - - - - - - A N F R K Y Y L P T P N   1060

Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   1062 W R - - D P K L R L - - H Y G V I S L R E D E R L G F W F W K D E G G N Q - - - - - - - - - - - -   1094
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995        971 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - G G G G S - - - - - - - - - - - -   975
AgrA/1-995        971 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - G G G G S - - - - - - - - - - - -   975
PaBglu 50A/1-749  745 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - D S A G S - - - - - - - - - - - -   749
Vad G925/1-1425   1061 W R A L G P D L R T F R S F G A I S V Y E Q - - - G A W - - - - N G G G S V S D L P E L R T W L T   1102

Fsa 1572/1-454        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Fsa 2534/1-1154   1095 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - H P D C - -   1098
Aga 50D/1-793         - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgaA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AgrA/1-995            - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
PaBglu 50A/1-749      - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Vad G925/1-1425   1103 A R L L W N P D L D T D T L I D E F L T G Y Y G P G A D A V R T Y M K L M N G A A D R H P E V A G   1151
```

43

```
Fsa 1572/1-454         · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Fsa 2534/1-1154        · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Aga 50D/1-793          · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgaA/1-995          976 · · · · · A G W L S · · · · · · · · · · · · · · · · · · · · · · · · · · · · L L G L A · · · · G V 987
AgrA/1-995          976 · · · · · A G W L S · · · · · · · · · · · · · · · · · · · · · · · · · · · · L L G L A · · · · G V 987
PaBglu 50A/1-749       · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Vad G925/1-1425    1152 S G F L S T T A A W L E E S T L L E A W Q A V E T A A R Q F R N D P V Y G P R L A M A T V P V G A 1200

Fsa 1572/1-454      451 · · · E K R T · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · 454
Fsa 2534/1-1154    1099 · · · H R R L E G T L K P G E R W Q P K E P E · · · · · · · · · · · P V V A V F G A F E T D · · · 1130
Aga 50D/1-793          · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgaA/1-995          988 F L L R R R · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · 993
AgrA/1-995          988 F L L R R R · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · 993
PaBglu 50A/1-749       · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Vad G925/1-1425    1201 A L L E R R · · · · · K P L E A W K I T V P D L R K L S T A G Y L A E T V A R M K A G G G D R L R 1244

Fsa 1572/1-454         · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Fsa 2534/1-1154    1131 · · · · · D N P R P W · S D L I · · · · · · · · · · · · · · · · · · · · · · · · · · · W W · · · · 1142
Aga 50D/1-793          · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgaA/1-995             · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgrA/1-995             · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
PaBglu 50A/1-749       · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Vad G925/1-1425    1245 E E S F H N T P D E W A A D L T A E M Q V F N R V L P N D G P A P A V A A G K P G A A W W R V E R 1293

Fsa 1572/1-454         · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Fsa 2534/1-1154    1143 L R S W A K I G V K V F · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · 1154
Aga 50D/1-793          · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgaA/1-995          994 · · · · · · · · · K V · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · 995
AgrA/1-995          994 · · · · · · · · · K V · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · 995
PaBglu 50A/1-749       · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Vad G925/1-1425    1294 I A D F H E V G R K V F P E N D P A A S T G R A V R I A D A E P G W N V Q V R R F P H G R F R F C 1342

Fsa 1572/1-454         · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Fsa 2534/1-1154        · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Aga 50D/1-793          · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgaA/1-995             · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgrA/1-995             · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
PaBglu 50A/1-749       · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Vad G925/1-1425    1343 A E V R G D L A K N A S G D A L E I V A F N W N S K E K T T R R F P V S K I G R S H Y R I V D F G 1391

Fsa 1572/1-454         · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Fsa 2534/1-1154        · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Aga 50D/1-793          · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgaA/1-995             · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
AgrA/1-995             · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
PaBglu 50A/1-749       · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
Vad G925/1-1425    1392 P V T L R D D L G V M L I P A R N P A V R N L R I D R I I L V P E R                             1425
```

Percent identity (%)

43   71   86

Figure S1: Multiple sequence alignment of GH50 enzymes. MAFFT-DASH was used to generate a multiple sequence alignment of seven GH50s: Fsa1572, Fsa2534, Aga50D, AgaA, AgrA, PaBglu50A, and VadG925. The alignment was visualized in Jalview v2.11.2.6. Two conserved glutamic acid residues of GH-A are highlighted in red, and the percent identity is indicated by purple.

