Establishing Activity Based Protein Profiling as a robust Method for Bioprospecting of Active Hydrolases in (Hyper)thermophilic Environments

Dissertation

Zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften

- Dr. rer. nat. –

Vorgelegt von **Thomas Martin Klaus**

geboren in Wuppertal

Umweltmikrobiologie und Biotechnologie

Arbeitsgruppe Molekulare Enzymtechnologie und Biochemie

Fachbereich Chemie

Der Universität Duisburg-Essen

2022

DuEPublico
Duisburg-Essen Publications online

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

ub | universitäts bibliothek

Diese Dissertation wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt und liegt auch als Print-Version vor.

DOI: 10.17185/duepublico/78416
URN: urn:nbn:de:hbz:465-20250108-114252-7

Alle Rechte vorbehalten.

Die dieser Thesis zugrunde liegenden Arbeiten wurden im Zeitraum von Mai 2018 bis Dezember 2021 im Labor von Frau Prof. Dr. Bettina Siebers im Arbeitskreis für Molekulare Enzymtechnologie und Biochemie in der Abteilung für Umweltmikrobiologie und Biotechnologie an der Fakultät für Chemie der Universität Duisburg-Essen durchgeführt.

Tag der Disputation: 04.05.2023

Gutachter: Prof. Dr. Bettina Siebers

Prof. Dr. Markus Kaiser

Vorsitzender: Prof. Dr. Mathias Ropohl

„Es wird ja fleißig gearbeitet und viel mikroskopiert, aber es müsste mal wieder einer einen gescheiten Gedanken haben.“ - Rudolf Virchow

Table of content

1. Introduction

1.1. Thermophiles and other extremophiles

Several decades ago, environments providing harsh conditions which are not favorable for most vertebrates were barely thought to host living cells. This general adoption started to change drastically, no later than with the discovery of *Thermus aquaticus* in 1969 – an bacterium with an optimal growth temperature at 70-79 °C [1]. Since then, many different species thriving under extreme conditions, such as high or low temperatures and pH-values as well as high pressure, salt concentrations or radiation, have been described. Consequently, organisms which are able to live or prefer to live in these habitats, are classified as (hyper)thermophilic, psychrophilic, alkaliphilic, acidophilic, barophilic, halophilic, and radioresistant species. Notably, some species even show the ability to cope with multiple extreme conditions [2]. Most of these species belong to archaeal and bacterial phyla but there are also extremophilic eukaryotes and even some multicellular eukaryotes have prooven to survive under extreme conditions [3]. Interestingly, prokaryotic extremophiles are oftenly considered to have been the first representatives of life on earth, coinciding with the harsh conditions which were prevalent on earth and theories about origin of life around deep sea thermal vents – a theory which is still widely accepted [4]. However, in order to cope with extreme conditions, cells must posses molecular adaptations such as e.g. themostable and pH-stable proteins [5,6], altered membrane compositions [7], production of compatible solutes [8], adapted metabolic pathways - especially in archaea [9,10] and also ways to maintain a favorable pH value within the cytoplasma [11]. Because of these adaptations, study of extremophiles offers not only great potential for fundamental research about physiology, genetics and evolutions of microbes, but also for industrial applications [12,13]. Remarkably, the use of biocatalysts from thermophiles for biotechnological applications has clearly been demonstrated by one of the first ever isolated thermophilic proteins – the thermstable DNA polymerase from *Thermus aquaticus* [14], which soon led to a revolution of the PCR technique [15]. Thus, not suprisingly, many extensive studies on thermophiles and their habitats have been done over the last decades, continously broadening the knowledge about these extreme heat tolerant organisms and leading to the discovery of many more. Usually, organisms growing best at temperatures above 45 °C are classified as thermophiles and species growing best at temperatures higher than 80 °C are refered to as hyperthermophiles [16]. Whereas there are some examples for moderate thermophilic eukaryotes, such as fungi and algae [17,18], no fungi growing at temperatures significantly higher than 62 °C have been reported yet – A proposed reason for that is the inability to form thermostable

organellar membranes [19–21]. Thus, hyperthermophily is exclusively found in archaea and bacteria, yet the majority of hyperthermophilic species known today, belongs the the archaeal domain of life. Bacterial phylogenetic groups with representatives of hyperthermophiles are the Thermotogae, Aquificae and Thermodesulfobacteria. Archaeal phyla containing hyperthermophiles on the other hand are the Nanoarchaeota, Euryarchaeota, Thaumarchaeota and Crenarchaeota, as depicted schematically in figure 1. Interestingly, these phyla cover the deepest and shortest lineages within the phylogenetic tree, thus displaying hyperthermophiles as the most primitive existing organisms and giving rise to the hypothesis, that the last common ancester (LUCA) was a hyperthermophile [22–24]. This assumption is backened up by the theory, that early earth was a place with high temperatures and first live on earth developed in thermophilic environments [25]. While hyperthermophiles need at least a temperature of about 60 °C to grow, or >85 °C for *Pyrolobus fumarii* [26], growth optimum for hyperthermophilic species is usually seen between 80-110 °C. An upper limit ist currently set at 122 °C as reported for *Methanopyrus kandleri* [27], although mere survival without further growth was described at a temperature as high as 130 °C for 2 h for *Geogemma barossii* [28]. Noteworthy, it has to be differentiated between thermotolerant species, which can survive under high temperatures to some extend and true thermophilic species, which actually need elevated temperatures to flourish. Thermophiles and hyperthermophiles can be commonly found in a broad and diverse set of aquatic and terestrial habitats, such as hot springs [29], solfataras [30], deserts [31], biogas plants [32], compost [33], and hydrothermal vents [34]. Thus, microbes living in these environments are highly diverse but also show some specific thermophilic traits, such as the production of thermostable structures like proteins, membranes and other complex structures. Moreover, also their metabolism has to cope with challenges caused by these extreme conditions, such as low pathway intermediate stability. Commonly found strategies in thermophiles are e.g. the use of adapted pathway topologies, new pathways and lower concentration of metabolites with low thermostability [35].

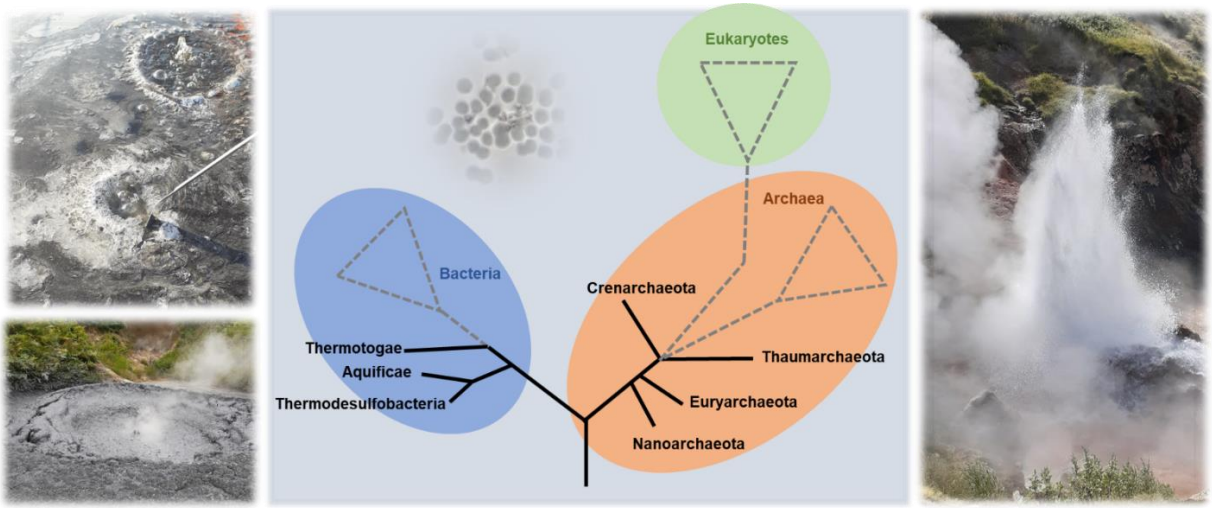


Figure 1. Hyperthermophilic origin/root of the two-domain tree of life. *A scheme depicting the two - domain of life tree with hyperthermophilic phyla (highlighted with bold balck lines) as well as mesophophilic phyla (dotted lines) is shown. The individual domains are marked Bacteria (blue), Archaea (orange) and Eukaryotes (green). Pictures show three hot springs in Kamtschatka (Russia).*

1.1.1. Archaeal unique features

As the most dominant representatives of extremophilic prokaryotes, the constantly diversifying and expanding domain of archaea plays an important role in diverse extreme environments, but also in mesophilic ecosystems [36,37] and even in the human microbiome [38]. However, since archaea and bacteria are relatively similiar in terms of size and shape, they have not been differentiated initially and archaea were defined as a seperate domain of life as late as 1977, when Carl Woese proposed a new three-domain phylogenetic tree, based on 16S/18S rRNA analysis [39]. Although there are different hypothesises about the origin of eukaryotes, phylogenomic data gathered over the last years widely favor a two domain tree of life model, in which the eukaryotes originate from the archaeal domain as a sister group from the Asgard Archaea [40,41]. The understanding of the archaeal domain itself has been expanding over the last years due to a constantly growing number of investigated and published genomes [36]. Thus, today Archaea are divided into the Euryarchaeal phylum and the Superphyla TACK (containing amongst others the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota) [42], DPANN (consisting of the phyla Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota as well as Woesearchaeota, Pacearchaeota and Altiarchaeota) [43–45] and the Asgard Archaea (containing the phyla Lokiarchaeota, Thorarchaeota, Odinarchaeota, and Heimdallarchaeota) [46], which are as beforementioned probably phylogenomically closely related to Eukaryotes [47]. However, there are several divergent phylogenetic models about the most accurate grouping within the Archaea. Thus for example, it has been proposed that the DPANN form a polyphyletic group within the

Euryarchaeota, instead of beeing an separate superphylum [48]. Archaea can be described as mosaic-like organisms, displaying features of both, bacteria and eukaryotes and some unique archaeal traits [49]. While the cell morphology, cell organisation and general structure of the DNA resembles those of bacteria, genetic information processing -transcription, and translation- is more similiar to Eukaryotes [50]. On the other hand, archaeal species display several features which are unique to archaea, such as their membrane composition. Contrarily to membranes in bacteria and eukaryotes, which have membranes consisting mainly of phospholipids with saturated or unsaturated fatty acids, that are linked to sn-glycerol-3-phosphate via ester-linkages, the archaeal membrane lipids consist of isoprenoide chains linked to sn- glycerol- 1- phosphat via ether-bonds [51]. Furthermore, Archaea lack the typical bacterial cellwall component murein, while some archaea have pseudomurein, N-acetylglucosamin, methanochondroitin, glutaminylglycan and sulfated heteropolysaccharides as cell wall components [52]. Most commonly, archaeal cells are surrounded by a paracrystalline protein surface layer, reffered to as S-layer [53]. Overall these archaeal cell wall features, enable higher thermal stability of the cell wall, due to a lower extend of segmental motion of the isoprenoid chains [54]. The physiology of mesophilic and thermophilic archaea shows likewise unique metabolic features and ways of sequestering nutrion and metals, that are neither present in bacteria, nor in eukarya [55,56]. Thus for instance, archaea utilize modified Emden-Meyerhof-Parnas and Entner-Doudoroff pathways for glycolysis, using unusual enzymes which are not present in bacteria or eukaryotes [9] such as adenosin-diphosphate (ADP)-dependent glucokinase [57], ADP-dependent phosphofructokinase [58] and a a glyceraldehyde 3-phosphate ferredoxin oxidoreductase catalyzing the oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate in a single step [59]. Further unusual metabolic traits can be found for example in gluconeogenesis where archaea employ a bi-functional class V fructose-1,6-bisphosphate aldolase/phosphatase additionally to a class I fructose-1,6-bisphosphate which is solely used in glycolysis [60]. Also, in most archaea, including the *Thermococcus sp.* 2319x1E strain described in 3.1, the classical pentose phosphate pathway, which is needed in bacteria and eukaryotes to maintain carbon homeostasis and providing precursors for nucleotide and amino acid biosynthesis, is absent or only partly present [61,62]. Yet except for these differences in basic metabolic processes, there are other unique physiological capabilities with a high impact on geothermal cycles, such as the ability to produce methane in order to conserve energy for the generation of adenosine triphosphate (ATP) [63].

1.1.2. Thermococcus sp. and their potential for biotechnology

Thermococcus represents a hyperthermophilic euryarchaeal genus within the order of *Thermococcales*, together with the phyla *Pyrococcus* and *Archaeococcus*. Noteworthy, especially *Pyrococcus* and *Thermococcus* species have been serving as model organisms for the study of microbial life at extremely high temperatures for decades now, since they can be easily grown on plain organic media to sufficient cell yields [64]. Members of the *Thermococcales* are usually obligate anaerobic heterotrophs which metabolize organic compounds such as pyruvate, amino acids, peptides, oligo- and polysaccharides, while terminal electrons from the respiratory chain are used for sulfur reduction or fermentative hydrogen production. [65,66]. Except of some *Thermococcales* species from freshwater springs, the majority of reported *Thermococcales* species originates from deep see hydrothermal vents where they are ubiquitiously present [67–69]. Today, there are more than 30 validly published *Thermococcus* genomes available (according to the KEGG database October 2022; https://www.genome.jp/kegg/), containing polyextremophilic or polyextremotolerant species such as *T. barophilus* which can also withstand high pressure or the radiation tolerant *T. radiotolerans* [70,71] . This high abundance of genomic information and also the availability of genetic tools, mainly developed for the extensively studied *Thermococcales* model organism *T. kodakarensis* [72], enables comprehensive studies on protein biochemistry and physiology of these cells for fundamental research and biotechnological applications. Several *Thermococcus* species assimilate and degrade different organic compound, especially sugars via unusual pathways or use hydrolases which are unusual for bacteria and eukarya for hydrolysis of peptides as well as oligo- and polysaccharides, such as the recently described *Thermococcus sp.* 2318x1 [73]. Since *Thermococcus* cells commonly thrive at temperatures between 50 °C to 100 °C, all of their cellular components like membrane lipids and proteins are adapted to extreme heat. Therefore, their enzymes are valuable biocatalysts for industrial processes and a variety of the involved enzymes, particularly glycoside hydrolases and peptidases have been heterologously expressed and were biochemically characterized [64], e.g. a GH57 family glycogen branching enzyme from *T. kodakarensis* KOD1 [74], a GH57 family amylopullulanase from *T. hydrothermalis* [75] and a pyrrolidone carboxylpeptidase from *T. litoralis* [76].

1.2. Industrial application of enzymes from (hyper)-thermophiles

The use of enzymes in various industrial applications is oftenly desired due to their ability to selectively and specifically catalyze different chemical reactions and has constantly been expanding for decades. In total there is a tremendous resoirvoir of enzymes in nature, especially in microbes, which is far from beeing completely examined – the entirety of enzymes in nature without assigned function is usually reffered to as *catalytic dark matter* [77]. Oftenly, the implementation of enzymes or even whole organisms into chemical, industrial processes reduces the amount of energy and chemicals needed, lowers the amount of by-products and is thus a cost-effective and sustainable alternative to processes which solely rely on synthetic chemicals [78]. Today, biocatalytic processes are commonly found in in food processing, in the production of platform and bulk chemicals, the production of finechemicals, in the pharmaceutical industry, in waste water treatment, in the paper industry and are used for the breakdown of biopolymers such as polysaccharides [79–82]. However, due to the harsh conditions prevailing during many of these processes, elevated enzyme stability towards high temperatures, denaturing agents and solvents is favoreds, thus making enzymes from (hyper)-thermophiles preferable biocatalysts. Since the implementation of the heatstable DNA-polymerase from *Thermus aquaticus* for routinous application in PCR, numerous applications of other enzymes from thermophiles have found their way into biotechnological use suchs e.g. the l-aminoacylase from *Thermococcus litoralis* for the efficient hydrolysis of different amino acid side chains at 85 °C [83]. Especially other thermostabile hydrolases for the processing of starch, xylan, cyclodextrins and other polycarbohydrates have been exploited from thermophiles like *Geobacillus stearothermophilus* [84]. Driven by the high impact of the topic of protein stability at high temperatures during industrial and biotechnological processes, scientists have been trying to answer the question about how this high degree of stability is achieved on a molecular level, resulting in the finding that there is no general strategy but several combined factors [85]. First there are some structural features contributing to higher thermal protein stability, such as a higher degree of salt-bridge networks and ion pairing, more hydrophobic and aromatic interactions, less and smaller surface loops, different packing and reduction of solvent exposed surface area and variations in the α-helix and β-sheet content. Moreover, a different amino acid composition, featuring less thermolabile residues, such as asparagine and cysteine and also the binding of metal ions and substrates can stabilize the protein structure [85]. Yet interestingly, amino acid sequences of (hyper)thermophilic enzymes are oftenly quite similiar to their homologous, mesophilic counterparts and their structures are

superimposable, despite distinctly different optimal reaction temperatur and thermostability [86]. Besides these structural, protein specific factors, proteins in the cytosol of (hyper)thermophilic cells can furthermore be protected and stabilized by the chemical environment consisting of salts and organic compunds such as metabolites and compatible solutes (e.g. mannosyl-glycerate and trehalose) [87,88]. Also, biophysical effects like macromolecular crowding contribute to thermostability *in vivo*, which are oftenly not possible to be properly reproduced *in vitro* [89]. Arising from the understanding of factors that contribute to thermostability of proteins, strategies for the targeted design of thermostable proteins have been established. For example site directed mutagenesis and directed evolution approaches were applied succesfully to reduce configurational unfolding entropy, thus enhancing protein stability [90–92]. Although techniques for the rational (*de novo*) design of proteins are constantly developing, bioprospecting of natural resources offers great promise due to the indetification of novel, thermostable enzymes with so far unkown biocatalytic properties.

1.2.1. Glycoside hydrolases

Glycoside hydrolases (GHs, sometimes also refered to as glycosidases or glycosyl hydrolases) are a class of enzymes which catalyze the hydrolysis of glycosidic bonds in glycosides, resulting in the formation of a sugar hemiacetal or hemiketal and a free aglycon (Fig. 2A). Together with glycosyl transferases (GTs) they are the key enzymes for the synthesis and degradation of glycosidic bonds in nature, as prevalent in biopolymers like cellulose or hemicelluloses. Hence, GHs are of high interest for biotechnological use in biorefining of natural resources. They are grouped together with other enzyme classes, such as polysaccharide lyases, carbohydrate esterases and auxiliary activities as carbohydrate active enzymes (CAZYmes). Classification of GHs is usually done based on several categories, including the catalytic mechanism (Fig. 2B), if they are endo- or exo-acting (Fig. 2C), based on the substrate beeing hydrolyzed and using sequence similarity. From a mechanistic point of view, there are GHs catalyzing hydrolysis reactions which end up with a change of anomeric conformation in the resulting hemiacetal on the one side (Fig. 2B I) and retaining GHs keeping the anomeric conformation upon hydrolysis (Fig. 2B II), which is achieved by a two-step reaction. For most GHs, this reaction mechanism - either inverting or retaining- involves two amino acid residues, one acting as a proton donor for the reaction, the other as a nucleophile/base as depicted in figure 3A I and 3A II [93]. Some GHs conduct the cleavage of complex sugars on the terminus of a carbohydrate chain, which is usually the reducing sugar end (*exo acting*) while other are able to hydrolyse glycosidic bonds in the middle of a polysaccharide (*endo acting*), such as cellulases (Fig. 2C) [93]. However, the

most significant classifications are the assignment to Enzyme Commision numbers (EC numbers) representing biochemically proven enzyme catalyzed reactions with defined substrates [94] and the sequence similiarity derived grouping into GH families. The EC-system is commonly used for all enzymes and those hydrolysing O- and S-glycosyl compounds (GHs) are assigned mostly with the numbers starting with EC 3.2.1.X with more than 200 different entries reported in the CAZY-database today (October 2022) (http://www.cazy.org/Glycoside-Hydrolases.html). The classification based on amino acid sequence and folding similarities was proposed by Henrissat and others about thirty years ago and has since then continously been refined [95], resulting in more than 170 defined GH families today, which are subdivided into 18 clans (GH-A to GH-M) with individual structural characteristics. This sequence based approach does not only reflect structural and mechanistic features of certain hydrolases and helps to display evolutionary relationships between these proteins, but also illustrates that substrate specificity does not necessarily correlates with family membership [95,96].