46
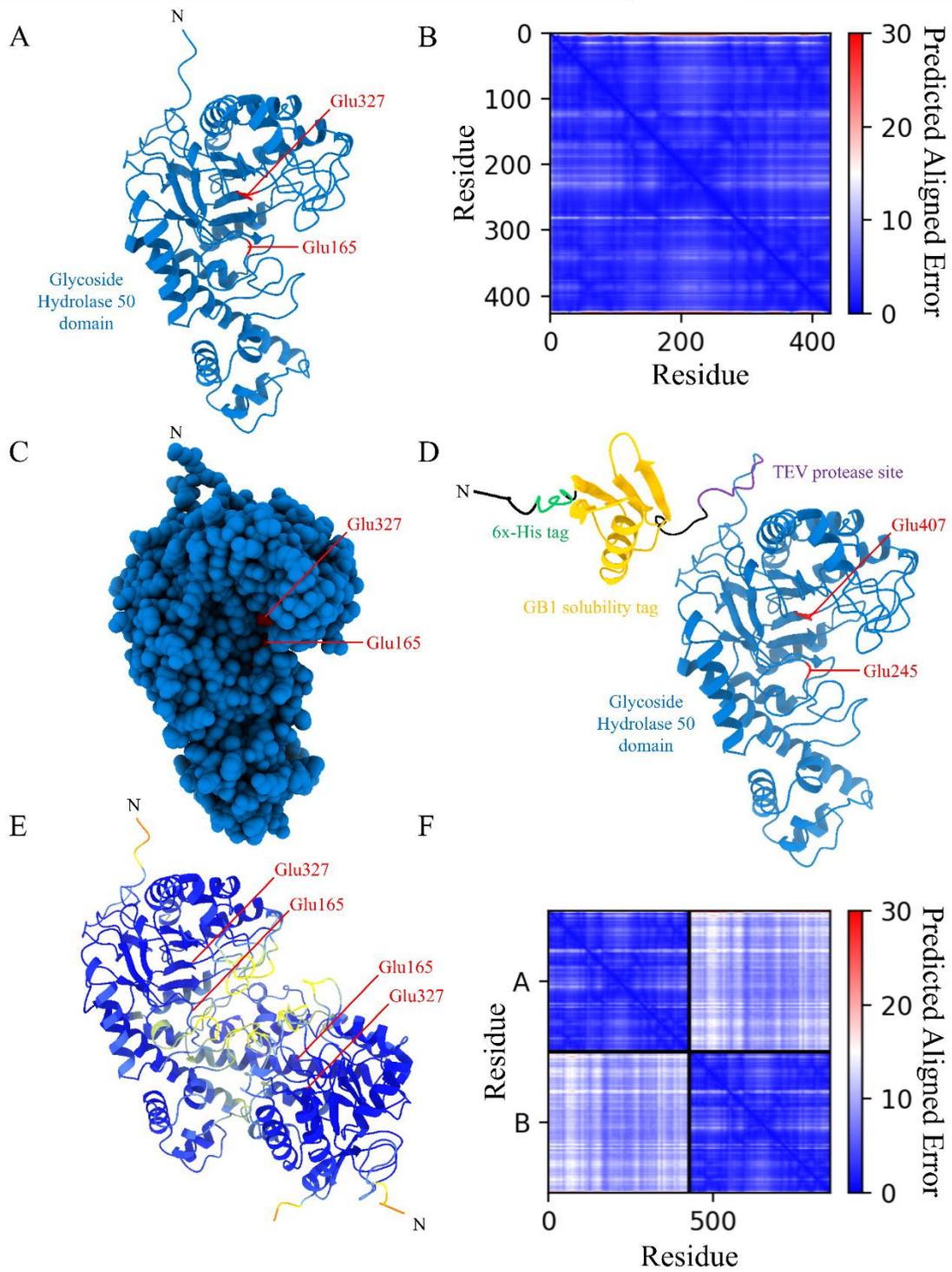
Figure S3: AlphaFold 2 structure predictions of Fsa1572. ColabFold integrated with ChimeraX v 1.5 was used to generate four Fsa1572 structure predictions using AlphaFold 2. (A) Predicted structure of native, mature Fsa1572 (blue) with conserved residues Glu165 and Glu327 (red). (B) PAE plot of the native mature Fsa1572 predicted structure showing the distance error for each residue pair. (C) Predicted structure surface representation of native, mature Fsa1572 (blue) with conserved residues Glu165 and Glu327 (red). (D) Predicted structure of non-native Fsa1572 containing an N-terminal 6x-His tag (green), GB1 solubility tag (yellow), and TEV protease site (purple), with conserved residues Glu245 and Glu407 (red). (E) Predicted homodimer structure of native, mature Fsa1572 with predicted local distance difference test (pLDDT) coloring where residue confidence ranges from low (red) to high (blue). (F) PAE plot of the native, mature Fsa1572 homodimer predicted structure revealing moderate confidence in the homodimer structure.

REFERENCES

Akram, F., and Haq, I. (2020). Overexpression and characterization of TnCel12B, a hyperthermophilic GH12 endo-1,4-β-glucanase cloned from *Thermotoga naphthophila* RKU-10T. *Anal. Biochem*. 599, 113741. doi: 10.1016/j.ab.2020.113741

Araki, T., Hayakawa, M., Lu, Z., Karita, S., and Morishita, T. (1998). Purification and characterization of agarases from a marine bacterium, *Vibrio* sp. PO-303. *J Mar Biotechnol*. 6, 260-265.

Aspinall, G.O. (1970). *Polysaccharides*. Elmsford, NY: Pergamon Press, 43-53.

Bhalla, A., Bansal, N., Kumar, S., Bischoff, K., and Sani, R. (2013). Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresour. Technol*. 128, 751-759. doi: 10.1016/j.biortech.2012.10.145

Bhattacharya, T., Ghosh, T., and Mande, S. (2015). Global profiling of carbohydrate active enzymes in human gut microbiome. *PloS One*. 10, E0142038. doi: 10.1371/journal.pone.0142038

Blank, C., Cady, S., and Pace, N. (2002). Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park. *Appl Environ Microbiol*. 68, 5123-5135. doi: 10.1128/AEM.68.10.5123-5135.2002

Blumer-Schuette, S., Kataeva, I., Westpheling, J., Adams, M.W., and Kelly, R.M. (2008). Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr. Opin. Biotechnol*. 19, 210–217. doi: 10.1016/j.copbio.2008.04.007

Cabrera, E., Muñoz, M.J., Martín, R., Caro, I., Curbelo, C., and Díaz, A.B. (2014). Alkaline and alkaline peroxide pretreatments at mild temperature to enhance enzymatic hydrolysis of rice hulls and straw. *Bioresour. Technol*. 167, 1–7. Doi: 10.1016/j.biortech.2014.05.103

Chen, Z., Lin, H., Huang, W., Hsuan, S., Lin, J., and Wang, J. (2018). Molecular cloning, expression, and functional characterization of the β-agarase AgaB-4 from *Paenibacillus agarexedens*. *AMB Express*. 8, 49-10. doi: 10.1186/s13568-018-0581-8

Chen, Z., Pereira, J.H., Liu, H., Tran, H.M., Hsu, N.S., Dibble, D., et al. (2013). Improved activity of a thermophilic cellulase, Cel5A, from *Thermotoga maritima* on ionic liquid pretreated switchgrass. *PLOS ONE*. 8, e79725. doi: 10.1371/journal.pone.0079725

Crosby, J.R., Laemthong, T., Bing, R.G., Zhang, K., Tanwee, T.N.N., Lipscomb, G.L., et al. (2022). Biochemical and regulatory analyses of xylanolytic regulons in

Caldicellulosiruptor bescii reveal genus-wide features of hemicellulose utilization. *Appl. Environ. Microbiol.* 88, e0130222. doi: 10.1128/aem.01302-22

Cunha, J.T., Soares, P.O., Baptista, S.L., Costa, C.E., and Domingues, L. (2020). Engineered *Saccharomyces cerevisiae* for lignocellulosic valorization: a review and perspectives on bioethanol production. *Bioengineered.* 11, 883–903. doi: 10.1080/21655979.2020.1801178

Do Vale, L.H.F., Filho, E.X.F., Miller, R.N.G., Ricart, C.A.O., and Sousa, M.V. (2014). "Chapter 16 - cellulase systems in *Trichoderma*: an overview," in *Biotechnology and Biology of Trichoderma,* ed V.K. Gupta (Oxford: Elsevier), 229-244. doi: 10.1016/B978-0-444-59576-8.00016-3

Drula, E., Garron, M-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* 50, 571–D577. doi: 10.1093/nar/gkab1045

Du, Z., Wang, J., Yang, L., and Chen, G. (2011). Identification of a marine agarolytic bacterium *Agarivorans albus* QM38 and cloning and sequencing its beta-agarase genes. *Acta Oceanol.* 30, 118–124. doi: 10.1007/s13131-011-0098-3

Fu, X., Lin, H., and Kim, S. (2008). Purification and characterization of a novel β-agarase, AgaA34, from *Agarivorans albus* YKW-34. *Appl. Microbiol. Biotechnol.* 78, 265-273. doi: 10.1007/s00253-007-1303-3

Giles, K., Pluvinage, B., and Boraston, A.B. (2017). Structure of a glycoside hydrolase family 50 enzyme from a subfamily that is enriched in human gut microbiome *Bacteroidetes*. *Proteins: Struct. Funct. Genet.* 85, 182–187. doi: 10.1002/prot.25189

Guerriero, G., Hausman, J., Strauss, J., Ertan, H., and Siddiqui, K. (2016). Lignocellulosic biomass: biosynthesis, degradation, and industrial utilization. *Eng. Life Sci*, 16, 1-16. doi: 10.1002/elsc.201400196