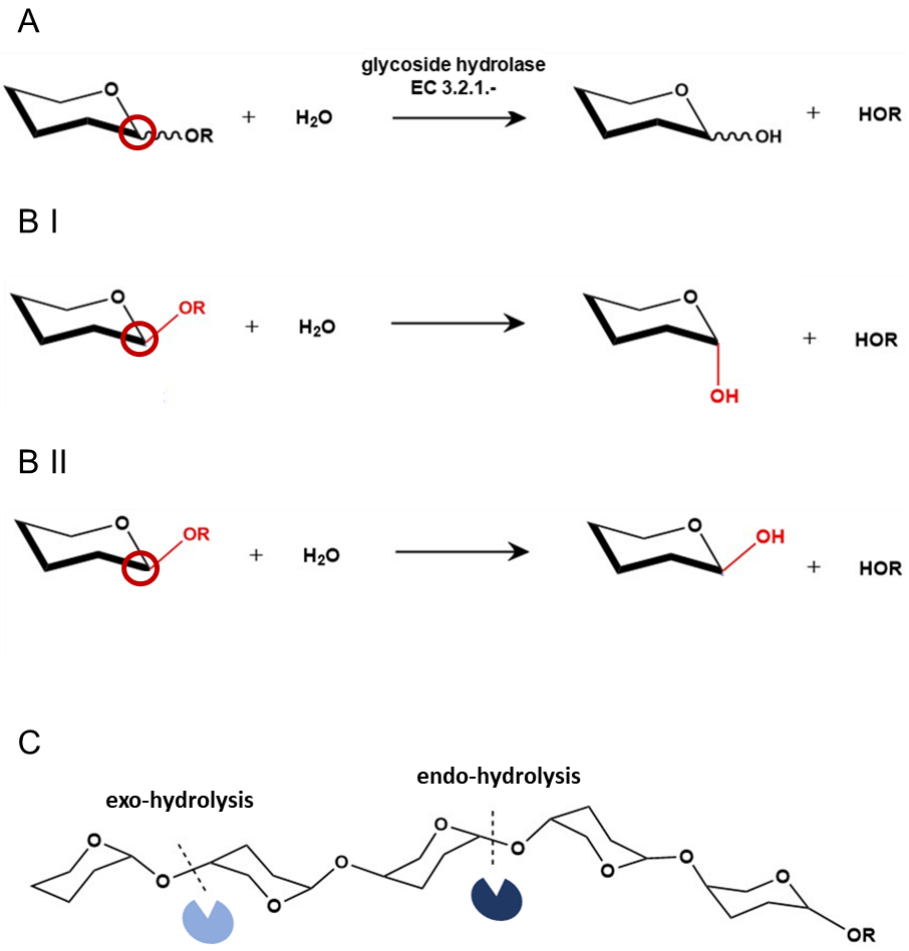


Figure 2: Classification of glycoside hydrolases. ***(A)*** *Glycoside hydrolases (EC 3.2.1.-) catalyze hydrolytic cleavage of glycosidic bonds either in a inverting* ***(B I)*** *or a retaining manner* ***(B II)****, resulting either in keeping the configuration of the anomeric carbon atom (red circle)****. (C)*** *Glycoside hydrolases are either exo- acting (bright blue) or endo-acting (dark blue). The mechanism of inverting hydrolases.*

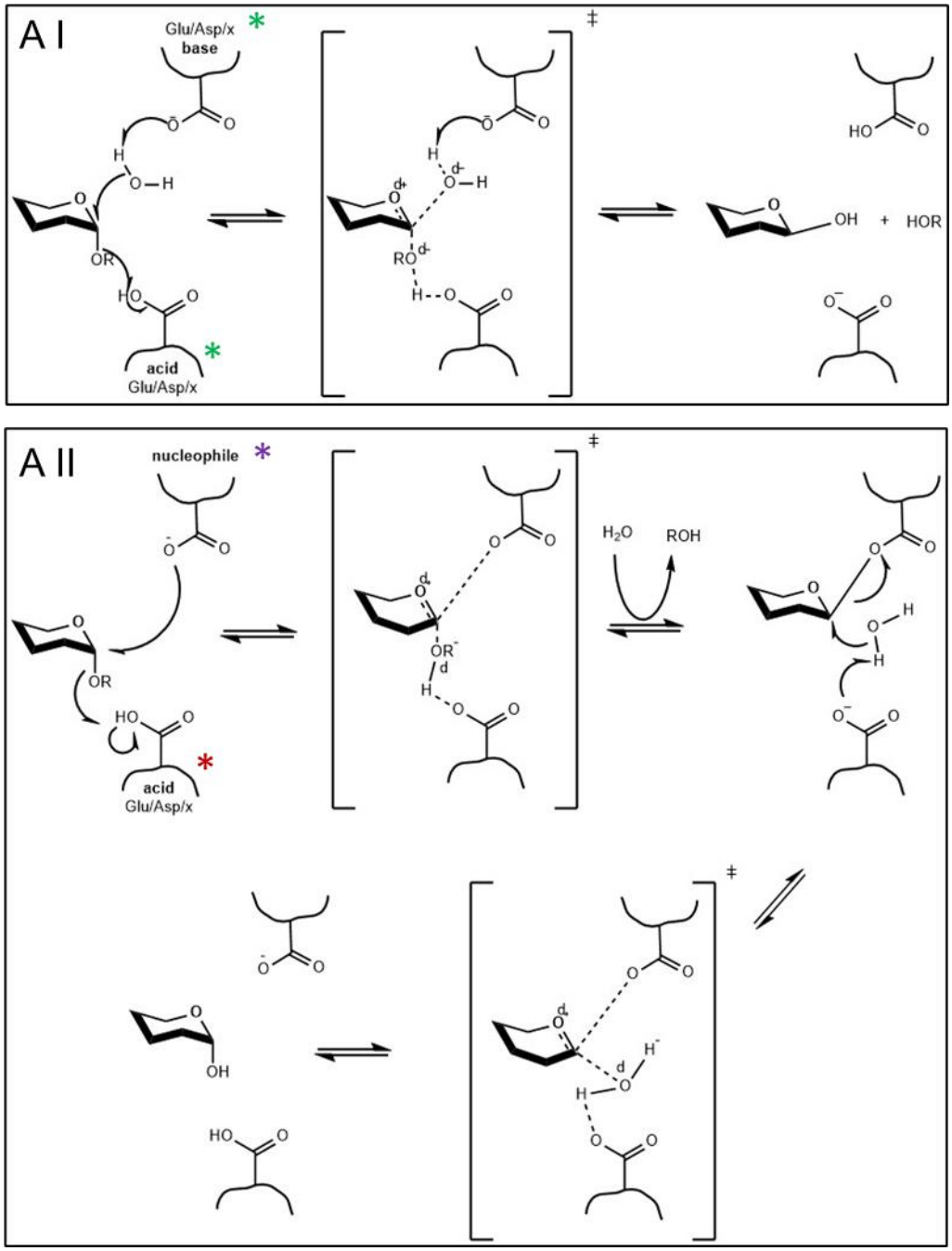


Figure 3: Reaction mechanisms of glycoside hydrolases. *The mechanism of inverting hydrolases* ***(A I)*** *runs via one intermediate state and the hydrolysis reaction is catalyzed by two amino acid residues (mostly Glu or Asp) serving as acid and base (green stars). The mechanism of retaining GHs contains two intermediate states* ***(A II)*** *and is catalyzed by a general nucleophile (purple star) and an acidic residue like Glu or Asp (red star).*

Naturally occuring GHs are not only used to degrade poly- and oligomeric substrates to gain influx for the energy metabolism of the cell, but also for the directed synthesis of secondary metabolites with special functions such as described for the synthesis of the disaccharide trehalose (α-D-glucopyranosyl-1,1-α-D-glucopyranoside), which serves as an osmoprotectant e.g. in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius* [97]. Because naturally

occuring polysaccharides like starch, cellulose and hemicelluloses represent highly abundant resources which also oftenly accumulate as waste products from industrial processes [98], GHs are valuable biocatalysts to transform these substrates into monosaccharides, which can then be furtherly processed to products like biofuels and platform chemicals. Due to harsh conditions during industrial biomass processing and high temperature pretreatment, GHs from (hyper)thermophiles are most suitable candidates for these applications and moreover for the application in bioanalytical processes as well as in the food-, feed- and textile industry [99].

1.2.2. Serine hydrolases

Serine hydrolases are one of the most diverse enzyme families and are commonly found with lots of different functions in higher eukaryotes but also within the microbial world. Their functions among others include proteases, peptidases, lipases, esterases, amidases and therefore plays important roles throughout nature [100,101]. Thus, not suprisingly extensive characterization and chemical biology profiling methods have been established e.g. to identify potential drug targets in humans [102,103]. Due to this pharmacological relevance of serine hydrolases, a comprehensive repertoir of chemical probes, targeting selectively different serine hydrolase functionalities, has been synthesized and tested for *in vitro* and *in vivo* labelling. Hydrolysis of various substrates by serine hydrolases is achieved by a highly nucleophilic serine residue, that is part of a catylytic dyade (Ser-Asp/ Ser-Lys/ Ser-His) or triade (most commonly Ser-His-Asp and several less widespread variations) [104–106].

1.3. Bioprospecting of hot springs

The entire microbial diversity of (hyper)thermophiles and other extremophiles offers a huge reservoir of potentially interesting hydrolases for biotechnological purposes, which can hardly be approached and utilized by conventional, culture dependent methods. Hence, with continously improving sequencing, spectroscopic and bioinformatik techniques more microbial genomes are beeing published and sequence-based metagenomic screenings from microbial communities are carried out to acces the previously hidden diversity [107,108]. However, prospecting for novel enzymes by sequence-based analysis alone is restricted, since gene sequences constructed from metagenomic data are annotated based on previously annotated sequences. Thus, only enzymes with conserved sequences are routinously beeing detected with this procedure [109,110]. Therefore, for the discovery of novel enzymes with certain biocatalytic specificities, functional metagenomics represents a more determined approach. This screening commonly starts with the isolation of DNA from microbial samples, construction of a metagenomic library, subsequent cloning of obtained DNA-fragments into appropiate vectors

and heterologous expression in suitable host organisms [111]. In the past, several glycosyl hydrolases with decent activity on poly- and oligosaccharides have been identified from different mesophilic nd thermophilic metagenoms [112–114]. Still, the sufficient expression of active enzymes appears as a bottleneck for the outcome of this screening technique, due to unapropiate codon usage in the expression host, wrong post translational processing, missing adequat ribosomal binding sites or other factors [115,116]. Moreover, most commonly, mesophilic hosts, e.g. *E. coli*, are beeing used for the heterologous expression of genes. Hence many hyperthermophilic proteins are not detected, since they sometimes need special adaptations in order to sustain their naturally function [117]. Another strategy to narrow down the complexity of microbial communities to communities with desired features is the use of enrichment cultures, which is based on the targeted cultivation of environmental samples with specific substrates to preferably enrich the culture with microorganisms which are capable of degrading the applied substrate [118,119]. Recently, a novel xylanolytic *Thermococcus* strain, beeing reported as the first xylanolytic hyperthermophilic euryarchaeon, was isolated with this approach performed *in situ* in a hot vent [73]. In the genome of this strain, namely *Thermococcus sp.* 2391x1, an unusual multidomain glycosidase, consisting of the domains GH5-12-12-CBM2-2 was identified, heterologously expressed and biochemically characterized, revealing β-1,3/β-1,4 glucosidase activity and minor xylanase activity. However, despite thorough genomic analysis, it remained unclear, which hydrolases are involved in xylan degradation in this strain. Yet another strain with xylanolytic activity was isolated from the same enrichment culture and was analyzed with respect to its glycoside hydrolases in in chapter 3.1 of this thesis, in which the establishment of activity based protein profiling (ABPP) as a screening method for the function specific detection of active hydrolases *in vivo* under (hyper)thermophilic conditions is described. Despite beeing able to grow on xylan or D-xylose as its sole carbon source, no common xylose degradation pathway such as the oxido-reductase pathway, the isomerase pathway, or the Weimberg/Dahms pathway could be identified in the genome of *Thermococcus sp.* 2391x1 or *Thermococcus sp.* 2391x1E. Initial investigations on enzymes putatively involved in d-xylose degradation in *Thermococcus sp.* 2391x1E comprised comparative proteomics from cells grown on different carbon sources, biochemical assays of cell crude extract and *in silico* methods. These results are presented in the appendix (IV Appendix).

1.3.1. Activity based protein profiling for the funtional identification of active hydrolases

As described previously, (meta) genomic analysis of microbial cultures and communities do not provide functional information and are also accompanied by other drawbacks. Thus, activity based protein profiling is a promissing technique to add functional information to genomic data in general. Furthermore, we demonstrate how this chemical profiling method can be used to gain physiologically and ecologically relevant information from microbes in hot environments.

ABPP is a chemical proteomic method, which aimes to function-specifically detect active proteins in their natural settings by using site directed probes and a subsequent proteomic workflow for identification and quantification of labelled proteins [120]. Generally, applied activity based probes (ABPs) consist of a war head, which is functionally derived from mechanism based inhibitors and covalently bind to the reactive residues of a targeted enzyme. The reactive warhead is furtherly connected via a linker region, which defines the probes specificity, to a reporter tag, which can be used to purify the labelled enzyme or to visualize it on a gel [121,122]. The choice of the reporter tag depends on the intended further processing of the samples – thus, there are common affinity tags such as biotin or for targeted purification, whereas fluorescent tags such as rhodamine are preferred for easy visualization e.g. on a agarose gel. The variety of the reactive warehead on the other hand is distinctly more diverse and highly depends on the targeted enzyme class and the desired specificity, since binding of the probes depends on mechanistic and structural active site features of the target. Therefore, this targeted design allows to adress even low-abundance proteins in rather complex proteome samples [122,123]. Most warheads, especially for serine hydrolases derive from electrophilic, mechanism-based irreversible inhibitors [124]. Thus, fluorophosphonate-biotin ist one of the most commonly applied ABPs for serine hydrolases and has already been widely applied, mainly for chemical proteomic analysis of human proteases [121]. Due to the bulky nature of common probe tags (0.7- 1.0 kDa), efficient probe uptake into the cell is oftenly hindered. Thus, tag-less probes offer better membrane permeability and thus higher *in vivo* labelling efficiancy. A reporter tag can then later be added by a bioorthogonal coupling reaction [125]. Most commonly, a $CuSO_4$ Huisgen type cycloaddition ist carried out in wich the alkine group of the probe reacts with the azide group of the added tag to form a triazol linker between the probe head and the tag [126].

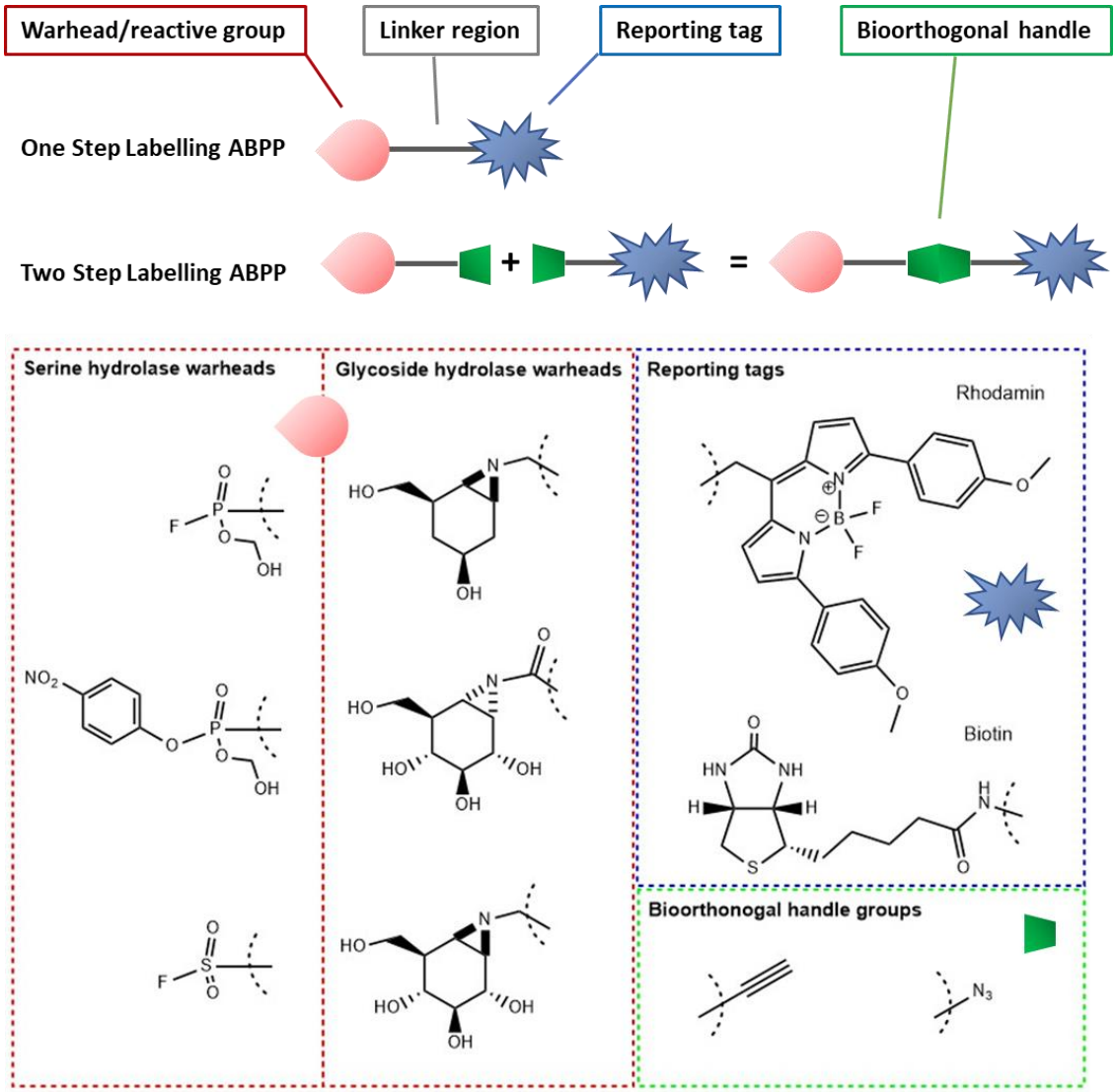


Figure 4: General design of Activity Based Probes (ABPs). *Every ABP consists of a reactive group (red drop) which is designed to bind to certain reactive aminoacids of a target protein. This group is also referred to as the ABP warhead. The applied probe for a one-step-labelling approach is connected to a reporting tag (blue star) by a linker region (black), which also contributes to the probes selectivity. The reporting tag is usually an affinity tag or a fluorophor and is beeing used for binding to a matrix or for visualization of the labelled protein.*

Combined with (meta)genomics, (meta)proteomics and biochemistry ABPP is a valuable and versatile method, which has already been applied in higher eukaryotes such as mamals, plants and also on mesophilic microorganisms to reveal information about activity state of certain enzymes under physiological conditions and enzyme affinity. In this thesis we have broadened the application of ABPP with a focus on bioprospecting of new enzymes with functions which cannot easily be derived from sequence alone. Contrarily to most previous works the here described combined workflows have been performed in (hyper)thermophilic systems such as *Thermococcus* cultures and thermal springs in Kamchatka, thus establishing this method for extreme conditions and complex environmental samples.