Guindon, S., Dufayard, J-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95-98. doi:10.14601/Phytopathol_Mediterr-14998u1.29

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309–316. doi: 10.1042/bj2800309

Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino- acid sequence similarities. *Biochem. J.* 293, 781-788. doi: 10.1042/bj2930781

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* 316, 695–696. doi: 10.1042/bj3160695

Heredia-Ponce, Z., de Vicente, A., Cazorla, F.M., and Gutiérrez-Barranquero, J.A. (2021). Beyond the wall: exopolysaccharides in the biofilm lifestyle of pathogenic and beneficial plant-associated *Pseudomonas*. *Microorganisms (Basel)*. 9, 445. doi: 10.3390/microorganisms9020445

Hess, M., Sczyrba, A., Egan, R., Kim, T., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*. 331, 463-467. doi: 10.1126/science.1200387

Hsu, S.C., and Lockwood, J.L. (1975). Powdered chitin agar as a selective medium for enumeration of actinomycetes in water and soil. *Appl. Microbiol.* 29, 422–426. doi: 10.1128/AM.29.3.422-426.1975

Hu, Z., Lin, B.K., Xu, Y., Zhong, M.Q., and Liu, G.M. (2009). Production and purification of agarase from a marine agarolytic bacterium *Agarivorans sp.* HZ105. *J Appl Microbiol*. 106, 181-90. doi: 10.1111/j.1365-2672.2008.03990.x

Huber, R., Langworthy, T., König, H., Thomm, M., Woese, C., Sleytr, U., and Stetter, K. (1986). *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* 144, 324-333. doi: 10.1007/BF00409880

Jha, P., Pathan, M., Iyer, A., Jobby, R., and Desai, N. (2020). "Role of glycosyl hydrolases in breakdown of lignocellulosic waste and its industrial applications," in *Industrial Applications of Glycoside Hydrolases*, ed S. Shrivastava (Singapore: Springer), 131-149. doi: 10.1007/978-981-15-4767-6_4

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. 596, 583-589. doi: 10.1038/s41586-021-03819-2

Kato, S., Sakai, S., Hirai, M., Tasumi, E., Nishizawa, M., Suzuki, K., and Takai, K. (2018). Long-term cultivation and metagenomics reveal ecophysiology of previously uncultivated thermophiles involved in biogeochemical nitrogen cycle. *Microbes Environ*. 33, 107-110. doi: 10.1264/jsme2.ME17165

Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Kelly, R., Abelson, J.N., Simon, M.I., and Adams, M.W. (2001). *Hyperthermophilic Enzymes (Vol. 330)*. US: Elsevier Science & Technology, 11-24.

Kim J.J., Kwon Y-K., Kim J.H., Heo S-J, Lee Y., Lee S-J, et al. (2014). Effective microwell plate-based screening method for microbes producing cellulase and xylanase and its application. *J Microbiol Biotechnol*. 24, 1559-65. doi: 10.4014/jmb.1405.05052

Kim, H.T., Lee, S., Lee, D., Kim, H.S., Bang, W.G., Kim, K.H., and Choi, I.G. (2010). Overexpression and molecular characterization of Aga50D from *Saccharophagus degradans* 2-40: an exo-type beta-agarase producing neoagarobiose. *Appl Microbiol Biotechnol*. 86, 227-34. doi: 10.1007/s00253-009-2256-5

Kirimura, K., Masuda, N., Iwasaki, Y., Nakagawa, H., Kobayashi, R., and Usami, S. (1999). Purification and characterization of a novel beta-agarase from an alkalophilic bacterium, *Alteromonas* sp. E-1. *J Biosci Bioeng*. 87, 436-41. doi: 10.1016/s1389-1723(99)80091-7

Ko, H.J., Park, E., Song, J., Yang, T.H., Lee, H.J., Kim, K.H., and Choi, I.G. (2012). Functional cell surface display and controlled secretion of diverse agarolytic enzymes by *Escherichia coli* with a novel ligation-independent cloning vector based on the autotransporter YfaL. *Appl Environ Microbiol*. 78, 3051-8. doi: 10.1128/AEM.07004-11

Kukolya, J., Nagy, I., Laday, M., Toth, E., Oravecz, O., Marialigeti, K., and Hornok, L. (2002). *Thermobifida cellulolytica* sp. nov., a novel lignocellulose-decomposing actinomycete. *Int. J. Syst. Evol. Microbiol*. 52, 1193-1199. doi: 10.1099/ijs.0.01925-0

Kumar, R., Henrissat, B., and Coutinho, P. (2019b). Intrinsic dynamic behavior of enzyme:substrate complexes govern the catalytic action of β-galactosidases across clan GH-A. *Sci. Rep*. 9(1), 10346-14. doi: 10.1038/s41598-019-46589-8

Kumar, V., Hainaut, M., Delhomme, N., Mannapperuma, C., Immerzeel, P., Street, N., et al. (2019a). Poplar carbohydrate-active enzymes: Whole-genome annotation and functional analyses based on RNA expression data. *Plant J*. 99, 589-609. doi: 10.1099/ijs.0.01925-0

Lee, D., Park, G., Kim, N.Y., Lee, E., Jang, M.K., Shin, Y.G., et al. (2006). Cloning, expression, and characterization of a glycoside hydrolase family 50 [beta]-agarase from a marine *Agarivorans* isolate. *Biotechnol*. 28, 1925-32. doi: 10.1007/s10529-006-9171-y

Lee, D.G., Jang, M.K., Lee, O.H., Kim, N.Y., Ju, S.A., and Lee, S.H. (2008). Over-production of a glycoside hydrolase family 50 beta-agarase from *Agarivorans* sp. JA-1 in *Bacillus subtilis*

and the whitening effect of its product. *Biotechnol Lett*. 30, 911-8. doi: 10.1007/s10529-008-9634-4

Li, J., Hu, Q., Li, Y., and Xu, Y. (2015). Purification and characterization of cold-adapted beta-agarase from an Antarctic psychrophilic strain. *Braz J Microbiol*. 46, 683-90. doi: 10.1590/S1517-838246320131289

Liang, S.S., Chen, Y.P., Chen, Y.H., Chiu, S.H., and Liaw, L.L. (2014). Characterization and overexpression of a novel β-agarase from *Thalassomonas agarivorans*. *J Appl Microbiol*. 116, 563-72. doi: 10.1111/jam.12389

Liang, Y., Ma, X., Zhang, L., Li, F., Liu, Z., and Mao, X. (2017). Biochemical characterization and substrate degradation mode of a novel exotype β-agarase from *Agarivorans gilvus* WH0801. *J Agric Food Chem*. 65, 7982-7988. doi: 10.1021/acs.jafc.7b01533

Liao, L., Xu, X.W., Jiang, X.W., Cao, Y., Yi, N., Huo, Y.Y., et al. (2011). Cloning, expression, and characterization of a new beta-agarase from *Vibrio* sp. strain CN41. *Appl Environ Microbiol*. 77, 7077-9. doi: 10.1128/AEM.05364-11

Lin, B., Lu, G., Zheng, Y., Xie, W., Li, S., and Hu, Z. (2011). Gene cloning, expression and characterization of a neoagarotetraose-producing β-agarase from the marine bacterium *Agarivorans* sp. HZ105. *World J Microbiol Biotechnol*. 28, 1691-7. doi: 10.1007/s11274-011-0977-y

Liu, G., Shen, J., Chang, Y., Mei, X., Chen, G., Zhang, Y., and Xue, C. (2023). Characterization of an endo-1,3-fucanase from marine bacterium *Wenyingzhuangia aestuarii*: the first member of a novel glycoside hydrolase family GH174. *Carbohydr. Polym*. 306, 120591–120591. doi: 10.1016/j.carbpol.2023.120591

Liu, N., Mao, X., Du, Z., Mu, B., and Wei, D. (2012a). Cloning and characterisation of a novel neoagarotetraose-forming-β-agarase, AgWH50A from *Agarivorans gilvus* WH0801. *Carbohydr Res*. 388, 147-51. doi: 10.1016/j.carres.2014.02.019

Liu, N., Mao, X., Yang, M., Mu, B., and Wei, D. (2012b). Gene cloning, expression and characterisation of a new β-agarase, AgWH50C, producing neoagarobiose from *Agarivorans gilvus* WH0801. *World J Microbiol Biotechnol*. 30, 1691-8. doi: 10.1007/s11274-013-1591-y

Lykidis, A., Mavromatis, K., Ivanova, N., Anderson, I., Land, M., DiBartolo, G., et al. (2007). Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *J. Bacteriol.* 189, 2477-2486. doi: 10.1128/JB.01899-06

Matthews, C., Crispie, F., Lewis, E., Reid, M., O'Toole, P., and Cotter, P. (2019). The rumen microbiome: A crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency. *Gut Microbes.* 10, 115-132. doi: 10.1080/19490976.2018.1505176

McCarthy, T., Hanniffy, O., Savage, A., and Tuohy, M. (2003). Catalytic properties and mode of action of three endo-β-glucanases from *Talaromyces emersonii* on soluble β-1,4- and β-1,3;1,4-linked glucans. *Int. J. Biol. Macromol*. 33, 141-148. doi: 10.1016/S0141-8130(03)00080-1