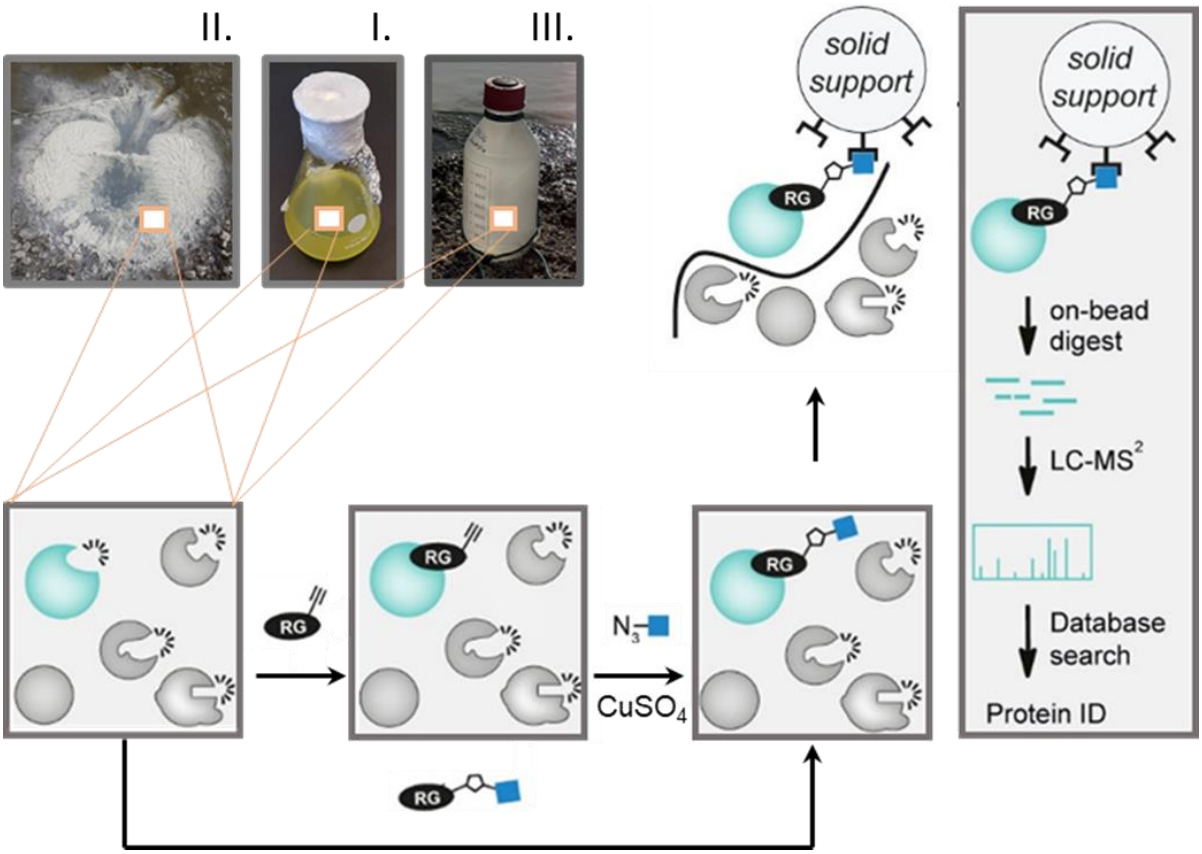


Figure 5: General workflow of ABPP experiments for the functional identification of active hydrolases in microbial cells. *ABP Labelling of pure microbial cultures (I.) for the identification of active enzymes has ben relatively well established in the recent years. ABPP studies on environmental samples (II.) and enrichment cultures (III.) on the other hand have been less often described. ABPP on the enzymes present in the sample material can be performed* in vivo *or after lysing the cells first to extract the proteome. Active enzymes can then be labelled in a one step approach, using an ABP in which the reactive group is already linked to a reporter group or in a two step approach, which can allow a higher labelling efficiancy. In the two step approach, the proteome is first labelled with a reactiv group connected to a bioorthogonal group, such as an alkine. In the second step, a reporter group e.g. linked to an azide group reacts with the other handle group. This is usually done via a copper catalyzed huisgen type cyclo addition. Labelled proteins are then furtherly purified using affinity purification. After on-bead digestion the proteins are analyzed by a LC-MS/MS method and identified using available sequence information. E.g. for eABPP a database needs to be created via metagenome analysis.*

2. Scope of the thesis

In this dissertation three chapters about the application of ABPP in thermophilic cells or environments are presented and my contribution to these manuscripts is described below.

In Chapter 3.1 the manuscript with the title „Activity-Based Protein Profiling for the Identification of Novel Carbohydrate-Active Enzymes Involved in Xylan Degradation in the Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1E“ is presented. In this manuscript, glycoside hydrolases which are active in Thermococcus sp. 2319x1E during growth on xylan were identified and characterized. Therefore, genomic analysis, comparativ proteomics and ABPP experiments were conducted and genes encoding proteins of interest were heterologously expressed, the proteins were purified and biochemicaly characterized. I performed all cultivation experiments of the cells, cloning, heterologous expression and purification of the proteins of interest and did the biochemical characterization of heterologously expressed enzymes and determined activities in cell lysates/crude extracts. Additionally, together with Dr. Sabrina Ninck, I conducted the *in vivo* chemical labelling with the applied ABPs, as well as subsequent protein extraction, affinity enrichment of the respective proteins and data evaluation.

In Chapter 3.2, “Function-based identification of novel enzymes from environmental microbial communities using activity-based protein profiling“ a proof of concept study for the application of ABPP on microbial mats in a hot spring was conducted. Thus, chemical labelling with an ABP designed for serine-hydrolases was done *in situ* and combined with thourough metagenomic analysis, microbial community analysis and heterologous expression of a selected target protein. For this project, I performed the in-field labelling experiments, subsequent protein extraction and affinity enrichment experiments. Furtherly, I contributed to the interpretion of the metagenom and metaproteome data and finally, I did the heterologous expression, and biochemical characterization of the selected serinehydrolase.

In the chapter 3.3, “Identification and characterization of a prevalent thermophilic β- glucosidase from a hot spring enrichment metagenome in Kamschatka identified with ABPP“, an enrichment culture with amorphous cellulose was examined with respect to its polysaccharolytic capacity. Therefore, metagenomic and metaproteomic analysis was combined with the ABPP workflow which was established in chapter 3.2., using a probe targeting preferably β-glucosidases. In the end, the best labelled β-glucosidase was heterologously expressed and characterized as an highly thermostable glycosidase with

interesting features for industrial application. For this manuscript, I prepared the enrichment culture and subsequently performed the *in situ* chemical labelling experiments. Furtherly, I performed protein extraction and affinity enrichment experiments in collaboration with Dr. Sabrina Ninck. Additionally, I conducted the heterologous expression and thorough biochemical characterization of the selected glucosidase.

Chapter 3.1

Activity-Based Protein Profiling for the Identification of Novel Carbohydrate-Active Enzymes Involved in Xylan Degradation in the Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1E

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fmicb.2021.734039

Activity-Based Protein Profiling for the Identification of Novel Carbohydrate-Active Enzymes Involved in Xylan Degradation in the Hyperthermophilic Euryarchaeon *Thermococcus* sp. Strain 2319x1E

Thomas Klaus[1†], Sabrina Ninck[2†], Andreas Albersmeier[3], Tobias Busche[3], Daniel Wibberg[3], Jianbing Jiang[4‡], Alexander G. Elcheninov[5], Kseniya S. Zayulina[5], Farnusch Kaschani[2], Christopher Bräsen[1], Herman S. Overkleeft[4], Jörn Kalinowski[3], Ilya V. Kublanov[5], Markus Kaiser[2] and Bettina Siebers[1]**

[1] Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Faculty of Chemistry, Centre for Water and Environmental Research (CWE), University of Duisburg-Essen, Essen, Germany, [2] Department of Chemical Biology, Center of Medical Biotechnology, Faculty of Biology, University of Duisburg-Essen, Essen, Germany, [3] Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany, [4] Section of Bio-Organic Synthesis, Leiden Institute of Chemistry, Leiden University, Leiden, Netherlands, [5] Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia

Activity-based protein profiling (ABPP) has so far scarcely been applied in Archaea in general and, especially, in extremophilic organisms. We herein isolated a novel *Thermococcus* strain designated sp. strain 2319x1E derived from the same enrichment culture as the recently reported *Thermococcus* sp. strain 2319x1. Both strains are able to grow with xylan as the sole carbon and energy source, and for *Thermococcus* sp. strain 2319x1E (optimal growth at 85°C, pH 6–7), the induction of xylanolytic activity in the presence of xylan was demonstrated. Since the solely sequence-based identification of xylanolytic enzymes is hardly possible, we established a complementary approach by conducting comparative full proteome analysis in combination with ABPP using α- or β-glycosidase selective probes and subsequent mass spectrometry (MS)-based analysis. This complementary proteomics approach in combination with recombinant protein expression and classical enzyme characterization enabled the identification of a novel bifunctional maltose-forming α-amylase and deacetylase (EGDIFPOO_00674) belonging to the GH57 family and a promiscuous β-glycosidase (EGIDFPOO_00532) with β-xylosidase activity. We thereby further substantiated the general applicability of ABPP in archaea and expanded the ABPP repertoire for the identification of glycoside hydrolases in hyperthermophiles.

Keywords: activity-based protein profiling, archaea, *Thermococcus*, xylan, hemicellulose degradation, glycoside hydrolases

OPEN ACCESS

Edited by:
Mirko Basen,
Universität Rostock, Germany

Reviewed by:
Haruyuki Atomi,
Kyoto University, Japan
Roderick Ian Mackie,
University of Illinois at Urbana-Champaign, United States

***Correspondence:**
Markus Kaiser
markus.kaiser@uni-due.de
Bettina Siebers
bettina.siebers@uni-due.de

[†] These authors have contributed equally to this work

[‡] Present address:
Jianbing Jiang,
Health Science Center, School of Pharmacy, Shenzhen University, Shenzhen, China

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *30 June 2021*
Accepted: *22 November 2021*
Published: *12 January 2022*

Citation:
Klaus T, Ninck S, Albersmeier A, Busche T, Wibberg D, Jiang J, Elcheninov AG, Zayulina KS, Kaschani F, Bräsen C, Overkleeft HS, Kalinowski J, Kublanov IV, Kaiser M and Siebers B (2022) Activity-Based Protein Profiling for the Identification of Novel Carbohydrate-Active Enzymes Involved in Xylan Degradation in the Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1E. Front. Microbiol. 12:734039. doi: 10.3389/fmicb.2021.734039

INTRODUCTION

Activity-based protein profiling (ABPP) is a powerful technique for the class-specific detection of active enzymes. With the help of an activity-based probe (ABP) that is composed of a reactive inhibitor group, a linker and a reporter group, a target enzyme is covalently modified at the active site and can thereby be detected or identified in downstream analyses (Cravatt et al., 2008). ABPP has been proven as an elaborate tool for several mesophilic enzyme families, allowing to unravel the activity state of enzymes under different conditions without knowledge of their natural substrates or enzyme functions, and thus also helps to deduce their function in certain cellular processes (Cravatt et al., 2008; Willems et al., 2014). This methodology can be applied to various types of biological samples and is even suitable for the *in vivo* study of enzyme activities under native conditions (Speers and Cravatt, 2004).

Recently, its scope of application has been expanded to the identification of serine hydrolases in extremophilic Archaea using phosphonate-derived ABPs (Zweerink et al., 2017). Labeling of α- and β-specific retaining glycoside hydrolases, on the other hand, has been successfully conducted in different Eukaryotes (like mammals, fungi and plants) by employing ABPs, which were designed to react with the active site nucleophiles of retaining glycosidases to form a covalent and irreversible bond (Wu et al., 2019). The two ABPs used in this work, JJB384 and JJB111, are based on two epimeric cyclitol aziridines. JJB384 is composed of 1,6-epi-cyclophellitol aziridine having a biotin moiety attached to the aziridine nitrogen (N_2). It is a structural isostere of an α-glucopyranoside warhead and thanks to this feature preferably targets retaining α-glucosidases (Jiang et al., 2016). JJB111, the biotinylated, *N*-alkylated aziridine analog of the natural product cyclophellitol, structurally and configurationally resembles β-glucopyranosides and preferentially reacts with β-glucosidases (Kallemeijn et al., 2012). However, neither of the two ABPs are fully in-class selective. JJB384 was found to label, besides α-glucosidases, also several β-glucosidases (Husaini et al., 2018). JJB111 in turn showed activity toward a variety of β-glycosidases, including β-xylosidases such as a bifunctional α-L-arabinofuranosidase/β-D-xylosidase, β-galactosidases, and β-glucuronidases (Chandrasekar et al., 2014; Husaini et al., 2018).

The heteromorphic polysaccharide xylan belongs to the hemicelluloses and, together with other polymers such as cellulose and lignin, is a major constituent of the plant cell wall (Ebringerová and Heinze, 2000). It is therefore one of the most abundant polysaccharides found in nature, and moreover, it is believed to account for about one third of the Earth's renewable organic carbon (Prade, 1996). In plant cell walls, xylans and other hemicelluloses, such as mannans and galactans, are covalently linked to lignin layers and non-covalently to cellulose fibers (Biely, 1985; Heredia et al., 1995). Xylan structures in plants are highly variable and depend on the phylogenetic position of the species and on their position in primary or secondary cell walls (Harris and Stone, 2008). Many xylans are thereby composed of a linear backbone of 1,4-linked β-D-xylopyranosyl residues. In cell walls, they form heteroxylans with side chains containing other monosaccharides, such as α-L-arabinofuranose and 4-*O*-methyl-α-D-glucuronic acid or α-D-glucuronic acid, as well as oligosaccharides, acetyl groups and, in some taxa, phenolic acid esters, such as ferulate or *p*-coumarate esters (Peña et al., 2016). In hardwood xylan, for instance, some D-xylose molecules are linked to 4-*O*-methylglucuronic acids at the C2 position and are highly acetylated at the C2 and C3 positions. Xylans from softwoods have an even higher content of 4-*O*-methylglucuronic acids, but instead of acetylation, α-L-arabinofuranose units are linked to C3 atoms of the D-xylose backbone (Puls, 1997). For algae and seaweed species, xylans have only been found in small amounts and feature a different structure: β-1,3-linked D-xylose backbones or mixtures of β-1,3 and β-1,4 bonds with diverse side chains (Usov, 2011; Hsieh and Harris, 2019).

Due to the complex and heterogeneous structure of xylan, the complete enzymatic hydrolysis into its monomeric sugars requires different classes of hydrolases such as endo-β-1,4-xylanases (EC 3.2.1.8), β-xylosidases (EC 3.2.1.37), α-glucuronidases (EC 3.2.1.139), α-arabinofuranosidases (EC 3.2.1.55) and furthermore acetylxylan esterases (EC 3.1.1.6) for deacetylation of the xylan backbone and released sugar moieties (**Figure 1**). Endo-β-1,4-xylanases and β-xylosidases are collectively referred to as xylanases since they are required for hydrolysis of the xylan backbone into D-xylose monomers (Juturu and Wu, 2012). According to the CAZy database (see the Carbohydrate Active Enzymes database[1]; Henrissat et al., 1989; Lombard et al., 2014), all so far reported endoxylanases and β-xylosidases are grouped into 9 and 14 different GH families, respectively. However, most of them originate from fungi or bacteria, while only a few archaeal (hyper)thermophilic xylanases are known so far, that mostly belong to the GH10 and GH11 families (Collins et al., 2005; Thomas et al., 2017). The presence of xylanolytic enzymes has been demonstrated only for a few representatives of the Crenarchaeota and Euryarchaeota, however, examples for a proper characterization of xylan degrading enzymes from those phyla remain scarce (Niehaus et al., 1999; Uhl and Daniel, 1999; Cannio et al., 2004; Collins et al., 2005). In *Thermococcus zilligii,* a xylanolytic enzyme was purified from the culture supernatant and described as the first archaeal hemicellulase discovered, with an activity of 5.07 U mg^{-1} protein using oat spelt xylan as substrate (Uhl and Daniel, 1999). However, *T. zilligii* is not able to grow on xylan. Hence, the recently described Archaeon *Thermococcus* sp. strain 2319x1 (Gavrilov et al., 2016) and a new *Thermococcus* isolate designated sp. strain 2319x1E described in this work, both originating from the same enrichment culture, are the first Euryarchaeota known to be able to utilize xylan as sole carbon and energy source. We thus applied (chemo)proteomics,

Abbreviations: ABP, activity-based probe; ABPP, activity-based protein profiling; CAZymes, carbohydrate-active enzymes; CMC, carboxymethyl cellulose; DAPI, 4′,6-diamidino-2-phenylindole; DNSA, 3,5-dinitrosalicylic acid; HRP, horseradish peroxidase; IPTG, isopropyl-β-D-thiogalactopyranosid; LB, lysogeny broth; LFQ, label-free quantification; MβCD, 6-*O*-α-maltosyl-β-cyclodextrin; MS, mass spectrometry; ONPG, *ortho*-nitrophenyl-β-D-galactopyranoside; PBS, phosphate buffered saline; PNP, *para*-nitrophenol; PNPA, *para*-nitrophenyl-acetate; PNPG, *para*-nitrophenyl-β-D-glucopyranoside; PNPX, *para*-nitrophenyl-β-D-xylopyranoside.

[1]http://www.cazy.org/

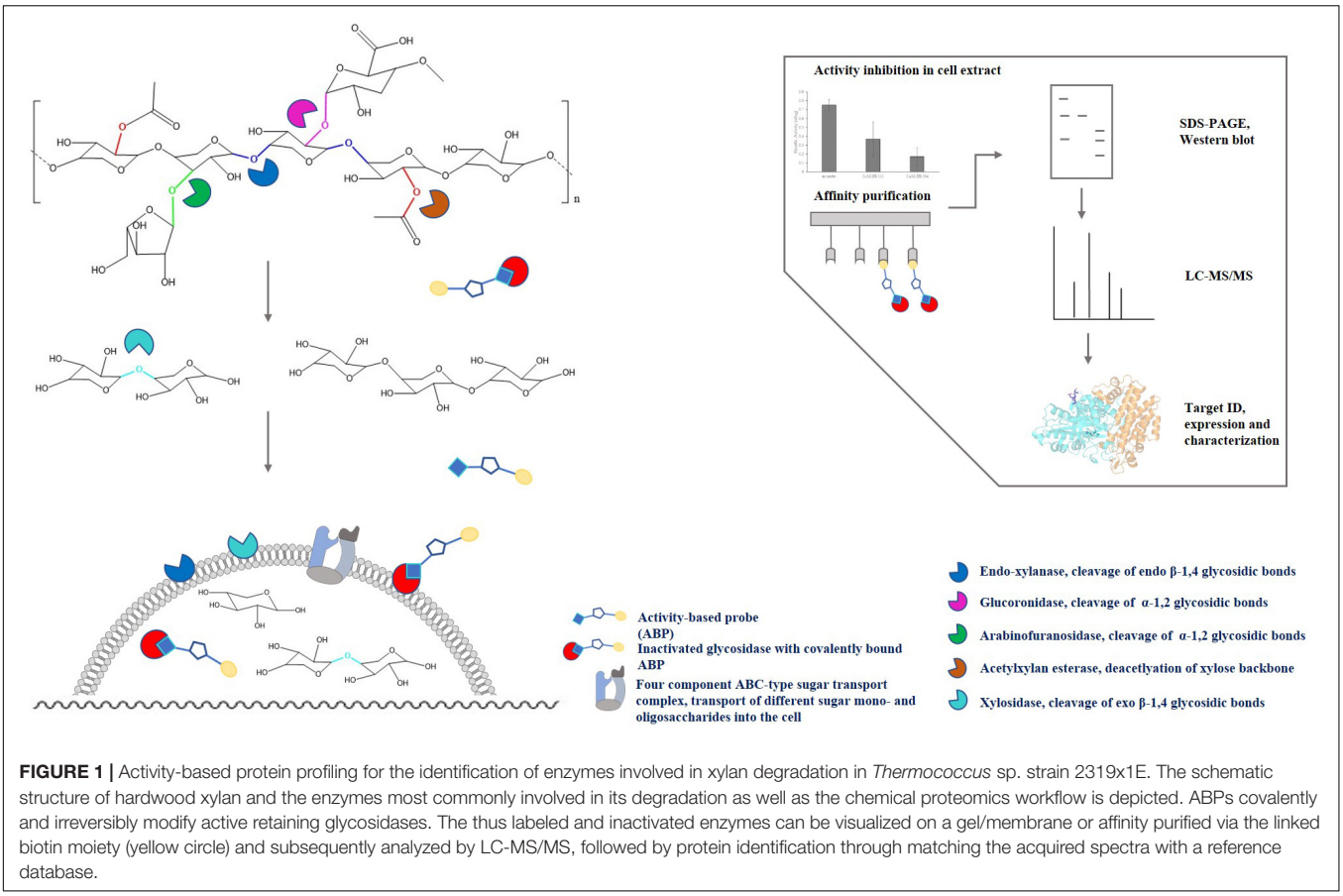


FIGURE 1 | Activity-based protein profiling for the identification of enzymes involved in xylan degradation in *Thermococcus* sp. strain 2319x1E. The schematic structure of hardwood xylan and the enzymes most commonly involved in its degradation as well as the chemical proteomics workflow is depicted. ABPs covalently and irreversibly modify active retaining glycosidases. The thus labeled and inactivated enzymes can be visualized on a gel/membrane or affinity purified via the linked biotin moiety (yellow circle) and subsequently analyzed by LC-MS/MS, followed by protein identification through matching the acquired spectra with a reference database.

i.e., ABPP using glycosidase-selective ABPs complemented by a comparative full-proteome analysis, for the identification of xylan degrading enzymes from the new isolate. The respective workflow of ABPP-based chemical proteomics as well as the degradation of the xylan polymer by different classes of enzymes are depicted in **Figure 1**. This approach extends the repertoire of ABPP in archaea, and hyperthermophiles in particular, by the identification of glycoside hydrolases, highlighting the importance of ABPP for deciphering metabolic processes and identifying novel enzymes.