Meng, D-D., Wang, B., Ma, X., Ji, S., Lu, M., and Li, F. (2016). Characterization of a thermostable endo-1,3(4)-β-glucanase from *Caldicellulosiruptor* sp. strain F32 and its application for yeast lysis. *Appl. Microbiol. Biotechnol*. 100, 4923-4934. doi: 10.1007/s00253-016-7334-x

Minh, B., Schmidt, H., Chernomor, O., Schrempf, D., Woodhams, M., Von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol*. 37, 1530-1534. doi: 10.1093/molbev/msaa015

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nat. Methods*. 19, 679-682. doi: 10.1038/s41592-022-01488-1

Mudgil, D. (2017). "Chapter 3 - the interaction between insoluble and soluble fiber," in *Dietary Fiber for the Prevention of Cardiovascular Disease*, ed R.A. Samaan (Elsevier Inc), 35-59. doi: 10.1016/B978-0-12-805130-6.00003-3

Nakajima, M., Yamashita, T., Takahashi, M., Nakano, and Y., Takeda, T. (2012). A novel glycosylphosphatidylinositol-anchored glycoside hydrolase from *Ustilago esculenta* functions in β-1,3-glucan degradation. *Appl Environ Microbiol*. 78, 5682-9. doi: 10.1128/AEM.00483-12.

Nikapitiya, C., Oh, C., Lee, Y-D., Lee, S-K., Whang, I-S., and Lee, J-H. (2010). Characterization of a glycoside hydrolase family 50 thermostable β-agarase AgrA from marine bacteria *Agarivorans* sp. AG17. *Fish. Aquat. Sci*. 13, 36–48. doi: 10.5657/FAS.2010.13.1.036

Nou, N.O. (2022) Taxonomic and physiological insights into novel bacterial class *Fervidibacteria*. Master's Thesis. Las Vegas (NV): University of Nevada, Las Vegas

Ochs, R. (2022). *Biochemistry (Second ed).* Boca Raton, FL: CRC Press, 125-126.

Ohta, Y., Hatada, Y., Ito, S., and Horikoshi, K. (2005). High-level expression of a neoagarobiose-producing beta-agarase gene from *Agarivorans* sp. JAMB-A11 in *Bacillus subtilis* and enzymic properties of the recombinant enzyme. *Biotechnol Appl Biochem.* 41, 183-91. doi: 10.1042/BA20040083

Paudel, D., Dhungana, B., Caffe, M., and Krishnan, P. (2021). A review of health-beneficial properties of oats. *Foods.* 10, 2591. doi: 10.3390/foods10112591

Payne, C., Resch, M., Chen, L., Crowley, M., Himmel, M., Taylor II, L., et al. (2013). Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proc. Natl. Acad. Sci. USA.* 110, 14646-14651. doi: 10.1073/pnas.1309106110

Peng, X., Su, H., Mi, S., and Han, Y. (2016). A multifunctional thermophilic glycoside hydrolase from *Caldicellulosiruptor owensensis* with potential applications in production of biofuels and biochemicals. *Biotechnol. Biofuels.* 9, 98. doi: 10.1186/s13068-016-0509-y

Peristiwati, Natamihardja, Y.S., and Herlini, H. (2018). Isolation and identification of cellulolytic bacteria from termites gut (*Cryptotermes* sp.). *J. Phys. Conf. Ser.* 1013, 12173. doi: 10.1088/1742-6596/1013/1/012173

Pettersen, E., Goddard, T., Huang, C., Meng, E., Couch, G., Croll, T., et al. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70-82. doi: 10.1002/pro.3943

Pluvinage, B., Hehemann, J-H., and Boraston, A.B. (2013). Substrate recognition and hydrolysis by a family 50 exo-β-agarase, Aga50D, from the marine bacterium *Saccharophagus degradans.* *J. Biol. Chem.* 288, 28078–28088. doi: 10.1074/jbc.M113.491068

Pluvinage, B., Robb, C.S., Jeffries, R., and Boraston, A.B. (2020). The structure of PfGH50B, an agarase from the marine bacterium *Pseudoalteromonas fuliginea* PS47. *Acta Crystallogr.* 76, 422–427. doi: 10.1107/S2053230X20010328

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLOS ONE.* 5, e9490–e9490. doi: 10.1371/journal.pone.0009490

Rastogi, G., Muppidi, G.L., Gurram, R.N., Adhikari, A., Bischoff, K.M., Hughes, S.R., et al. (2009). Isolation and characterization of cellulose-degrading bacteria from the deep subsurface of the Homestake gold mine, Lead, South Dakota, USA. *J. Ind. Microbiol. Biotechnol.* 36, 585–598. doi: 10.1007/s10295-009-0528-9

Ren, G., Cao, L., Kong, W., Wang, Z., and Liu, Y. (2016). Efficient secretion of the β-galactosidase Bgal1-3 via both tat-dependent and tat-independent pathways in *Bacillus subtilis*. *J. Agric. Food Chem.* 64, 5708-5716. doi: 10.1021/acs.jafc.6b01735

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N., Anderson, I., Cheng, J., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature.* 499, 431-437. doi: 10.1038/nature12352

Robert, C., Chassard, C., Lawson, P., and Bernalier-Donadille, A. (2007). *Bacteroides cellulosilyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int. J. Syst. Evol. Microbiol.* 57, 1516-1520. doi: 10.1099/ijs.0.64998-0

Rossi, F., and De Philippis, R. (2015). Role of cyanobacterial exopolysaccharides in phototrophic biofilms and in complex microbial mats. *Life (Basel, Switzerland).* 5, 1218-1238. doi: 10.3390/life5021218

Rozewicki, J., Li, S., Amada, K.M., Standley, D.M., and Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* 47, W5–W10. doi: 10.1093/nar/gkz342

Santos, C., Costa, P., Vieira, P., Gonzalez, S., Correa, T., Lima, E., et al. (2020). Structural insights into β-1,3-glucan cleavage by a glycoside hydrolase family. *Nat. Chem. Biol.* 16, 920-929. doi: 10.1038/s41589-020-0554-5

Schultz-Johansen, M., Bech, P.K., Hennessy, R.C., Glaring, M.A., Barbeyron, T., Czjzek, M., and Stougaard, P. (2018). A novel enzyme portfolio for red algal polysaccharide degradation in the marine *bacterium Paraglaciecola hydrolytica* S66T encoded in a sizeable polysaccharide utilization locus. *Front Microbiol.* 3, 839. doi: 10.3389/fmicb.2018.00839

Strazzulli, A., Fusco, S., Cobucci-Ponzano, B., Moracci, M., and Contursi, P. (2017). Metagenomics of microbial and viral life in terrestrial geothermal environments. *Rev. Environ. Sci. Biotechnol.* 16, 425–454. doi: 10.1007/s11157-017-9435-0

Sugano, Y., Matsumoto, T., Kodama, H., and Noma, M. (1993). Cloning and sequencing of agaA, a unique agarase 0107 gene from a marine bacterium, *Vibrio* sp. strain JT0107. *Appl. Environ. Microbiol.* 59, 3750–3756. doi: 10.1128/aem.59.11.3750-3756.1993

Summers, E., Moon, C., Atua, R., and Arcus, V. (2016). The structure of a glycoside hydrolase 29 family member from a rumen bacterium reveals unique, dual carbohydrate-binding domains. *Acta Crystallogr.* 72, 750-761. doi: 10.1107/S2053230X16014072

Tai, S-K., Lin, H-P.P., Kuo, J., and Liu, J-K. (2004). Isolation and characterization of a cellulolytic *Geobacillus thermoleovorans* T4 strain from sugar refinery wastewater. *Extremophiles.* 8, 345–349. doi: 10.1007/s00792-004-0395-2

Temuujin U., Chi W.J., Park J.S., Chang Y.K., Song J.Y., and Hong S.K. (2012a). Identification and characterization of a novel β-galactosidase from *Victivallis vadensis* ATCC BAA-548, an anaerobic fecal bacterium. *J Microbiol*. 50, 1034-40. doi: 10.1007/s12275-012-2478-6

Temuujin, U., Chi, W., Chang, Y., and Hong, S. (2012b). Identification and biochemical characterization of Sco3487 from *Streptomyces coelicolor* A3(2), an exo- and endo-type β-agarase-producing neoagarobiose. *J. Bacteriol. Res*. 194, 142-149. doi: 10.1128/JB.05978-11

Teufel, F., Almagro Armenteros, J., Johansen, A., Gíslason, M., Pihl, S., Tsirigos, K., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol*. 40, 1023-1025. doi: 10.1038/s41587-021-01156-3

Tiwari, P., Misra, B.N., and Sangwan, N.S. (2013). β-Glucosidases from the fungus *Trichoderma*: an efficient cellulase machinery in biotechnological applications. *Biomed Res. Int.* 2013, 203735–10. doi: 10.1155/2013/203735

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*. 50, D439-D444. doi: 10.1093/nar/gkab1061

Viborg, A.H., Terrapon, N., Lombard, V., Michel, G., Czjzek, M., Henrissat, B., and Brumer, H. (2019). A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J. Biol. Chem.* 294, 15973–15986. doi: 10.1074/jbc.RA119.010619

Vidal, R., and Venegas-Calerón, M. (2019). Simple, fast and accurate method for the determination of glycogen in the model unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *J. Microbiol. Methods*. 164, 105686. doi: 10.1016/j.mimet.2019.105686.