MATERIALS AND METHODS

Chemicals

Chemicals for cultivation of *Escherichia coli* and *Thermococcus* sp. strain 2319x1E, such as yeast extract, lysogeny broth (LB), beechwood xylan, D-glucose and D-xylose were obtained from Carl Roth (Germany). Avicel® cellulose, carboxymethyl cellulose (CMC), maltodextrin and soluble starch were purchased from Sigma-Aldrich (United States), xylobiose, *ortho*-nitrophenyl-β-D-galactopyranoside (ONPG), *para*-nitrophenyl acetate (PNPA), *para*-nitrophenyl-β-D-xylopyranoside (PNPX) and *para*-nitrophenyl- β-D-glucopyranoside (PNPG) from Megazyme (Ireland). Other chemicals which were used in this work are 4',6-diamidino-2-phenylindole (DAPI; Thermo Fisher Scientific, United States), urea (GE Healthcare Life Sciences, Germany), ammonium bicarbonate (ABC; Sigma-Aldrich, United States), dithiothreitol (DTT; Sigma-Aldrich, United States), iodoacetamide (IAM; Sigma-Aldrich, United States), formic acid (FA; Fischer Chemical, Germany) and bovine serum albumin (BSA; VWR Chemicals, United States).

Genome Sequencing, Assembly and Annotation

Thermococcus sp. strain 2319x1E was sequenced based on short-read and long-read sequencing methods. In a first step, whole-genome-shotgun PCR-free libraries for Illumina sequencing were constructed from 2 μg of gDNA with the Illumina TruSeq DNA PCR-free Sample Preparation Kit (Illumina, United States) based on the manufacturer's protocol. The libraries were quality controlled by analysis on an Agilent 2000 Bioanalyzer with Agilent High Sensitivity DNA Kit (Agilent Technologies, United States) for fragment sizes of 550 bp. Sequencing was performed on a Illumina MiSeq (2 × 300 bp/v3 chemistry) in paired-end mode. Adapters and low-quality reads were removed by an in-house software pipeline prior to polishing as recently described (Wibberg et al., 2016).

In a second step, the Oxford Nanopore sequencing library from *Thermococcus* sp. strain 2319x1E gDNA was prepared using the Nanopore Native barcoding genomic DNA kit with

native barcodes (SQK-LSK109 with EXP-NBD104) according to the manufacturer's instructions. The resulting libraries were sequenced on an Oxford Nanopore GridION Mk1 sequencer using an R9.4.1 flow cell, which was prepared according to the manufacturer's instructions. MinKNOW was used to control the run using the 24 h sequencing run protocol; live base calling was performed using guppy v3.2.6. The assembly was performed as described recently with smaller modifications - Unicycler v0.4.6. was used for the hybrid-assembly and no additional polishing was performed (Wibberg et al., 2020). The finished genome sequence was imported into the annotation platforms Prokka (Seemann, 2014) and GenDB (Meyer et al., 2003) for automatic prediction of genes and functional annotation as described previously (Tomazetto et al., 2018).

Phylogenetic Analysis

For an extensive phylogenic analysis, all available genomes of cultivated species of *Thermococcus*, *Pyrococcus*, *Palaeococcus,* and *Methanothermobacter thermautotrophicus* Delta H (used as an outgroup) were downloaded from GenBank. In total, 122 archaeal-specific conserved proteins were identified from 34 genomes and aligned using gtdb-tk v.1.0.2 (Parks et al., 2018; Chaumeil et al., 2019). The concatenated alignment was refined using Gblock v.0.91b with the most gentle parameters and complete gap elimination option (Castresana, 2000). The final alignment contained 27,083 amino acids. A phylogenetic tree was constructed in RAxML v.8.2.12 (Stamatakis, 2014) with the PROTGAMMAILG model of amino acid substitutions. Local support values were 1000 rapid bootstrap replications.

Genomic Analysis and Bioinformatics Methods

Carbohydrate-active enzyme (CAZyme) genes were searched with dbCAN2 (Zhang et al., 2018) using HMMER (Potter et al., 2018) with default parameters. For selected CAZymes, phylogenetic analysis was performed as follows: the top 5 BLAST homologs for each sequence and all proteins from the Swiss-Prot database (evidence at protein level only) affiliated to the corresponding family were aligned in Muscle (Edgar, 2004). Phylogenetic trees were constructed in RAxML v.8.2.12, as described above (*Phylogenetic analysis* section). Manual curation of protein function prediction was conducted using BlastP (McGinnis and Madden, 2004), HHpred (Zimmermann et al., 2018), HHMMER (phmmer and hmmscan) (Potter et al., 2018), and SUPERFAMILY 2 (Pandurangan et al., 2019). Secretion signal peptides and transmembrane domains were predicted using SignalP (Almagro Armenteros et al., 2019), Phobius (Käll et al., 2007), and TMHMM (Krogh et al., 2001). Search for orthologous gene clusters in *Thermococcus* sp. strains 2319x1 and 2319x1E was performed with OrthoVenn2 (Xu et al., 2019).

Cultivation and Growth Experiments

Thermococcus sp. strain 2319x1E was isolated from the same *in situ* enrichment as the previously described *Thermococcus* sp. strain 2319x1 (Gavrilov et al., 2016). Cultivation as well as growth experiments were conducted in a strictly anaerobic modified Pfennig medium, consisting of 0.33 g l^{-1} $MgCl_2$, $CaCl_2$, KCl, NH_4Cl and KH_2PO_4, 9 g l^{-1} NaCl, 2 g l^{-1} $NaHCO_3$, 16 g l^{-1} $Na_2S \cdot 9\ H_2O$, 0.5 ml l^{-1} trace elements (Kevbrin and Zavarzin, 1992), 1 ml l^{-1} vitamin solution (Wolin et al., 1963), 1 g l^{-1} sulfur, 0.1 g l^{-1} yeast extract and 1 g l^{-1} of the respective growth substrate (beechwood xylan, Avicel® cellulose, D-glucose, D-xylose). All growth experiments were performed at 85°C and pH 7.0 under a N_2-atmosphere, either in 500 ml Schott bottles without shaking or in a 10 l bioreactor (Bioengineering, Switzerland) with constant stirring (100 rpm) and N_2 infusion (0.5 l min^{-1}). Growth was determined by counting cells in a Neubauer improved counting chamber (Brand, Germany) on a Zeiss Axioscope microscope (Zeiss, Germany). Cells that were further used for assays, ABPP experiments or proteomic analysis, were grown on the respective substrate for at least three consecutive transfers to the same fresh medium.

Determination of Enzyme Activity in *Thermococcus* Cells

Thermococcus sp. strain 2319x1E cells were collected from grown cultures by centrifugation at up to 8,000 × *g* for 10 min at 25°C, washed two times with 1× PBS (8 g l^{-1} NaCl, 0.2 g l^{-1} KCl, 1.44 g l^{-1} Na_2HPO4, 0.24 g l^{-1} KH_2PO_4, pH 7.0) and stored at –80°C. After thawing, the cell pellets were resuspended in 1× PBS pH 7.0 (2 ml per gram wet weight) and sonicated with a UP 200S sonicator (Hielscher Ultrasonics GmbH, Germany) for 3× 5 min (cycle 0.5, amplitude 50) followed by centrifugation at 12,000 × *g* for 30 min at room temperature to remove intact cells and cell debris. To separate the cytosolic protein fraction from the membrane protein fraction, ultracentrifugation of the crude protein extract was done (30 min, 100,000 × *g*, 4°C). To retrieve the respective membrane fraction, the pellet was resuspended in 1 × PBS pH 7.0 after collecting the cytosolic protein-containing supernatant. Protein concentrations were determined using the Bradford assay (Bio-Rad, United States) with bovine serum albumin (BSA) as standard (Bradford, 1976).

For determining glucosidase activities of *Thermococcus* sp. strain 2319x1E from either crude extracts or cytosolic and membrane fractions, the 3,5-dinitrosalicylic acid (DNSA) assay was applied (Miller, 1959). The enzyme assays were performed in 50 mM Tris-HCl pH 7.0 with a total reaction volume of 550 µl, containing 250 µl substrate solution in H_2O (1% w/v for xylan and CMC, 0.04% w/v for xylobiose). The reactions were started by the addition of 7.5 µg of protein from the respective sample followed by incubation for 30 min at 85°C. The reaction mixture was cooled on ice for 5 min, mixed with 700 µl DNSA reagent (Miller, 1959) and subsequently incubated at 100°C for 10 min. The concentration of reducing sugars was then calculated from the absorption of the samples at 575 nm by using a D-xylose or D-glucose calibration curve. Background absorption of cell extracts and substrate solution was subtracted from the absorption of the samples. One unit of enzyme

activity was defined as the amount of enzyme required to release 1 μmol D-xylose equivalents per minute. All absorbance measurements were done in 96-well plates using a Tecan infinite M200 plate reader (Tecan trading AG, Switzerland).

Cloning of Glycoside Hydrolases for Protein Expression in *Escherichia coli*

The *Thermococcus* sp. strain 2319x1E genes *EGIDFPO_00674* (1800 nt) and *EGIDFPO_00532* (2196 nt) were amplified from gDNA extracted with the DNeasy Blood & Tissue Kit (Qiagen, Germany) using Q5® polymerase (NEB, United States) and the following gene specific primers (Eurofins Genomics, Germany): for *EGIDFPOO_00674* 5′-GCGGCCCATGGGCTACCAGAAGTTTGGATATCATTTTCATGC-3′ and 5′-ATTAGAATTCCTAGTGGTGGTGGTGGTGGTGAACCCTGGCCCTATGCTCGCACCATT-3′ (*Nco*I and *Eco*RI restriction sites underlined); for *EGIDFPOO_00532* 5′-CTGGTGCTAGCCTGATGCTTGTCTTTCCCGATTCCTTCCTCT-3′ and 5′-CTATCCTCGAGTCATAGCTCCGGAAGTCCATACTTTG-3′ (*Nhe*I and *Xho*I restriction sites underlined). PCR products were afterwards purified using the Wizard® SV Gel and PCR clean-up kit (Promega, United States).

The genes were cloned into the pET-28b(+) vector (Novagen, United States) for subsequent expression, providing the recombinant proteins with a C-terminal (EGIDFPOO_00674) or N-terminal 6× His-tag. After restriction digest of the purified PCR products and the empty vector with the respective restriction enzymes (NEB), PCR product and vector were used in a molar ratio of 1:6 for *sticky end* ligation using T4 DNA ligase (NEB) at 16°C over night. *E. coli* DH5α cells (Novagene, United States) were transformed with the obtained constructs and the presence of successfully cloned genes was confirmed by colony PCR with the transformed cells as template and T7 promotor and terminator primers. The correctness of the sequence was confirmed by Sanger sequencing of both strands (Eurofins Genomics, Germany). The codon optimized gene (*E. coli*) *EGIDFPOO_00563* was synthesized and cloned into a pET28b(+) vector with N-terminal 6× His-tag (BioCat GmbH, Germany).

Heterologous Expression and Purification of Glycoside Hydrolases

Recombinant expression of the glycoside hydrolases EGIDFPOO_00674, EGIDFPOO_00532 and EGIDFPOO_00563 was performed in *E. coli* Rosetta cells (Novagen, United States) which were transformed with the respective plasmids. A freshly inoculated 1 l culture in LB medium, supplemented with 50 μg ml^{-1} kanamycin and 50 μg ml^{-1} chloramphenicol, was grown to an OD_{600} of 0.4 at 37°C with constant shaking (180 rpm) upon subsequent induction of protein expression with 500 μM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were incubated at 18°C for a further 16 h, harvested by centrifugation (8,000 × *g* for 20 min at 4°C) and resuspended in 3–5 mL of buffer (50 mM Tris-HCl pH 7.0) per gram wet weight of the pellet for further protein purification. Cell lysis was performed by sonication as described above, followed by centrifugation at 12,000 × *g* for 45 min at 4°C. For further purification of EGIDFPOO_00674 or EGIDFPOO_00563, the cleared lysate was passed through a 0.45 μm filter, followed by an affinity purification using a 5 ml Ni-IDA column (GenScript, GenScript Biotech, United States) equilibrated with buffer A (50 mM Tris-HCl pH 7.8, 250 mM KCl, 10 mM imidazole). Protein elution was performed using a linear gradient of 10–350 mM imidazole in buffer A. The elution buffer was exchanged for storage buffer (50 mM Tris-HCl pH 8.0, 20 mM KCl, 5 mM $MgCl_2$) using Amicon® centrifugal filter devices (30 or 50 kDa cutoff, Merck, Germany). For long-time storage at −80°C, 50% (v/v) glycerol was added and the protein solution was flash frozen in liquid N_2. EGIDFPOO_00532 was purified from inclusion bodies using a high pH-buffer purification strategy (Singh et al., 2015). The pellet obtained after sonication was washed with 50 mM Tris-HCl pH 8.0, taken up in 7.5 ml resuspension buffer (50 mM Tris-HCl pH 12.5, 2 M urea) per gram (wet weight) and subsequently incubated for 30 min at room temperature. The solubilized pellet was then centrifuged (8,000 × *g*, 10 min, 4°C) and the supernatant was diluted stepwise 1:20 into refolding buffer [50 mM Tris-HCl pH 8.0, 2 M urea, 10% (w/v) sucrose] on ice. Afterwards, a heat precipitation (20 min, 80°C) with subsequent centrifugation (8,000 × *g*, 10 min, 4°C) was performed. The obtained supernatant was directly used for further enzyme assays.

Activity Assays With Recombinant Glycosidases

The activity of the heterologously expressed proteins EGIDFPOO_00532 and EGIDFPOO_00674 was determined using a discontinuous assay with the colorimetric nitrophenyl substrates PNPA, PNPX, PNPG, OPNG. Upon enzymatic cleavage of the nitrophenyl substrates, the absorbance of the free nitrophenol was measured at 400 nm in 96-well plates using a plate reader. Absorbance values of controls lacking the enzyme were subtracted from the absorbance values of the samples. Enzyme activity was calculated from a calibration curve with *para*-nitrophenol (PNP). The pH-dependent activity was measured at 348 nm in McIlvaine buffers with the respective pH values (McIlvaine, 1921) as described previously (Kallnik et al., 2014). Deacetylase activity of EGIDFPOO_00674 was measured in 50 mM Tris-HCl pH 8.0 using 2.5 μg of the purified enzyme and up to 8.8 mM PNPA in a total sample volume of 250 μl. α-amylase activity of EGIDFPOO_00674 was determined in 50 mM Tris-HCl pH 8.0 using the DNSA assay as described above, with 10 μg of the purified enzyme and 3 mM 6-*O*-α-maltosyl-β-cyclodextrin (MβCD) per 550 μl reaction volume. The β-glycosidase activity of EGIDFPOO_00532 was assessed in refolding buffer containing 25 μg of the purified enzyme and up to 26.7 mM of the respective PNP-substrate. Kinetics were determined at the optimal pH and temperature which was pH 8.0 and 100°C for both enzymes. To assess the thermostability of the enzymes, the PNP-substrate assay was performed as described above, but the enzymes were preincubated at 60,

80, and 100°C for up to 4 h. The glycogen phosphorylase activity of EGIDFPOO_00563 was assayed using a discontinuous enzyme assay; First, 500 µl samples containing 50 mM Tris-HCl pH 7, 20 µg purified EGIDFPOO_00563, 10 mM NaH_2P0_4 and 0.05% (w/v) substrate (glycogen, maltodextrin or starch) were incubated for 5 min at 100°C. Subsequently, the samples were stored on ice for 5 min. The amount of glucose 1-phosphate formed in the glycogen phosphorylase reaction was then determined by the indicator reaction, containing 4 U phosphoglucomutase (Sigma-Aldrich, United States), 3 U glucose-6-phosphate dehydrogenase (Roche, Switzerland), 10 mM $MgCl_2$, and 0.5 mM $NADP^+$ in 50 mM Tris-HCl pH 7 (total volume of 500 µl). The assay was incubated for 3 min at 30°C and the amount of glucose 1-phosphate was calculated from the absorption at 340 nm using a Specord 210® photometer (Analytik Jena, Germany).

Activity Inhibition With Activity-Based Probes

Protein solutions were diluted with their respective buffer to a final protein concentration of 0.02 mg protein ml^{-1} for native crude extract and 0.03 mg protein ml^{-1} for heterologously expressed purified protein. To examine activity inhibition upon covalent binding of the glycosidase probes JJB384 or JJB111, the samples (containing heterologously expressed proteins or *Thermococcus* sp. Strain 2319x1E crude extract) were incubated with 2 and 4 µM JJB384 or JJB111, respectively, for 1 h at 85 °C and subsequently used for activity determination assays with DNSA or nitrophenyl substrates.

Sampling Cells for Proteomics Studies

Cells for proteomics studies were grown in Schott bottles as described above (Cultivation and Growth Experiments). For every carbon source under evaluation (beechwood xylan, Avicel® cellulose, D-glucose, D-xylose), four biological replicates have been prepared. Therefore, 200 ml growth medium supplemented with the respective substrate were inoculated from a preculture which has been grown on that substrate for at least three consecutive transfers. Cells were harvested from cultures by centrifugation up to 8,000 × *g* for 20 min at 4°C after 10–12 h of growth during early mid-log phase. For storage at –80°C, the pellets were resuspended in 2 ml of the culture supernatant, transferred to fresh 15 ml Falcon tubes and centrifuged 6,500 × *g* for 10 min at 4°C. After removing the supernatant, the pellets were directly flash frozen in liquid N_2.

Protein Extraction for Full Proteome Studies and Activity-Based Protein Profiling Experiments

Cell pellets were resuspended in an adequate volume (approx. 2000–4000× less than the original culture volume) of 50 mM Na_2HPO_4 pH 8.0 supplemented with 1× MS-SAFE Protease and Phosphatase Inhibitor (Sigma-Aldrich, United States) and transferred to 1.5 ml reaction tubes. Cell lysis was performed by conducting a two-step sonication protocol. First, the cells were sonicated 7× 1 min in an ultrasonic bath filled with ice with short breaks in between to vortex the samples, followed by sonication with a Bioruptor UCD-200 (Diagenode, Belgium) using the following conditions: 1 min pulse and 30 s break in 10 cycles with the intensity set to "high" while cooled with ice. The extracts were cleared twice by centrifugation (21,000 × *g*, 20 min and 5 min, 4 °C) to remove cell debris and the protein concentration of the cleared extracts was determined by a modified Bradford assay with ROTI®Nanoquant (Carl Roth, Germany).

Sample Preparation for Full Proteome Analysis

For each sample, 15 µg of total protein were subjected to a methanol-chloroform precipitation (Wessel and Flügge, 1984). The pellet was washed with methanol twice, dried on air and resuspended in 25 µl 8 M urea in 100 mM ABC, followed by reduction of disulfide bonds with 5 mM DTT in 100 mM ABC and incubation at 23°C for 30 min shaking at 1,000 rpm. Subsequently, alkylation of cysteine residues was done using 20 mM IAM in 100 mM ABC and incubation at 37°C for 30 min shaking in the dark at 1,000 rpm. The reaction was quenched by adding DTT to a final concentration of 25 mM. Afterwards, the proteins were digested with 500 ng Lys-C (FUJIFILM Wako Chemicals, Japan) dissolved in 100 mM ABC for 3 h at 37°C shaking at 1,000 rpm, followed by overnight digestion (∼16 h) with 500 ng trypsin (Thermo Scientific, United States) dissolved in 50 mM acetic acid at 37°C shaking at 1,000 rpm. Prior to sample clean-up for LC-MS/MS, formic acid (FA) was added to the samples to a final concentration of 5% (v/v).

In variation, the pellet of the samples from cells grown on xylan was washed once with 10% DMSO/10% ethanol and thrice with methanol after precipitation. Furthermore, the samples were dried in a vacuum concentrator (Eppendorf, Germany) after acidification and the peptides were dissolved in 0.5% (v/v) FA prior to desalting.

In vivo Activity-Based Protein Profiling of Glycoside Hydrolases

All chemical probes were dissolved in DMSO. Labeling of active glycoside hydrolases was performed using a 2 ml liquid culture which was obtained by concentration of the growth culture (50× for samples analyzed on gel; 500× for samples dedicated for MS-based target identification) by centrifugation at up to 8,800 × *g* for 20 min at 25°C. For labeling, 2 µM of JJB384 or JJB111 were added to the liquid culture, followed by incubation at 78°C for 2 h shaking at 180 rpm. Afterwards, the cells were harvested by centrifugation at 8,800 × *g* for 10 min at 25°C. Protein extraction was performed as described above. For MS-based target identification, each sample was prepared in triplicates.

Detection of Labeled Glycosidases by Western Blot Analysis

Labeled protein extracts (10–15 µg protein) were mixed with 4× lithium dodecyl sulfate (LDS) gel loading dye [423 mM Tris-HCl, 563 mM Tris base, 8% (w/v) LDS, 40% (w/v) glycerol, 2 mM EDTA, 0.075% (w/v) SERVA Blue G250; freshly supplemented with 100 mM DTT], incubated at 70°C for 15 min and separated

by denaturing polyacrylamide gel electrophoresis on 11% Bis-Tris resolving gels. The separated proteins were transferred on a PVDF membrane (Merck, United States) using a tank blot setup (Bio-Rad, United States) and the membrane was washed thrice with TBS-T [20 mM Tris base pH 7.5, 150 mM NaCl, 0.2% (w/v) Tween20]. Blocking with 3% BSA (w/v) in TBS-T was done overnight at 4°C, followed by incubation with a streptavidin-horse radish peroxidase (HRP) conjugate (Sigma-Aldrich, United States) directly added into the blocking solution (1:25,000) for 2.5 h at room temperature. The membrane was washed 6× with TBS-T and the labeled proteins were detected by enhanced chemiluminescence (ECL) with a mix of the SuperSignal® West Pico Chemiluminescent Substrate and the SuperSignal® West Femto Maximum Sensitivity Substrate (4:1; Thermo Scientific, United States) using an Amersham Imager 600 (GE Healthcare Life Sciences, Germany).

Enrichment of Labeled Glycoside Hydrolases and Sample Preparation for Mass Spectrometry-Based Proteomics

Prior to affinity enrichment, the protein solution (225 μg protein in a final volume of 500 μl 50 mM Na_2HPO_4 pH 8.0 supplemented with 1× MS-SAFE Protease and Phosphatase Inhibitor) was cleaned-up by precipitation with a 4× volume of methanol overnight at –20°C. The proteins were collected by centrifugation at 21,000 × *g* for 10 min at 4°C and washed with a 2× volume of methanol. The air-dried pellets were dissolved in 850 μl 2% (w/v) SDS in 1× PBS (155 mM NaCl, 3 mM Na_2HPO_4, 1.06 mM KH_2PO_4, pH 7.4) by incubation at 37°C with agitation at 1,000 rpm and diluted with 1× PBS pH 7.4 to a final concentration of 0.2% (w/v) SDS. The resulting protein solutions were incubated with avidin agarose beads (Thermo Scientific, United States), equilibrated in the same buffer, for 1 h at room temperature while gently tumbling. The beads were collected by centrifugation (400 × *g*, 5 min), washed 5× with 1% (w/v) SDS in MS-quality water and finally 4× with pure MS-quality water. Afterwards, tryptic on-bead-digestion was conducted. Thereto, the beads were taken up in 100 μl 0.8 M urea in 50 mM ABC buffer. Reduction of disulphide bonds with 10 mM DTT in 50 mM ABC was done by incubation at room temperature for 1 h shaking at 1,500 rpm, followed by alkylation of cysteine residues with 25 mM IAM in 50 mM ABC and incubation at room temperature for 1 h in the dark with shaking at 1,500 rpm. The reaction was quenched by increasing the DTT concentration to 35 mM and further 10 min incubation. For digestion of proteins, 10 μl of a 100 ng μl^{-1} trypsin stock solution in 50 mM acetic acid were added and the samples were incubated at 37°C for 16 h with shaking at 1,250 rpm. The beads were collected by centrifugation at 3,000 × *g* for 5 min and FA was added to the recovered supernatants to a final concentration of 5% (v/v). The beads were washed with 50 μl 1% (v/v) FA at room temperature for 15 min shaking at 1,500 rpm and collected by centrifugation at 3,000 × *g* for 5 min. Both supernatants were combined and subsequently cleared by passing over home-made tips containing two disks of glass microfiber (GE Healthcare, Life Sciences, Germany; poresize 1.2 μm; thickness 0.26 mm).

Sample Clean-Up for LC–MS/MS

Peptides were desalted on home-made C_{18} StageTips (Rappsilber et al., 2007) containing two layers of an octadecyl silica membrane (3M, United States). All centrifugation steps were carried out at room temperature. The StageTips were first activated and equilibrated by passing 50 μl of methanol (600 × *g*, 2 min), 80% (v/v) acetonitrile (ACN) with 0.5% (v/v) FA (600 × *g*, 2 min) and 0.5% (v/v) FA (800 × *g*, 3 min) over the tips. Next, the tryptic digests were passed over the tips (800 × *g*, 3–4 min). The flow through was collected and applied a second time (same settings). The immobilized peptides were then washed with 50 μl and 25 μl 0.5% (v/v) FA (800 × *g*, 3 min). Bound peptides were eluted from the StageTips by an application of two rounds of 25 μl 80% (v/v) ACN with 0.5% (v/v) FA (600 × *g*, 2 min). After elution from the StageTips, the peptide samples were dried using a vacuum concentrator (Eppendorf, Germany) and the peptides were dissolved in 15 μl 0.1% (v/v) FA prior to analysis by MS.

LC-MS/MS Analysis

Experiments were performed on an Orbitrap Elite or Orbitrap Fusion Lumos mass spectrometer (Thermo Fischer Scientific, United States) that were coupled to an EASY-nLC 1000 or 1200 liquid chromatography (LC) system (Thermo Fischer scientific, United States). The LCs were operated in the one-column mode. The analytical column was a fused silica capillary (inner diameter 75 μm × 36 or 46 cm) with an integrated PicoFrit emitter (New Objective, United States) packed in-house with Reprosil-Pur 120 C18-AQ 1.9 μm (Dr. Maisch, Germany). The analytical column was encased by a column oven (Sonation, Germany) and attached to a nanospray flex ion source (Thermo Fischer scientific, United States). The column oven temperature was adjusted to 45 or 50°C during data acquisition. The LC was equipped with two mobile phases: solvent A (0.1% FA, in water) and solvent B (nLC 1000: 0.1% FA in ACN; nLC 1200: 0.1% FA, 20% H_2O, in ACN). All solvents were of UHPLC (ultra-high-performance liquid chromatography) grade (Honeywell, Germany). Peptides were directly loaded onto the analytical column with a maximum flow rate that would not exceed the set pressure limit of 980 bar (usually around 0.5 – 0.8 $\mu l\ min^{-1}$). Peptide solutions were subsequently separated on the analytical column using different gradients (for details see **Supplementary Data Sheet 2**).

The mass spectrometers were operated using Xcalibur software (Elite: v2.2 SP1.48; Lumos: v4.3.7.3.11). The mass spectrometers were set in the positive ion mode. Precursor ion scanning (MS^1) was performed in the Orbitrap analyzer (FTMS; Fourier Transform Mass Spectrometry with the internal lock mass option turned on (lock mass was 445.120025 *m*/*z*, polysiloxane) (Olsen et al., 2005). MS^2 Product ion spectra were recorded only from ions with a charge higher than +1 and in a data dependent fashion in the ion trap mass spectrometry. All relevant MS settings (Resolution, scan range, AGC, ion acquisition time, charge states isolation window, fragmentation type and details, cycle time, number of scans performed, and

various other settings) for the individual experiments can be found in **Supplementary Data Sheet 2**.

Peptide and Protein Identification Using MaxQuant and Perseus

RAW spectra were submitted to an Andromeda search (Cox et al., 2011) in MaxQuant (v.1.6.2.6) using the default settings (Cox and Mann, 2008). Label-free quantification (LFQ) (Cox et al., 2014) and match between runs was activated. MS/MS spectra data were searched against the self-constructed TspE_proteome_AA_new.fasta file (2182 entries; see above for reference). All searches included a contaminants database (as implemented in MaxQuant, 246 sequences). The contaminants database contains known MS contaminants and was included to estimate the level of contamination. Andromeda searches allowed oxidation of methionine residues (16 Da) and acetylation of the protein N-terminus (42 Da) as dynamic modifications while carbamidomethylation of cysteine residues (57 Da, alkylation with IAM) was selected as static modification. Enzyme specificity was set to "Trypsin/P." The instrument type in Andromeda searches was set to Orbitrap and the precursor mass tolerance was set to ±20 ppm (first search) and ±4.5 ppm (main search). The MS/MS match tolerance was set to ±20 ppm. The peptide spectrum match FDR and the protein FDR were set to 0.01 (based on target-decoy approach). Minimum peptide length was 7 amino acids. For protein quantification, unique and razor peptides were allowed. In addition to unmodified peptides, modified peptides with dynamic modifications were allowed for quantification. The minimum score for modified peptides was set to 40.

Further data analysis and filtering of the MaxQuant output was done in Perseus (v.1.6.2.1) (Tyanova et al., 2016). LFQ intensities were loaded into the matrix from the proteinGroups.txt file and potential contaminants as well as reverse hits from the reverse database and hits only identified based on peptides with a modification site were removed. For all MS-based proteomics experiments, biological replicates were combined into categorical groups to allow comparison of the different treatment groups and the data was transformed to the $\log_2$-scale. For the full proteome analysis, the data was filtered to keep only hits with a valid LFQ intensity for at least three out of four replicates for samples from cells grown on xylan. For the identification of glycoside hydrolases enriched with JJB384, the data was separately filtered for the two sugar substrates under investigation. Only hits with a valid LFQ intensity for a minimum of two out of three of the probe-labeled sample replicates were retained for further analysis. Prior to quantification, missing values were imputed from a normal distribution (width 0.3, down shift 1.8). Comparison of normalized protein group quantities (relative quantification) between different MS runs was solely based on the LFQ intensities as calculated by MaxQuant (MaxLFQ algorithm) (Cox et al., 2014). Briefly, label-free protein quantification was switched on and unique and razor peptides were considered for quantification with a minimum ratio count of 2. Retention times were recalibrated based on the built-in non-linear time-rescaling algorithm. MS/MS identifications were transferred between LC-MS/MS runs with the "Match between runs" option in which the match time window was set to 0.7 min and the alignment time window to 20 min. The quantification was based on the "value at maximum" of the extracted ion current. At least two quantitation events were required for a quantifiable protein. The $\log_2$-fold enrichment of protein groups for samples from cells grown on xylan was calculated based on the mean LFQ intensity compared to the samples from cells grown on any of the other sugars. Statistical significance was calculated based on a two-sided Student's *t*-test. Full MS data for the comparative full proteome analysis can be found in **Supplementary Table 1**. The $\log_2$-fold enrichment of protein groups with JJB384 was calculated based on the mean LFQ intensity compared to the DMSO control. Protein groups with a negative fold enrichment were excluded from further analysis. Full MS data for the ABPP with JJB384 can be found in **Supplementary Table 2**.

RESULTS

Thermococcus sp. Strain 2319x1E Genome Analysis

Thermococcus sp. strain 2319x1E was isolated from the same enrichment culture as *Thermococcus* sp. strain 2319x1 (Gavrilov et al., 2016). The fully assembled genome of *Thermococcus* sp. strain 2319x1E consists of a single chromosome with the size of 1,989,851 bp and an average GC content of 44.39%. The sequence is deposited in the NCBI genome database under accession number LR778300. The average nucleotide identity (ANI) to the genome of *Thermococcus* sp. strain 2319x1 is 97.86%, thus being above the species level (95%) according to Goris et al. (2007), whereas the ANI to *Thermococcus litoralis,* which is considered the closest validly published *Thermococcus* species described, is only 90.01% and to the more distantly related *Thermococcus kodakarensis* 71.97%. The phylogenetic position of both *Thermococcus* isolates was determined using phylogenetic tree construction based on the "ar122" set of conserved proteins (**Figure 2**).

The predicted number of protein-coding sequences (CDS) in *Thermococcus* sp. strain 2319x1E is 2,195, and therefore is in a similar range as the 2319x1 genome (2,192 CDS). The comparison of both genomes revealed 1,886 single-copy orthologous gene clusters found in both strains and 266 and 360 unique genes in the genome of 2319x1E and 2319x1, respectively. Altogether, these numbers indicate that the two *Thermococcus* strains are relatively closely related to each other. With regard to enzymes involved in xylan degradation, the CAZymes present in both strains were compared. The *Thermococcus* sp. strain 2319x1E genome comprises 42 genes which putatively encode for CAZymes (Lombard et al., 2014), with 26 of them being predicted to encode for glycosyl transferases (GT), 13 for glycosyl hydrolases (GH) and 3 for carbohydrate esterases (CE) (**Table 1**). Remarkably, the genes encoding the unique multifunctional multidomain glycosidase described for the *Thermococcus* sp. strain 2319x1 isolate (Gavrilov et al., 2016) and 4 additional GHs, as well as 8 GTs are not present in the genome of the new isolate 2319x1E. Correspondingly, in the *Thermococcus* sp. strain

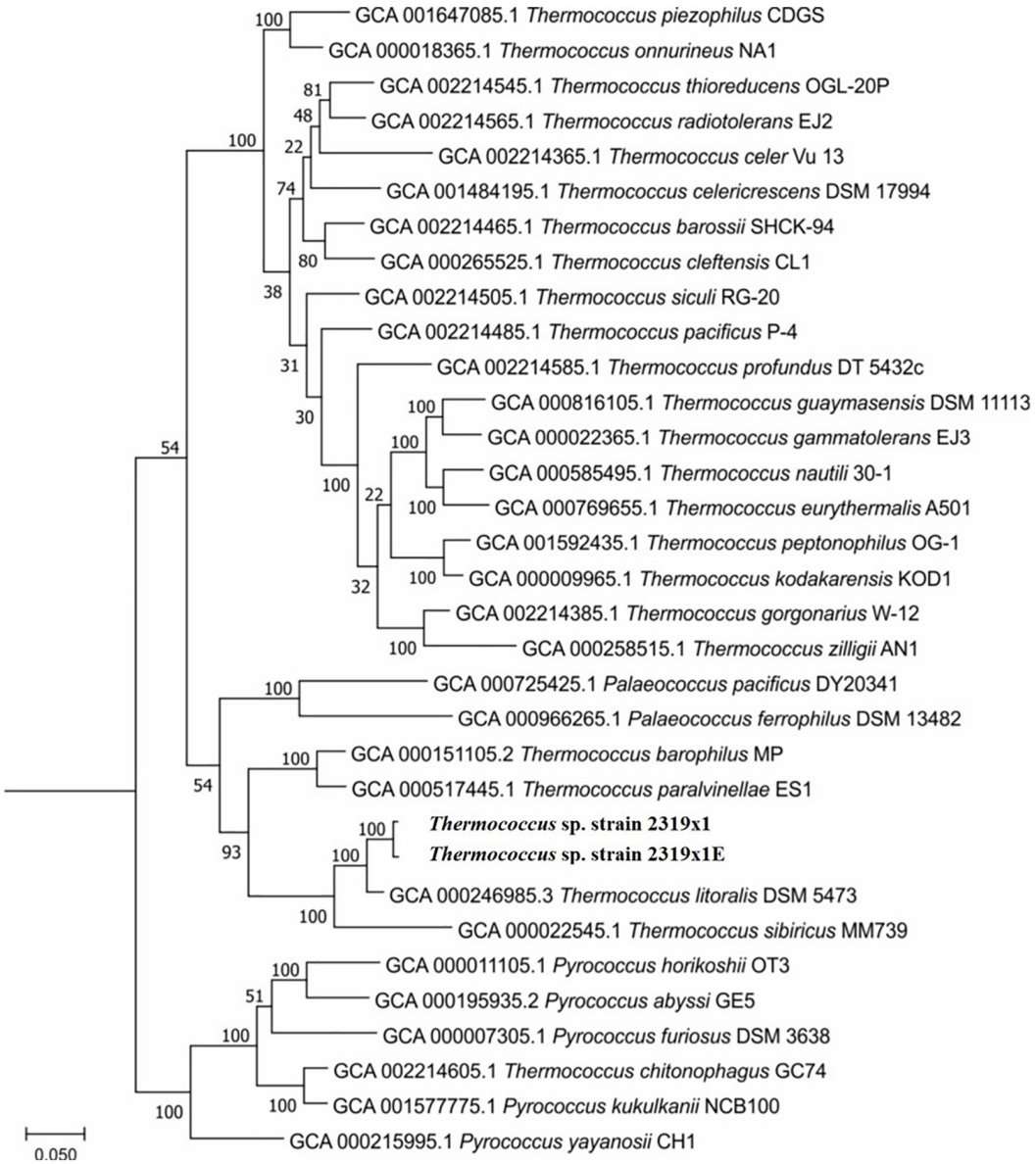


FIGURE 2 | Phylogenetic position of the novel *Thermococcus* sp. strain 2319x1E in comparison with all available cultivated genomes of *Thermococcus*, *Pyrococcus*, and *Palaeococcus* species. Tree construction was done in RAxML v. 8.2.12 with the PROTGAMMAILG model of amino acid substitutions.

2319x1, genes encoding 9 GTs and 2 GHs (EGIDFPOO_01323 and EGIDFPOO_01324) that are present in strain 2319x1E are missing (**Supplementary Table 1** in **Supplementary Data Sheet 1**).

Of the CAZymes identified in *Thermococcus* sp. strain 2319x1E, two putative GHs encoded by the genes EGIDFPOO_01845 and EGIDFPOO_01993 were predicted to be extracellular (**Table 1**), although it should be noted that the reliable prediction of archaeal secretion signals remains challenging (Bagos et al., 2009; Szabo and Pohlschroder, 2012). In addition, 11 putative GTs were predicted to be membrane-associated (**Table 1**) due to the presence of one or more predicted transmembrane domains. Among the 12 putative GHs, 7 are most likely involved in the hydrolysis of α-linked oligo- and polymers such as glycogen, starch, pullulan and maltooligosaccharides, as determined by BLAST against the Swiss-Prot database as well as phylogenetic analyses for some families (GH13 and GH57; **Supplementary Figures 1**, **2** in **Supplementary Data Sheet 1**). In addition, three enzymes were annotated as β-specific hydrolases: EGIDFPOO_00274 is predicted to act as a β-1,4-mannooligosaccharide phosphorylase (GH130), while the two GHs EGIDFPOO_00532 and EGIDFPOO_01324 are putative β-glucosidases (GH1). However, no homologs of enzymes known to be involved in xylan degradation, such as endo-xylanases, β-xylosidases, α-arabinofuranosidases, α-glucuronidases or acetylxylan esterases, were identified.

TABLE 1 | CAZymes identified in the *Thermococcus* sp. strain 2319x1E genome.