Wang, X-Y., and Fisher, P.B. (2015). "Chapter nine: scavenger receptors: emerging roles in cancer biology and immunology," in *Immunotherapy of Cancer,* ed P.B. Fisher (Cambridge, MA: Academic Press), 1-375.

Waterhouse, A., Procter, J., Martin, D., Clamp, M., and Barton, G. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 25, 1189-1191. doi: 10.1093/bioinformatics/btp033

Wertz, J-L., and Goffin, B. (2021). *Starch in the Bioeconomy (First edition).* Boca Raton, FL: CRC Press, Taylor and Francis Group, LLC, 1-18.

Wolfenden, R., and Snider, M. (2001). The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* 34, 938-945. doi: 10.1021/ar000058i

Yi, Y., Yan, Q., Jiang, Z., and Wang, L. (2018). A first glycoside hydrolase family 50 endo-β-1,3-d-glucanase from *Pseudomonas aeruginosa. Enzyme Microb. Technol.* 108, 34–41. doi: 10.1016/j.enzmictec.2017.09.002

Yoav, S., Barak, Y., Shamshoum, M., Borovok, I., Lamed, R., Dassa, B., et al. (2017). How does cellulosome composition influence deconstruction of lignocellulosic substrates in *Clostridium (Ruminiclostridium) thermocellum* DSM 1313? *Biotechnol.* 10, 222–222. doi: 10.1186/s13068-017-0909-7

Yun, E., Yu, S., Park, N., Cho, Y., Han, N., Jin, Y., and Kim, K. (2021). Metabolic and enzymatic elucidation of cooperative degradation of red seaweed agarose by two human gut bacteria. *Sci. Rep.* 11, 13955. doi: 10.1038/s41598-021-92872-y

Zakir, H.M., Hasan, M., Shahriar, S.M.S., Ara, T., M. Hossain, M. (2016). Production of biofuel from agricultural plant wastes: corn stover and sugarcane bagasse. *Chem. Eng. Sci.* 4, 5-11. doi:10.12691/ces-4-1-2

Zeng, Y., Himmel, M., and Ding, S. (2017). Visualizing chemical functionality in plant cell walls. *Biotechnol.* 10, 263. doi: 10.1186/s13068-017-0953-3

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, 95–101. doi: 10.1093/nar/gky418

Zhang, L., Yan, J., Fu, Z., Shi, W., Ninkuu, V., Li, G., et al. (2021). FoEG1, a secreted glycoside hydrolase family 12 protein from *Fusarium oxysporum*, triggers cell death and modulates plant immunity. *Mol. Plant Pathol.* 22, 522-538. doi: 10.1111/mpp.13041

Zhang, P., Zhang, J., Zhang, L., Sun, J., Li, Y., Wu, L., et al. (2019b). Structure-based design of agarase AgWH50C from *Agarivorans gilvus* WH0801 to enhance thermostability. *Appl. Microbiol. Biotechnol.* 103, 1289–1298. doi: 10.1007/s00253-018-9540-1

Zhang, Y., Fu, X., Duan, D., Xu, J., and Gao, X. (2019a). Preparation and characterization of agar, agarose, and agaropectin from the red alga *Ahnfeltia plicata. J. Oceanol. Limnol.* 37, 815–824. doi: 10.1007/s00343-019-8129-6

Zhou, Z., Liu, Y., Xu, W., Pan, J., Luo, Z-H., and Li, M. (2020). Genome- and community-level interaction insights into carbon utilization and element cycling functions of *Hydrothermarchaeota* in hydrothermal sediment. *mSystems*. 5. doi: 10.1128/mSystems.00795-19

CURRICULUM VITAE

Jonathan K. Covington

jonathankevincovington@gmail.com

https://orcid.org/0000-0002-8377-6722

Education

Master of Science – Biological Sciences (Ecology and Evolution)

University of Nevada, Las Vegas

May 2023

Bachelor of Science – Biological Sciences (Ecology and Evolution)

University of Nevada, Las Vegas

May 2020

Publications

Palmer, M., Covington, J.K., Zhou, E-M., Thomas, S., Habib, N., Seymour, C., et al. (2023). Thermophilic *Dehalococcoidia* with unusual traits shed light on an unexpected past. *ISME J.* doi: 10.21203/rs.3.rs-2143584/v1