CAZy domain	Predicted function	Gene	Secretion signal	Transmembrane domain
Glycosyl hydrolases				
GH13	α-1,4-glucan:maltose-1-phosphate maltosyltransferase	*EGIDFPOO_00018*		(+)
GH130	β-1,4 -mannooligosaccharide phosphorylase	*EGIDFPOO_00274*		
GH57	α-amylase/4-α-gluconotransferase	*EGIDFPOO_00375*		
GH1	β-glucosidase	*EGIDFPOO_00532*		
GH57		*EGIDFPOO_00534*		
GH57		*EGIDFPOO_00674*		
GH122	α-glucosidase	*EGIDFPOO_00753*		
GH57	1,4-α-branching enzyme	*EGIDFPOO_01266*		
	Amylo- α -1,6-glucosidase	*EGIDFPOO_01323*		
GH1	β-glucosidase	*EGIDFPOO_01324*		
GH57		*EGIDFPOO_01845*	Sec/SPI	+
GH13	α-amylase/neopullulanase	*EGIDFPOO_01849*		
GH13	α-amylase/neopullulanase	*EGIDFPOO_01993*	Sec/SPII	+
Carbohydrate esterases				
CE10		*EGIDFPOO_00955*		
CE10		*EGIDFPOO_01219*		
CE1	Putative carbohydrate esterase	*EGIDFPOO_01302*	Sec/SPII	
Glycosyl transferases				
GT2	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	*EGIDFPOO_00012*		++
GT55		*EGIDFPOO_00184*		
GT66		*EGIDFPOO_00197*	Sec/SPI	++
GT39		*EGIDFPOO_00332*	Sec/SPI	++
GT66		*EGIDFPOO_00336*		++
GT2	Undecaprenyl-phosphate mannosyltransferase	*EGIDFPOO_00399*		
GT35	Glycogen phosphorylase	*EGIDFPOO_00563*		
GT2	Phosphoglycolate phosphatase	*EGIDFPOO_00612*		
GT4	*N*-acetyl-α-D-glucosaminyl L-malate synthase	*EGIDFPOO_01181*		
GT4	*N*-acetyl-α-D-glucosaminyl L-malate synthase	*EGIDFPOO_01440*		+
GT81	Glucosylglycerate synthase	*EGIDFPOO_01674*		
GT5	Glycogen synthase	*EGIDFPOO_01851*		
GT2	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	*EGIDFPOO_01900*		++
GT66		*EGIDFPOO_02109*		++
GT2	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	*EGIDFPOO_02112*		++
GT4	D-inositol-3-phosphate glycosyltransferase	*EGIDFPOO_02119*		
GT4	D-inositol-3-phosphate glycosyltransferase	*EGIDFPOO_02121*		
GT4	D-inositol-3-phosphate glycosyltransferase	*EGIDFPOO_02122*		
GT4	Spore coat protein SA	*EGIDFPOO_02135*		
GT4		*EGIDFPOO_02136*		
GT4	D-inositol-3-phosphate glycosyltransferase	*EGIDFPOO_02137*		
GT2		*EGIDFPOO_02138*		++
GT2		*EGIDFPOO_02139*		+
GT4	D-inositol-3-phosphate glycosyltransferase	*EGIDFPOO_02196*		
GT2		*EGIDFPOO_02198*		+
GT55		*EGIDFPOO_02209*		

The table shows all putative CAZymes with their corresponding predicted function, predicted secretion signal and predicted occurrence of transmembrane domains. Sec/SPI or Sec/SPII stands for Sec-type signal peptides or lipoprotein signal peptides, respectively; + indicates the presence of one predicted transmembrane (TM) domain; ++ indicates at least two predicted TM domains; (+) indicates that a TM domain was only predicted with Phobius, but not with TMHMM.

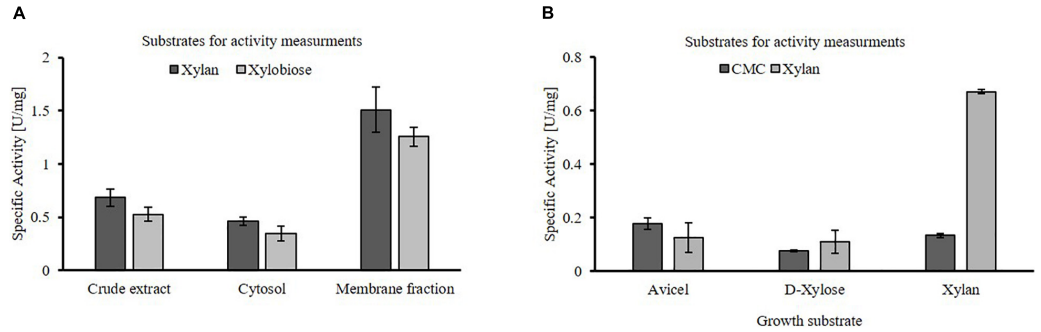


FIGURE 3 | Hydrolytic activity of *Thermococcus* sp. strain 2319x1E cell fractions after growth on different substrates. **(A)** Specific hydrolytic activity of distinct protein fractions from cells grown on beechwood xylan using xylan or xylobiose as substrate, determined by monitoring the increase of reducing sugars via the DNSA assay. The activity in crude extracts, the soluble cytosolic and the membrane fraction is shown. **(B)** Specific enzymatic activity in crude extracts of cells grown on either Avicel® cellulose, D-xylose or beechwood xylan using xylan or CMC as substrate. The enzymatic activity was determined after 30 min of incubation at 85°C using the DNSA assay.

Growth Characteristics of *Thermococcus* sp. Strain 2319x1E

Cells of *Thermococcus* sp. strain 2319x1E show a morphology typical for *Thermococcales* as they appear as irregularly shaped cocci with a size of 1–2 μm in diameter. This strain is an obligate anaerobe and grows optimal at 85°C and a pH of 7.0. When cultivated on xylan in batch mode, the generation time was 3.4 h (2.9 h for D-xylose, 3.7 h for Avicel® cellulose, 3.2 h for D-glucose) and the final growth yield ranged 1–1.5 × 10^7 cells ml^{-1} for all substrates (see **Supplementary Figure 3** in **Supplementary Data Sheet 1**). Notably, cell lysis was observed after 20 h of cultivation by DAPI staining, indicating a short stationary phase when the cells were grown in a closed bottle without stirring and gas exchange.

Proof of Native Xylanolytic Activity

Thermococcus sp. strain 2319x1E crude extracts as well as the isolated cytosolic and membrane fractions obtained from cells grown on xylan possess hydrolytic activity on xylan and xylobiose, similar to strain 2319x1 (**Figure 3A**). The specific xylanolytic activity of the membrane fraction was roughly three times higher than for the cytosolic fraction, thus indicating the presence of membrane-bound hydrolase activity. Furthermore, the hydrolytic activity in crude extracts of cells grown on xylan, D-xylose and Avicel® cellulose was compared using either xylan or CMC as substrate (**Figure 3B**). Whereas a higher xylanolytic activity is clearly linked to growth on xylan, only a slight effect of the growth substrate could be observed for the glucosidase activity on CMC compared to xylan. Of note, no activity was observed in the culture supernatant using the DNSA assay despite long incubation times. Additional efforts to concentrate the supernatant by means of centrifugation or filtration failed due to the viscosity of the residual xylan.

Changes in Protein Abundance in Response to Different Sugars

In order to first identify proteins from *Thermococcus* sp. strain 2319x1E that are regulated in response to the offered carbon source, we used a comparative proteomics approach. *Thermococcus* sp. strain 2319x1E cells were grown on either D-xylose, D-glucose, Avicel® cellulose or beechwood xylan with subsequent MS-based full proteome analysis to identify differentially abundant glycoside hydrolases based on LFQ (**Figure 4** and **Supplementary Table 1**).

Statistical evaluation of the identified proteins showed the strong alteration of protein abundances upon growth of *Thermococcus* sp. strain 2319x1E on beechwood xylan compared to D-glucose (**Figure 4A**), Avicel® cellulose (**Figure 4B**), and D-xylose (**Figure 4C**). The abundance of 107 proteins was significantly upregulated with a $\log_2$-fold change > 1 in the proteome of xylan-grown cells compared to cells grown on D-glucose (200 proteins compared to cellulose, 208 proteins compared to D-xylose) whereas 90 proteins were significantly downregulated with a $\log_2$-fold change < -1 (114 proteins compared to cellulose and D-xylose). Among the proteins that showed a differential abundance in response to growth on the different sugars, we found several CAZymes (for overview see **Figure 4D** and **Supplementary Table 2** in **Supplementary Data Sheet 1**). The α-specific CAZymes EGIDFPOO_00018 (GH13), EGIDFPOO_00375 (GH57), EGIDFPOO_00563 (GT35), EGIDFPOO_00674 (GH57) EGIDFPOO_01845 (GH57), and EGIDFPOO_01849 (GH13) were all found to be more abundant in cells grown on xylan compared to the other tested sugars, although this difference was not significant for EGIDFPOO_00674 compared to growth on D-xylose as well as for EGIDFPOO_01849 compared to growth on D-glucose. EGIDFPOO_00532, a predicted β-glucosidase, was less abundant in cells grown on xylan when compared to cells grown on D-glucose, but more abundant compared to cells grown on Avicel® cellulose or D-xylose. Moreover, the GH122 family hydrolase EGIDFPOO_00753 was slightly upregulated in xylan-grown cells compared to D-glucose and Avicel® cellulose, but slightly downregulated compared to D-xylose. Noteworthy, the genes for the two hydrolases *EGIDFPOO_01845* (GH57) and *EGIDFPOO_01849* (GH13) are in close genetic neighborhood to a predicted ATP-binding cassette (ABC) transporter comprising the genes

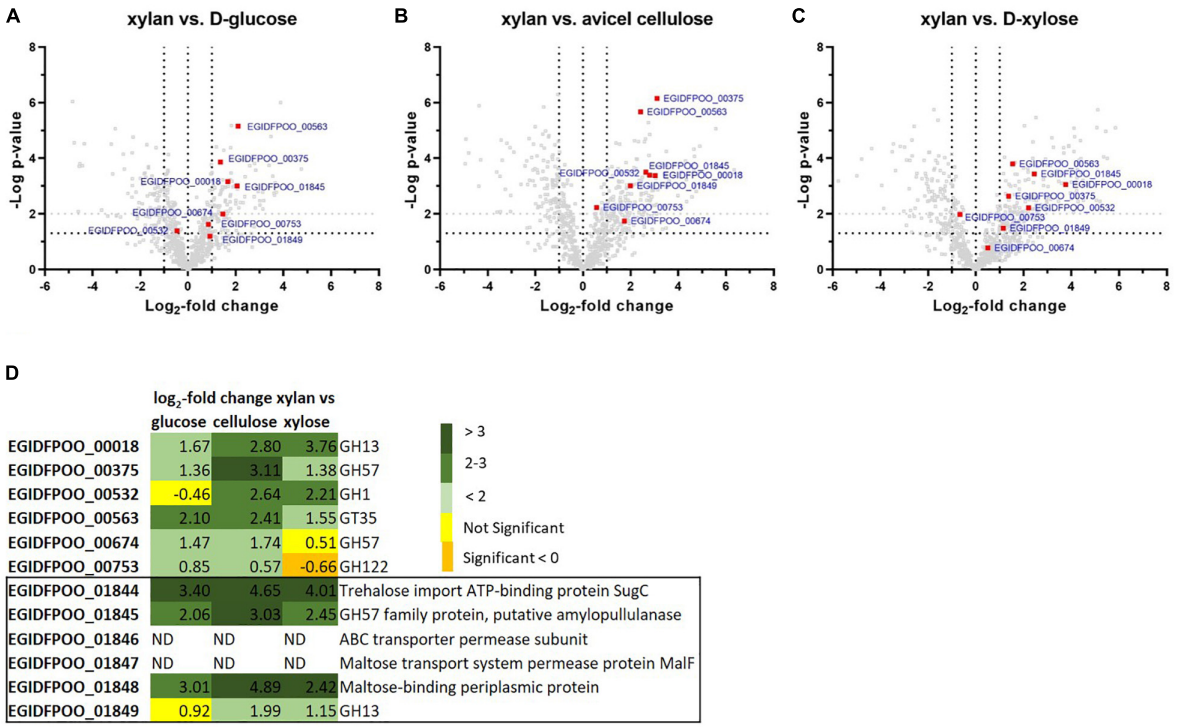


FIGURE 4 | Differentially abundant proteins in *Thermococcus* sp. strain 2319x1E cells grown on different carbon sources. The protein abundance profile of cells grown on beechwood xylan was compared with that of cells grown either on D-glucose **(A)**, Avicel® cellulose **(B)** or D-xylose **(C)**. The volcano plots show the *p*-value against the $\log_2$-fold change in protein abundance. Statistical significance was calculated based on a two-sample Student's *t*-test. Predicted CAZymes are depicted as red squares. Each treatment group comprised four biological replicates. **(D)** Table of differentially abundant CAZymes or selected sugar transport proteins with their respective $\log_2$-fold change and annotation. Yellow fields indicate non-significantly regulated proteins, orange fields indicate significantly downregulated proteins ($p < 0.05$), whereas green fields indicate significantly upregulated proteins ($p < 0.05$) with a $\log_2$-fold change > 1 (pale green: 1–2; moderate green: 2–3; dark green: >3). Proteins that were absent in the filtered data are displayed as ND (not detected). The two GHs EGIDFPOO_01845 and EGIDFPOO_01849 belong to a gene cluster encoding further sugar transport proteins that is boxed.

EGIDFPOO_01846 and EGIDFPOO_01847 (both permeases), as well as to EGIDFPOO_01844 and EGIDFPOO_01848, which are predicted to encode for an ATP binding protein and a putative sugar binding protein, respectively. These proteins are homologous to subunits of an ABC transporter for maltose, trehalose and probably a number of other mono- and disaccharides (Greller et al., 1999). The abundance of this gene cluster is upregulated in cells grown on xylan compared to cells grown on Avicel® cellulose, D-glucose or D-xylose. However, the putatively membrane bound sugar transport proteins EGIDFPOO_01846 and EGIDFPOO_01847 were removed from the data due to the filtering criteria applied (see methods; **Supplementary Table 1**). The $\log_2$-fold changes for all candidate proteins are reported (**Figure 4D**).

Identification of Active Enzymes Involved in Xylan Degradation by Activity-Based Protein Profiling

To shift the focus on the detection of active glycosidases involved in xylan degradation in *Thermococcus* sp. strain 2319x1E, we screened the two well-established biotin-tagged cyclophellitol aziridine-based glycosidase probes JJB384 and JJB111 (**Figure 5A**) for their ability to label active enzymes *in vivo* upon growth on xylan. Thereto, *Thermococcus* sp. strain 2319x1E cells grown on xylan were incubated with 2 μM of the respective probe for 2 h, followed by the detection of labeled proteins via Western blot analysis (**Figure 5B**). This revealed a distinctive band pattern for the samples labeled with JJB384 with two prominent bands at a MW around 70 kDa. While the upper band is absent in the DMSO control, there is a band in the control sample at about the same height as the lower band in the labeled sample, but considerably less pronounced. In contrast, only faint bands are visible for the labeling with JJB111. Furthermore, the inhibitory effect of both ABPs on the xylanolytic activity was tested. Crude extract of *Thermococcus* sp. strain 2319x1E cells grown on xylan was incubated under the conditions previously described for the *in vivo* labeling, after which activity determination was performed via the DNSA assay. Despite only a faint labeling was obtained from the cells incubated with JJB111 in the Western blot analysis, the hydrolytic activity of the crude extract on xylan was decreased to 49% by preincubation with the probe. However, this effect was distinctly stronger for JJB384, with a remaining activity of about 23% (**Figure 5C**). Consequently, JJB384 was selected for further identification of target enzymes by MS-based proteomics.

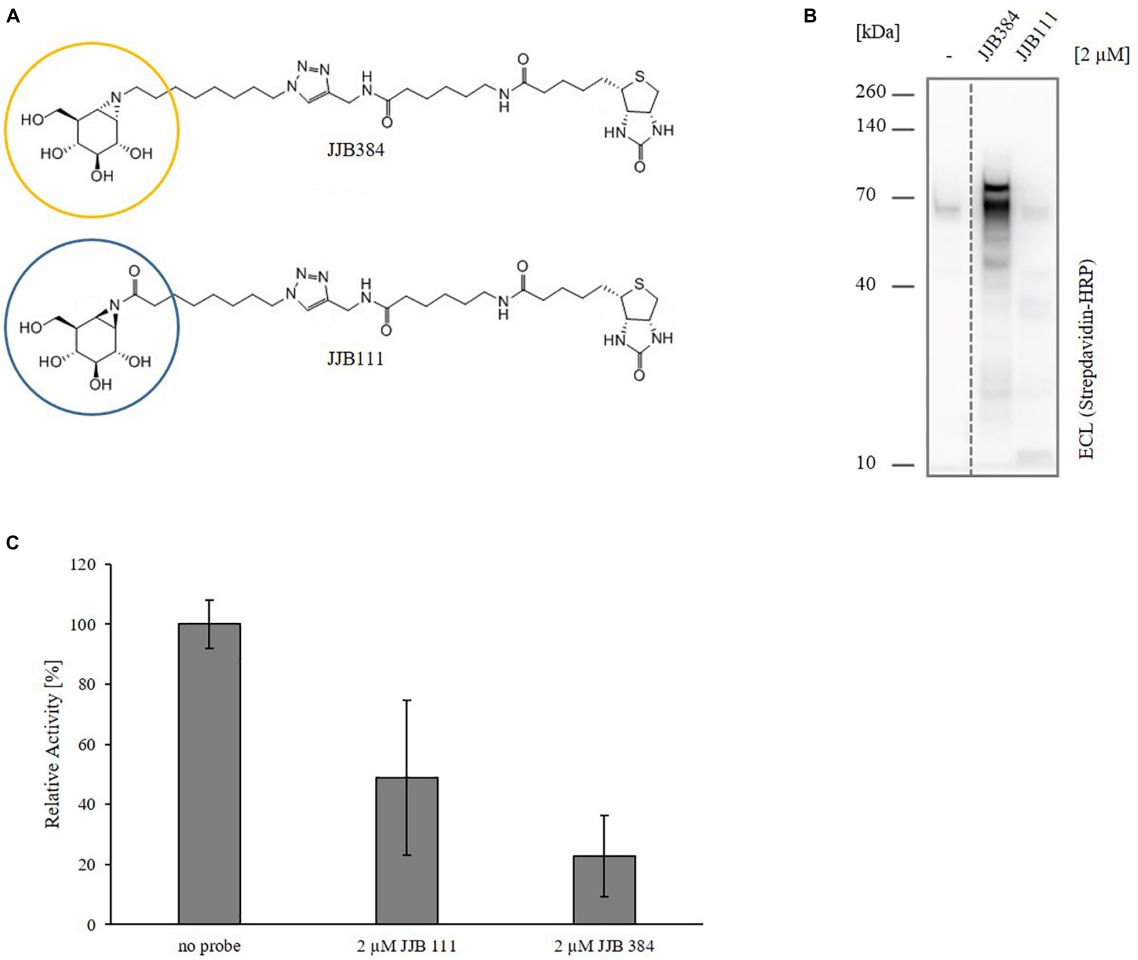


FIGURE 5 | ABPP of *Thermococcus* sp. strain 2319x1E with glycosidase-selective probes. **(A)** Chemical structure of the two employed glycosidase-selective ABPs JJB384 and JJB111. The circles indicate the cyclophellitol aziridine moieties (yellow, α-configured moiety of JJB384; blue, β-configured moiety of JJB111) which function as warheads covalently binding to the target proteins. **(B)** Western blot analysis of the in vivo profiled *Thermococcus* sp. strain 2319x1E proteome of cells grown on xylan after separation via Bis-Tris gel electrophoresis and detection with a streptavidine HRP conjugate. The cells were incubated with 2 µM of either of the two ABPs or with DMSO as control for 2 h at 78°C with constant shaking. **(C)** Xylanolytic activity of *Thermococcus* sp. strain 2319x1E cell extracts (20 µg protein) after incubation in presence or absence of 2 µM of the respective probe for 2 h at 78°C. Inhibition of xylanolytic activity was determined in triplicates.

To confirm that the band pattern obtained by labeling with JJB384 is specific for cells grown on xylan, a comparative *in vivo* ABPP of *Thermococcus* sp. strain 2319x1E grown on either D-xylose or beechwood xylan was conducted (**Figure 6A**). The Western blot analysis shows a differential labeling pattern of proteins from cells grown on the different carbon sources. In contrast to the labeling of cells grown on xylan, no differences in the band pattern between labeled cells and the control could be observed for cells grown on D-xylose. Thus, proteins that were not labeled with JJB384 after growth of cells on D-xylose show specific activity upon growth of cells on xylan and therefore may play a role in xylan degradation.

For identification of the JJB384 probe-labeled proteins, an affinity enrichment with subsequent tryptic on-bead digestion and LC-MS/MS analysis was performed using protein extracts from cells grown on D-xylose and beechwood xylan (**Figures 6B,C** and **Supplementary Table 2**). The initial data were filtered (see methods) to only keep protein groups that were identified in at least two out of three biological replicates of probe-labeled samples per carbon source. For the xylan samples, 122 protein groups were retained, of which 113 protein groups were enriched with the probe compared to the DMSO controls, whereas 68 out of 69 retained protein groups were enriched with JJB384 for the D-xylose samples. Among the xylan samples, 5 GHs (EGIDFPOO_00375, EGIDFPOO_00532, EGIDFPOO_00674, EGIDFPOO_01845, EGIDFPOO_01993) and 1 GT (EGIDFPOO_00563) were enriched with JJB384 (**Figure 6C**). Of these, only EGIDFPOO_01993 was not reported in the comparative full proteome analysis, as this protein did not pass the filtering criteria and was therefore removed from the initial dataset (see methods, **Supplementary Table 1**).

We next had a closer look at these active enzymes that likely participate in xylan degradation. EGIDFPOO_00532 is the only β-specific GH among the identified proteins and is closely related

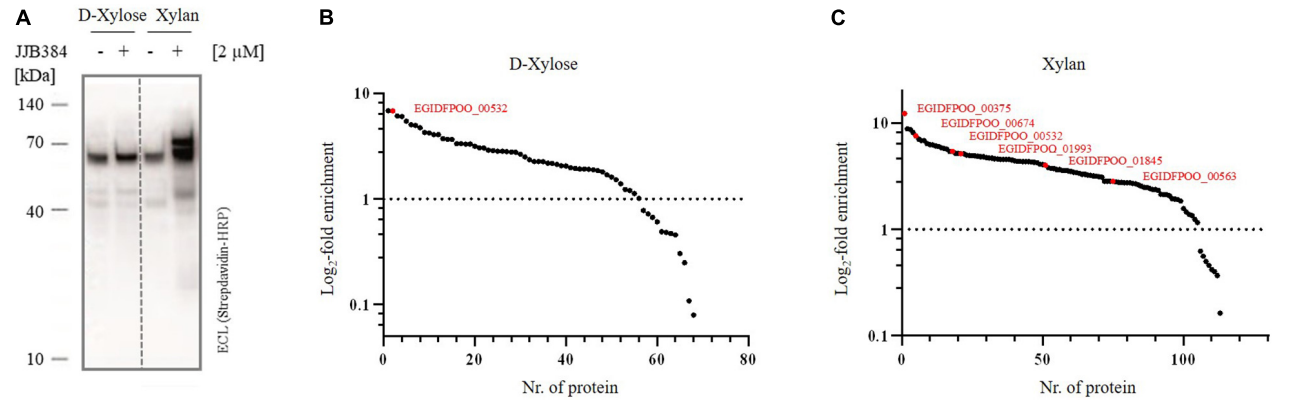


FIGURE 6 | CAZymes identified by ABPP with JJB384. **(A)** Western blot analysis of the *in vivo* profiled *Thermococcus* sp. strain 2319x1E proteome of cells grown on either D-xylose or xylan after separation via Bis-Tris gel electrophoresis and detection with a streptavidine-HRP conjugate. The cells were incubated with 2 µM JJB384 for 2 h at 78°C with constant shaking. **(B,C)** Log_2-fold changes for all proteins enriched with JJB384 in cells grown on D-xylose **(B)** or xylan **(C)** compared to the respective DMSO-controls lacking the probe. Predicted CAZymes are depicted in red. Each treatment group comprised three biological replicates.

to β-glucanases and β-galactosidases from other thermophilic archaea (**Supplementary Figure 4** in **Supplementary Data Sheet 1** and **Supplementary Table 3** in **Supplementary Data Sheet 1**). Notably, it is not only enriched in the xylan samples, but is also the only CAZyme enriched in the D-xylose samples (**Figure 6B**). The respective log_2-fold changes for all identified CAZymes are depicted in **Table 2**. Several α-specific CAZymes were thus only enriched in the xylan samples. The highest enrichment was observed for EGIDFPOO_00375 from the GH57 family (**Supplementary Figure 2** in **Supplementary Data Sheet 1**) that is homologous to the characterized pullulan hydrolase TK-PUL from *T. kodakarensis* (Ahmad et al., 2014), 4-α-glucanotransferase from *T. litoralis* (Jeon et al., 1997) and α-amylase from *P. furiosus* (Laderman et al., 1993). These enzymes have been reported to hydrolyze multiple α-linked polysaccharides, including starch, glycogen, dextrin, amylose, amylopectin, and different cyclodextrins. According to phylogenetic analysis and BLAST with biochemically characterized enzymes, other enriched hydrolase candidates were predicted to function as amylopullulanase (EGIDFPOO_001845, **Supplementary Figure 2** in **Supplementary Data Sheet 1**) or glucan-branching enzyme (EGIDFPOO_01993, **Supplementary Figure 1** in **Supplementary Data Sheet 1**) based on PFAM analysis and sequence homology to biochemically described enzymes (Han et al., 2013; Sun et al., 2015) and should therefore also be specific for hydrolysis of α-linkages in polymers. EGIDFPOO_00563 is highly similar to the GT35 family maltodextrin phosphorylase from *T. litoralis* (97% sequence identity) (Xavier et al., 1999). Notably xylanase activity was reported for the characterized homolog from *T. zilligii* (76% sequence identity) (Uhl and Daniel, 1999), which was, however, later questioned and the respective protein is now annotated as a maltodextrin-phosphorylase in the PDB database (Rolland et al., 2002). EGIDFPOO_00674 is another putative α-specific glycosidase enriched in the xylan samples that belongs to the highly diverse GH57 family and, like EGIDFPOO_00375 and EGIDFPOO_01845, forms an isolated cluster within the phylogenetic tree (**Supplementary Figure 2** in **Supplementary Data Sheet 1**). Unlike the other GH57 family proteins found with the ABPP approach, EGIDFPOO_00674 was not assigned a specific function by automated proteome annotation in Swiss-Prot (Gattiker et al., 2003) and was therefore selected for biochemical characterization along with the putative β-glucosidase EGIDFPOO_00532. The second predicted β-glucosidase EGIDFPOO_01324 (**Table 1**) was neither identified with the comparative full proteome survey nor with the ABPP approach.

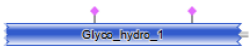
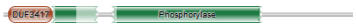
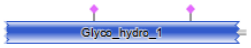
In summary, of the 13 GHs identified in the genome of *Thermococcus* sp. strain 2319x1E (**Table 1**), 7 GHs were shown to be upregulated in response to xylan compared to any of the other sugars, and of these, 5 GHs were additionally shown to be active upon growth on xylan by labeling with JJB384 (**Supplementary Table 2** in **Supplementary Data Sheet 1**). In addition, EGIDFPOO_01993 (predicted glucan-branching enzyme) was only confirmed via ABPP.

Novel Enzyme Activities of Carbohydrate-Active Enzymes Identified by Activity-Based Protein Profiling

Maltose-Forming α-Amylase and Deacetylase Activity of EGIDFPOO_00674

The putative GH57 family protein EGIDFPOO_00674 with uncertain function is potentially involved in xylan degradation as it was enriched with the probe JJB384 upon growth of cells on xylan and was likewise found to be upregulated in the proteome of cells grown on xylan compared to cells grown on D-glucose, Avicel® cellulose or D-xylose. EGIDFPOO_00674 (70.6 kDa, 599 amino acids) contains a GH57 family domain spanning the amino acid residues 52–300, with E153 and D253 acting as catalytic residues, as predicted by HMMER analysis. This protein is conserved in many *Thermococcus* species (see **Table 3**) as revealed by sequence similarity search with BLASTP, but for most

TABLE 2 | Overview of CAZymes enriched from *Thermococcus* sp. strain 2319x1E cultures grown on xylan or D-xylose using the α-selective probe JJB384.

Growth substrate	Gene	Protein family and predicted function	MW [kDa]	Log_2-fold change	Domain architecture (PFAM)
Xylan	*EGIDFPOO_00375*	GH57, **α (1–4)** amylase/ 4-α-glucanotransferase	47.7	12.27	
	EGIDFPOO_00674	GH57, **α**, putative GH57 family enzyme	70.6	7.57	
	EGIDFPOO_00532	GH1, **β (1–4)** glucosidase	49.3	5.39	
	EGIDFPOO_01993	GH13, **α (s1–4)** glucan branching enzyme	82.7	5.16	
	EGIDFPOO_01845	GH57, **α (1–4/6)** amylopullulanase	125.2	4.01	
	EGIDFPOO_00563	GT35, **α (1–4)** glycogen phosphorylase	96.8	2.82	
D-Xylose	*EGIDFPOO_00532*	GT1, **β (1–4)** glucosidase	49.3	6.82	

Log_2-fold changes indicate the difference in LFQ intensity between the labeled samples and the DMSO controls. The respective domain architectures, including predicted active site residues, were obtained with the HMMER function phmmer.

TABLE 3 | Distribution of glycosyl hydrolase homologs (according to BLASTP) of upregulated proteins identified by ABPP or comparative proteomics during growth on xylan in *Thermococcus* sp. strain 2319x1E in other members of the *Thermococcales*.

		Sequence identity of homologs [%]					
Protein family	***T.* sp. strain 2319x1E**	***T.* sp strain 2319x1**	***T. litoralis***	***T. sibiricus***	***T. zilligii***	***P. furiosus***	***T. kodakarensis***
GH1	EGIDFPOO_00532	97.62	92.58	77.62	76.81	81.15	77.94
GT35	EGIDFPOO_00563*	99.64	97.23	87.35	76.03	70.41	77.00
GH57	EGIDFPOO_00674	99.50	97.33	89.98	76.88	73.24	57.79
GH13	EGIDFPOO_00018	98.04	92.42	/	73.65	/	/
GH57	EGIDFPOO_00375	99.39	95.45	86.67	77.29	65.30	/
GH57	EGIDFPOO_01845	99.64	93.00	/	71.82	75.56	77.21
GH13	EGIDFPOO_01849	99.54	89.91	/	55.0	57.48	/
GH13	EGIDFPOO_01993	99.22	/	82.35	/	/	61.62

Enzymes with proposed function in xylan hydrolysis are highlighted in gray.
Xylanolytic activity for the GT35 homolog was reported only for T. zilligii (Uhl and Daniel, 1999), however, as previously suggested (Rolland et al., 2002) and in accordance with our studies, the enzyme possesses glycogen phosphorylase activity (Supplementary Figure 7** in **Supplementary Data Sheet 1**).*

homologs no specific function is assigned, except for the maltose-forming amylase PY04_RS04545 from *Pyrococcus* sp. strain ST04 (Jung et al., 2014), which has a sequence identity of 69% to EGIDFPOO_00674. Several other amylases were identified via HHpred analysis as structural homologs. However, remote structural homologs with assigned functions as mannosidases, glucanotransferases or polysaccharide deacetylases/esterases were also identified (**Supplementary Table 3** in **Supplementary Data Sheet 1**). Of note, the structurally similar regions of the homologs annotated as deacetylases roughly correspond to the GH57 domain region of EGIDFPOO_00674 with its catalytic residues.

To elucidate the precise function of EGIDFPOO_00674, the enzyme was heterologously expressed in *E. coli* and purified (**Supplementary Figure 5** in **Supplementary Data Sheet 1**). The homolog PY04_RS04545 from *Pyrococcus* sp. strain ST04 has been demonstrated to efficiently cleave α-1,6-linked maltose residues from β-cyclodextrins and thus exhibits maltose-forming α-amylase activity (Jung et al., 2014). Indeed, purified EGIDFPOO_00674 showed comparable activity with 6-*O*-α-maltosyl-β-cyclodextrin as substrate (14.1 U mg^{-1} protein, data not shown). Interestingly, deacetylase activity was also observed by using the artificial substrate PNPA, indicating that this enzyme might function as a bifunctional α-amylase/carbohydrate deacetylase. The highest deacetylase activity was detected at a pH of 8.0 (**Figure 7A**) and a temperature of 100°C (**Figure 7B**). Remarkably, EGIDFPOO_00674 retained 44% (100°C) to 72% (80°C) of its deacetylase activity after 4 h of incubation at the respective temperature (**Figure 7C**). A V_{max} of 28.19 U mg^{-1} protein and a K_m of 2.57 mM was determined for the hydrolysis of PNPA by kinetic measurements (**Figure 7D**). Furthermore, the deacetylase activity of EGIDFPOO_00674 was efficiently decreased in a concentration-dependent manner to 46% (2 μM JJB384) and 16% (4 μM JJB384) residual activity by preincubation with the α-glucosidase probe, thus confirming the specific

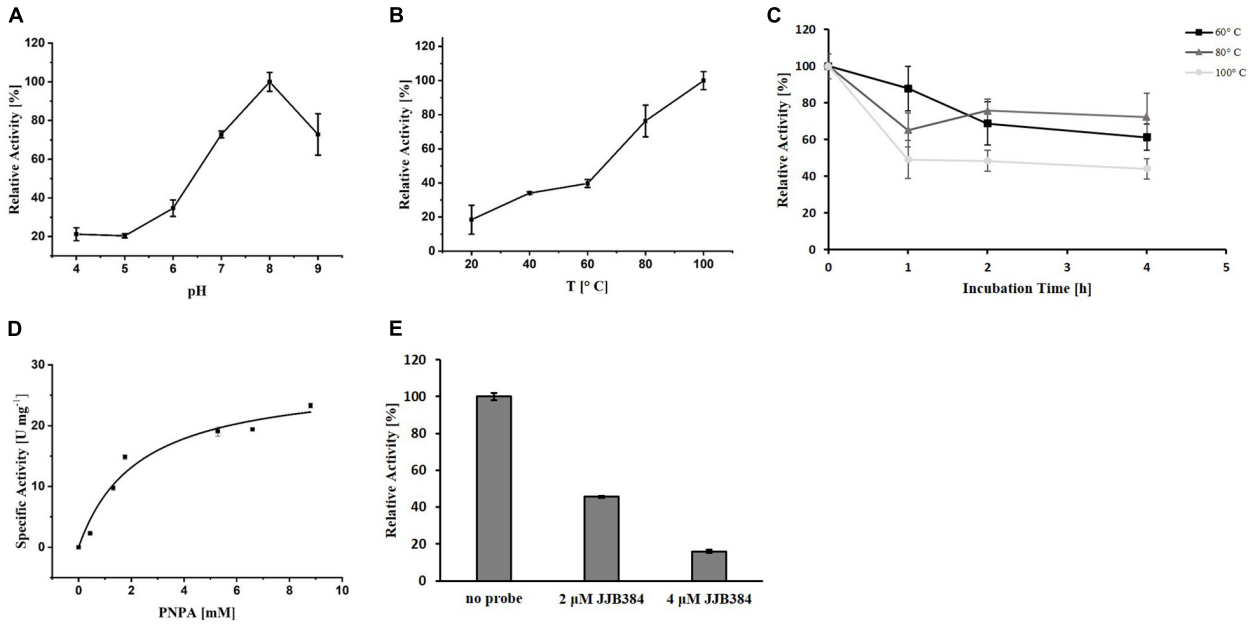


FIGURE 7 | Characterization of heterologously expressed EGIDFPOO_00674 using PNPA as substrate. Effect of pH **(A)** and temperature **(B)** on the deacetylase activity of EGIDFPOO_00674. **(C)** Thermostability of EGIDFPOO_00674. The residual deacetylase activity of EGIDFPOO_00674 after preincubation at 60°C (black line, ■), 80°C (dark gray line, ▲) or 100°C (bright gray line, ●) for up to 4 h was determined. **(D)** Kinetic characterization of EGIDFPOO_00674 using PNPA (0–8.8 mM) as substrate (pH 8.0, 100 °C). **(E)** Inhibition of EGIDFPOO_00674 deacetylase activity by preincubation with JJB384 (2 and 4 μM). The kinetic characterization and the inhibition assays were performed at pH 8.0 and 100°C. Deacetylase activity was determined by measuring the increase in absorbance at 400 nm. All measurements were performed in triplicates.

covalent binding of the ABP to the active site of the enzyme (**Figure 7E**).

Promiscuous β-Glycosidase Activity of EGIDFPOO_00532

EGIDFPOO_00532 (49.32 kDa, 420 amino acids) was predicted as β-glucosidase by Swiss-Prot annotation and is the only β-specific GH in *Thermococcus* sp. strain 2319x1E that was enriched with the probe JJB384. Indeed, a GH1 family domain spanning the amino acids 1–407 (e-value 1.2×10^{-90}) was predicted for this enzyme by HMMER analysis. EGIDFPOO_00532 is conserved in many different species of the *Thermococcales* (see **Table 3**), as determined by BLASTP searches and respective homologs have proposed functions as β-glucosidases or β-galactosidases (**Supplementary Figure 4** in **Supplementary Data Sheet 1**). The majority of close structural homologs are predicted by HHpred to possess either β-glucosidase or β-galactosidase activity, whereas β-xylanase activity was only reported for more remote (probability > 95%) structural homologs (**Supplementary Table 4** in **Supplementary Data Sheet 1**). Although no secretion signals or transmembrane regions were predicted for EGIDFPOO_00532, a close homolog from *P. horikoshii* (80.91% sequence identity) has been characterized as a putatively membrane bound β-glycosidase capable of hydrolyzing different alkyl-β-glycosides (Akiba et al., 2004).

Since EGIDFPOO_00532 was classified as a GH1 family β-glucosidase, its hydrolytic activity was determined using the chromogenic substrate PNPG (β-glucosidase activity), alongside with the further substrates ONPG (β-galactosidase activity) and PNPX (β-xylosidase activity) after heterologous expression of the enzyme in *E. coli* and denaturing purification from inclusion bodies (**Supplementary Figure 6** in **Supplementary Data Sheet 1**). The enzyme possesses broad substrate specificity and showed activity with all three substrates. We first determined the pH and temperature optimum monitoring the β-xylosidase activity of EGDIFPOO_00532. Highest activity with PNPX was observed at a pH of 8.0 (**Figure 8A**) and a temperature of 100°C (**Figure 8B**). The thermostability of EGDIFPOO_00532 was determined by preincubation of the enzyme up to 4 h at 60, 80, and 100°C, revealing 61, 32, and 7% residual enzyme activity, respectively (**Figure 8C**). The kinetic characterization was performed for the three substrates at 100°C and pH 8.0, revealing highest activity for PNPG with a V_{max} of 1.27 U mg^{-1} protein and a K_m of 0.18 mM. Slightly less activity was determined for ONPG (V_{max} = 1.14 U mg^{-1} protein, K_m = 6.23 mM) and roughly half of the activity for PNPX with a V_{max} of 0.53 U mg^{-1} protein and a K_m of 4.38 mM (**Figure 8D**). Accordingly, EGIDFPO_00532 is a promiscuous β-glucosidase with additional β-xylosidase and β-galactosidase activity. The β-xylosidase activity of EGDIFPOO_00532 was furthermore shown to be reduced by preincubation with the α- and β-selective glucosidase probes JJB384 and JJB111, respectively. While the application of 4 μM of the α-selective glucosidase probe JJB384 only resulted in a slight inhibition of the β-xylosidase activity (85% activity compared to the untreated control), a considerable decrease in activity to 38% was observed upon preincubation

with 4 μM of the β-glucosidase probe JJB111 (**Figure 8E**). This is in agreement with the identification of EGDIFPOO_00532 using JJB384, but furthermore confirms nicely the α- and β-preference of both probes.

DISCUSSION

The archaeal metabolism resembles that of Bacteria and lower Eukaryotes in complexity, but Archaea are characterized by unique metabolic features as well as modifications in common metabolic pathways. Their metabolic potential in combination with their prevalence under extreme environmental conditions renders their enzyme repertoire of particular interest for industrial use. Despite years of research, the metabolic complexity of Archaea has yet to be resolved and deeper understanding of their metabolic pathways is only available for a few archaeal model organisms. In order to unravel the metabolic potential of Archaea and their extremozymes for biotechnological applications, alternative approaches can be helpful. ABPP is an attractive methodology in many different fields of research and its application enables to bridge the gap between "native activity measurements" which do not provide information about the proteins that contribute to this activity, and polyomic studies, which lack information about the function of the genes and the activity state of the proteins identified. Although well established for eukaryotic (Niphakis and Cravatt, 2014; Morimoto and van der Hoorn, 2016) and bacterial (Keller et al., 2020) organisms, this methodology has only recently been applied in the Archaea for the first time and turned out to be likewise suitable for the identification of serine hydrolases under the extreme conditions many of the so far cultivated Archaea thrive in (Zweerink et al., 2017).

We herein demonstrated the identification of enzymes involved in xylan degradation in the hyperthermophilic *Thermococcus* isolate strain 2319x1E by ABPP, thereby expanding the applicability of this methodology by the identification of glycoside hydrolases in Archaea. We first showed that the organism, which was isolated from the same enrichment as the previously described isolate strain 2319x1 by *in situ* cultivation techniques (Gavrilov et al., 2016), likewise uses xylan as well as crystalline cellulose as its sole carbon and energy source. The utilization of these carbon sources is rare among the Archaea and has so far only been reported for a few organisms. Archaeal growth in pure culture on crystalline cellulose for example has only been demonstrated for *Desulfurococcus fermentans* (Perevalova et al., 2005) and some haloarchaeal species (Sorokin et al., 2015). Reports of Archaea growing on xylan on the other hand are limited to crenarchaeal species, such as *Thermosphaera aggregans* (Huber et al., 1998), *Sulfolobus solfataricus* (Cannio et al., 2004) and *Acidilobus saccharovorans* (Prokofeva et al., 2009). These two *Thermococcus* strains hence are the first representatives of hyperthermophilic Euryarchaea growing with xylan as growth substrate. Accordingly, the induction of xylanolytic activity, which was measured predominantly in cell membrane fractions, was demonstrated for *Thermococcus* sp. strain 2319x1E. Following these basic analyses, we then decided for a combined (chemo)proteomics approach using a comparative label-free full proteome analysis to detect enzymes involved in xylan degradation as well as ABPP in order to confirm their activity upon growth on xylan. This enabled the identification of a subset of upregulated proteins with activity toward the employed ABP(s). Since JJB384 is glycosidase-selective with a preference for targeting retaining α-glucosidases, it is likely that the identified CAZymes are somehow involved in the hydrolysis of the complex xylan polymer. To confirm this assumption, we further characterized two of the candidates, i.e., EGIDFPOO_00532 (GH1) and EGIDFPOO_00674 (GH57).

Based on our findings, we suggest that EGIDFPOO_00532 (GH1) and EGIDFPOO_00674 (GH57) are involved in xylan degradation, thus contributing to the xylanolytic activity in *Thermococcus* sp. strain 2319x1E cells that was shown to be inducible upon growth on xylan, predominantly localized in cell membrane fractions (**Figure 3**) and targeted by the ABPs JJB384 (preferably targeting α-specific GH) and JJB111 (preferably targeting β-specific GH) (**Figure 5**). Both glycosidases which have been described in this work display high thermostability and retain considerable activity over several hours at 60°C (EGIDFPOO_00532, **Figure 8C**) or even 100°C (EGIDFPOO_00674, **Figure 7C**).

EGIDFPOO_00674 is predicted to belong to the comparatively scarcely described GH57 family, which is frequently found in the genomes of hyperthermophilic Archaea, often occurring jointly with GH13 family enzymes (Janecek et al., 1999). So far, α-amylase, α-galactosidase, amylopullulanase, branching enzyme and 4-α-glucanotransferase activities have been reported for characterized GH57 family members, and it is likely that additional functionalities within this family will be described in the future (Blesák and Janeček, 2012). We could demonstrate that this protein can efficiently cleave α-1,6-linked maltose residues from β-cyclodextrins like it was reported for the closely related homolog from *Pyrococcus* sp. strain ST04 (Jung et al., 2014). Furthermore, the decent hydrolysis of PNPA indicates a potential role as carbohydrate esterase/deacetlyase (**Figure 7**), which cleaves acetyl residues from the xylan backbone and makes it thus more accessible for xylanases and xylosidases, respectively. There are several examples for bifunctional enzymes with two different glycoside hydrolase functions (Shi et al., 2010; Ferrara et al., 2014; Liang et al., 2018). Especially for xylanases, bi- or multifunctionality is common, including acetyl ester-xyloside hydrolases (Khandeparker and Numan, 2008; Cao et al., 2019). However, as far as we know, a bifunctional enzyme with maltose-forming α-amylase and deacetylase activity has not been reported yet. In addition, we suggest that for EGIDFPOO_00674 both functionalities are located at the same site within the GH57 domain or at least in close vicinity to each other, since the deacetylase activity is significantly decreased upon binding of the α-glucosidase-selective ABP JJB384 to this enzyme (**Figure 7**).

For EGIDFPOO_00532 we elucidated promiscuous β-glucosidase activity with significant side activity on galactosides and xylosides (**Figure 8**) contributing to the breakdown of xylan in *Thermococcus* sp. strain 2319x1E. Although no secretion signals or transmembrane domains were predicted for

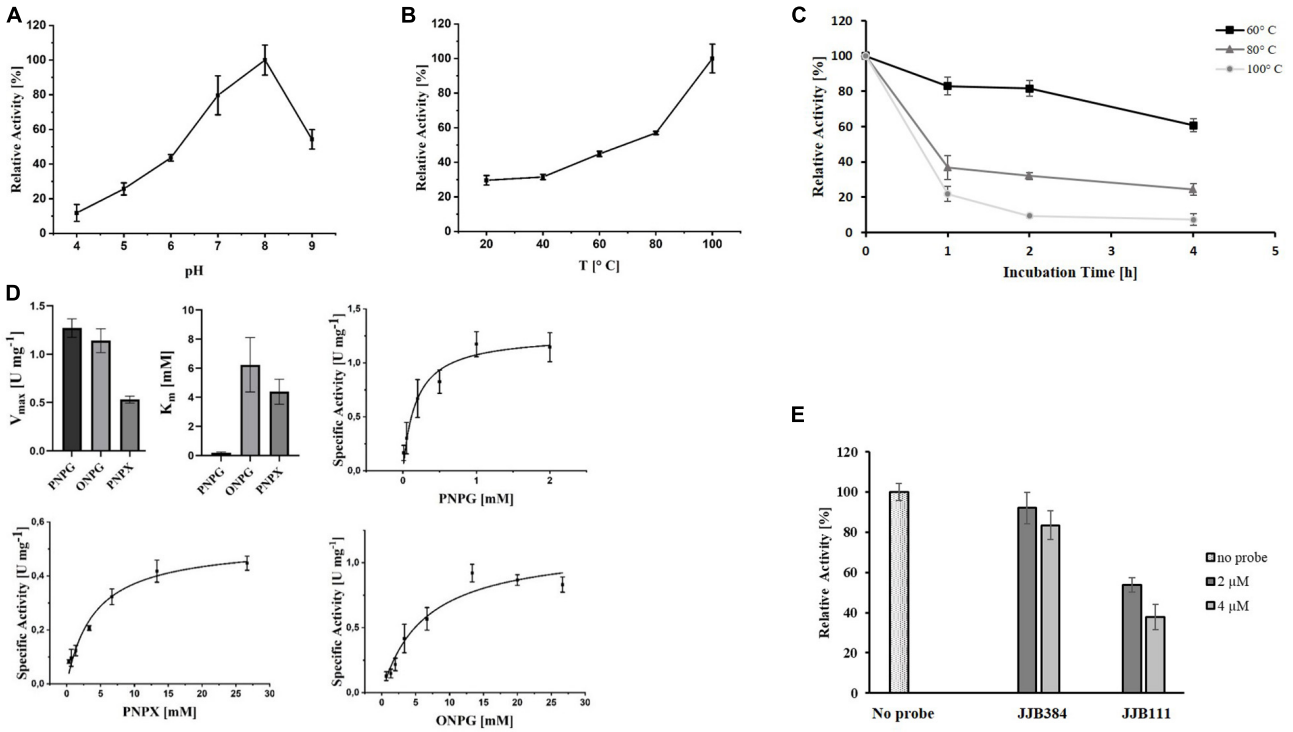


FIGURE 8 | Characterization of heterologously expressed EGIDFPOO_00532. Effect of pH **(A)** and temperature **(B)** on the β-xylosidase activity of EGIDFPOO_00532. **(C)** Thermostability of EGIDFPOO_00532. The residual β-xylosidase activity of EGIDFPOO_00532 after preincubation at 60°C (black line, ■), 80°C (dark gray line, ▲) or 100°C (bright gray line, ●) for up to 4 h was determined. **(D)** Kinetics characterization of EGIDFPOO_00532 with the nitrophenyl substrates PNPG, ONPG and PNPX (pH 8.0, 100°C). **(E)** Inhibition of β-xylosidase activity of EGIDFPOO_00532 with 2 or 4 µM JJB384 and JJB111, respectively.

EGIDFPOO_00532, it appears plausible that EGIDFPOO_00532 is indeed a membrane bound protein, since a close homolog in *P. horikoshii* is described as membrane associated (Matsui et al., 2000; Akiba et al., 2004) and its heterologous expression led to inclusion body formation. Thus, it is potentially exposed to the cell surface and can access large xylan polymers and xylooligosaccharides. While it is common for glycosidases of the GH1 family to possess both β-glucosidase and β-galactosidase activity, β-xylosidase/xylanase activity is unusual for this family (He and Withers, 1997). However, there are examples of other GH1 family hydrolases, which likewise show decent activity with the xylosidase substrate PNPX (Liu et al., 2018; Yin et al., 2020) and promiscuous glycoside hydrolases of other GH families, mainly from the GH family 3, that show β-glucosidase as well as β-xylosidase activity, have also been reported (Zhou et al., 2012; Gruninger et al., 2014; Patel et al., 2018).

Analysis of the prevalence of homologs of the CAZymes discussed in this work (EGIDFPOO_00532 and EGIDFPOO_00674) across other members of the *Thermococcales*, such as *T. kodakarensis*, *T. litoralis,* or *T. sibiricus*, revealed that these enzymes are widely distributed (**Table 3**), suggesting that these strains might be capable of utilizing xylan as carbon source. Furthermore, it remains to be investigated whether and how the other candidates that were identified via ABPP and/or comparative proteomics contribute to the xylanolytic activity of *Thermococcus* sp. strain 2319x1E. For instance, the enzymes EGIDFPOO_00018, EGIDFPOO_00375, EGIDFPOO_00563, EGIDFPOO_01845, EGIDFPOO_01849, and EGIDFPOO_01993 with predicted α-specific functionalities could possibly be involved in the cleavage of (α-linked) side branches of different xylans, such as mannose, arabinofuranose or glucuronic acid. Alternatively, we may have identified these candidates due to the co-occurrence of xylans with other polymers in nature, such as other hemicelluloses, starch or pectin (Tharanathan et al., 1987; Rodrigues Mota et al., 2018). Consistent with this assumption we expressed and purified EGIDFPOO_00563 and demonstrated glycogen phosphorylase activity, i.e., the formation of glucose 1-phosphate from glycogen, maltodextrin and starch in presence of phosphate, but no glycosidase activity was observed with xylan as substrate (**Supplementary Figure 7** in **Supplementary Data Sheet 1**). As mentioned above, this enzyme belongs to the GT 35 family (EC 2.4.1.1) and the homolog in *T. zilligii* (76% sequence identity) was erroneously reported to possess endo-xylanase activity (Uhl and Daniel, 1999; Rolland et al., 2002).

In general, it should be kept in mind that the outcome of such ABPP studies strongly depends on the probes used. The ABPs employed in this study have been reported to preferably target exo-glycosidases (Kallemeijn et al., 2012; Jiang et al., 2016; Husaini et al., 2018) with preference for α-linked (JJB384) or β-linked (JJB111) substrates, which is thus in line with the findings

of this study. However, xylans display a complex structure composed of a diverse set of monosaccharides connected by many different types of glycosidic bonds and thus, a broad and heterogenous set of hydrolytic enzymes is required for its degradation (**Figure 1**). However, we herein clearly demonstrated the benefits of (chemo)proteomics approaches to identify key players involved in polysaccharide degradation. We successfully showed that ABPP with glycoside hydrolase-specific ABPs can be utilized for the functional characterization of CAZymes in (hyper)thermophilic Archaea and thus further extended the scope of ABPP in extremophiles. ABPP therefore represents a promising novel approach for the identification of extremozymes, which may also be of biotechnological interest.

CONCLUSION

We successfully established ABPP with glycoside hydrolase-specific ABPs for the identification of active CAZymes in hyperthermophilic Archaea, here the novel isolate *Thermococcus* sp. strain 2319x1E. The challenging task of identifying novel biocatalysts from (hot) environments has frequently been addressed using (functional) metagenomics (Voget et al., 2003; Ferrer et al., 2016). These approaches, however, lack the option of verifying enzyme activities under physiological conditions. In addition, mesophilic bacteria that possess fundamental differences, for example, in information processing such as transcription (e.g., different eukaryal-like promoter structure and RNA polymerase), translation, post-translational modification (e.g., glycosylation) and protein transport (e.g., signal peptides) are typically used for functional screening. Therefore, this selective approach does not cover the entire biological diversity.

Due to its high abundancy, xylan, for example, is an attractive and cheap resource for various industrial applications in the food, pulp and paper industry and in the production of renewable fuels and chemicals (Butt et al., 2007; Goluguri et al., 2012; Menon and Datta, 2017; Walia et al., 2017), which comes along with ecological and economic benefits (Ayyachamy and Vatsala, 2007; Woldesenbet et al., 2012; Chandel(ed.), 2013; Motta et al., 2013). For biomass conversion, biocatalysts from thermophiles and hyperthermophiles are of particular interest since they generally possess high thermostability, often in combination with high general robustness and stability under harsh reaction conditions, including the presence of detergents and organic solvents (Egorova and Antranikian, 2005; Elleuche et al., 2014; Schocke et al., 2019). Moreover, it is noteworthy that industrial processes at elevated temperatures are often desired due to higher substrate solubility and accessibility as well as a lower risk of contamination (Cokgor et al., 2009). Therefore, the application of alternative screening strategies, such as ABPP in extremophilic environments, represents an alternative approach to address biological diversity and to screen for novel biocatalysts.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/pride/archive/, PXD026056 (Vizcaíno et al., 2016); https://www.ncbi.nlm.nih.gov/, LR778300.

AUTHOR CONTRIBUTIONS

IK, AE, and KZ performed the strain isolation, initial cultivation, and genome analysis. TK performed cultivation of the cells, determination of native enzyme activity, cloning and heterologous expression, purification as well as biochemical characterization of the genes of interest. SN and TK performed *in vivo* chemical labeling, protein extraction and affinity enrichment experiments. SN performed the MS analysis experiments. AA, TB, DW, and JK performed the genome sequencing, assembly and annotation. JJ synthesized ABPs in the HO laboratory. CB, FK, IK, HO, BS, and MK supervised the study. TK and SN wrote the manuscript under supervision of IK, MK, and BS. All authors contributed to the article and approved the submitted version.

FUNDING

BS, MK, and IK acknowledge funding within the DFG-RSF Cooperation for joint German-Russian projects by the DFG (SI 642/12-1, KA 2894/6-1) and RSF (18-44-04024). Furtherly, this work was supported by the DFG (INST 20876/322-1 FUGG, to MK and FK). AE and IK acknowledge support from the Russian Science Foundation Grant Number 20-14-00250.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.734039/full#supplementary-material

REFERENCES

Ahmad, N., Rashid, N., Haider, M. S., Akram, M., and Akhtar, M. (2014). Novel maltotriose-hydrolyzing thermoacidophilic type III pullulan hydrolase from *Thermococcus kodakarensis*. *Appl. Environ. Microbiol.* 80, 1108–1115. doi: 10.1128/AEM.03139-13

Akiba, T., Nishio, M., Matsui, I., and Harata, K. (2004). X-ray structure of a membrane-bound beta-glycosidase from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *Proteins* 57, 422–431. doi: 10.1002/prot.20203

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Ayyachamy, M., and Vatsala, T. M. (2007). Production and partial characterization of cellulase free xylanase by *Bacillus subtilis* C 01 using agriresidues and its application in biobleaching of nonwoody plant pulps. *Lett. Appl. Microbiol.* 45, 467–472. doi: 10.1111/j.1472-765X.2007.02223.x

Bagos, P. G., Tsirigos, K. D., Plessas, S. K., Liakopoulos, T. D., and Hamodrakas, S. J. (2009). Prediction of signal peptides in archaea. *Protein Eng. Des. Sel.* 22, 27–35. doi: 10.1093/protein/gzn064

Biely, P. (1985). Microbial xylanolytic systems. *Trends Biotechnol.* 3, 286–290. doi: 10.1016/0167-7799(85)90004-6

Blesák, K., and Janeček, S. (2012). Sequence fingerprints of enzyme specificities from the glycoside hydrolase family GH57. *Extremophiles* 16, 497–506. doi: 10.1007/s00792-012-0449-9

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1006/abio.1976.9999

Butt, M. S., Tahir-Nadeem, M., Ahmad, Z., and Sultan, M. T. (2007). Xylanases and their applications in baking industry. *Food Technol. Biotechnol.* 46, 22–31.

Cannio, R., Di Prizito, N., Rossi, M., and Morana, A. (2004). A xylan-degrading strain of *Sulfolobus solfataricus*: isolation and characterization of the xylanase activity. *Extremophiles* 8, 117–124. doi: 10.1007/s00792-003-0370-3

Cao, H., Sun, L., Huang, Y., Liu, X., Yang, D., Liu, T., et al. (2019). Structural insights into the dual-substrate recognition and catalytic mechanisms of a bifunctional acetyl ester–xyloside hydrolase from caldicellulosiruptor lactoaceticus. *ACS Catal.* 9, 1739–1747.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chandel, A. (ed.) (2013). *Sustainable Degradation of Lignocellulosic Biomass–Techniques, Applications and Commercialization*. Rijeka: InTech.

Chandrasekar, B., Colby, T., Emran Khan Emon, A., Jiang, J., Hong, T. N., Villamor, J. G., et al. (2014). Broad-range glycosidase activity profiling. *Mol. Cell Proteomics* 13, 2787–2800. doi: 10.1074/mcp.O114.041616

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Cokgor, E. U., Oktay, S., Tas, D. O., Zengin, G. E., and Orhon, D. (2009). Influence of pH and temperature on soluble substrate generation with primary sludge fermentation. *Bioresour. Technol.* 100, 380–386. doi: 10.1016/j.biortech.2008.05.025

Collins, T., Gerday, C., and Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol. Rev.* 29, 3–23. doi: 10.1016/j.femsre.2004.06.005

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics* 13, 2513–2526. doi: 10.1074/mcp.M113.031591

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805. doi: 10.1021/pr101065j

Cravatt, B. F., Wright, A. T., and Kozarich, J. W. (2008). Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* 77, 383–414. doi: 10.1146/annurev.biochem.75.101304.124125

Ebringerová, A., and Heinze, T. (2000). Xylan and xylan derivatives–biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. Macromol. *Rapid Commun.* 21, 542–556. doi: 10.1002/1521-3927(20000601)21:9<542::aid-marc542>3.0.co;2-7

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Egorova, K., and Antranikian, G. (2005). Industrial relevance of thermophilic Archaea. *Curr. Opin. Microbiol.* 8, 649–655. doi: 10.1016/j.mib.2005.10.015

Elleuche, S., Schröder, C., Sahm, K., and Antranikian, G. (2014). Extremozymes–biocatalysts with unique properties from extremophilic microorganisms. *Curr. Opin. Biotechnol.* 29, 116–123. doi: 10.1016/j.copbio.2014.04.003

Ferrara, M. C., Cobucci-Ponzano, B., Carpentieri, A., Henrissat, B., Rossi, M., Amoresano, A., et al. (2014). The identification and molecular characterization of the first archaeal bifunctional exo-β-glucosidase/N-acetyl-β-glucosaminidase demonstrate that family GH116 is made of three functionally distinct subfamilies. *Biochim. Biophys. Acta* 1840, 367–377. doi: 10.1016/j.bbagen.2013.09.022

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnol.* 9, 22–34. doi: 10.1111/1751-7915.12309

Gattiker, A., Michoud, K., Rivoire, C., Auchincloss, A. H., Coudert, E., Lima, T., et al. (2003). Automated annotation of microbial proteomes in SWISS-PROT. *Comput. Biol. Chem.* 27, 49–58. doi: 10.1016/s1476-9271(02)00094-4

Gavrilov, S. N., Stracke, C., Jensen, K., Menzel, P., Kallnik, V., Slesarev, A., et al. (2016). Isolation and characterization of the first xylanolytic hyperthermophilic euryarchaeon *Thermococcus* sp. strain 2319x1 and its unusual multidomain glycosidase. *Front. Microbiol.* 7:552. doi: 10.3389/fmicb.2016.00552

Goluguri, B. R., Thulluri, C., Cherupally, M., Nidadavolu, N., Achuthananda, D., Mangamuri, L. N., et al. (2012). Potential of thermo and alkali stable xylanases from *Thielaviopsis basicola* (MTCC-1467) in biobleaching of wood kraft pulp. *Appl. Biochem. Biotechnol.* 167, 2369–2380. doi: 10.1007/s12010-012-9765-x

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Greller, G., Horlacher, R., DiRuggiero, J., and Boos, W. (1999). Molecular and biochemical analysis of MalK, the ATP-hydrolyzing subunit of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Biol. Chem.* 274, 20259–20264. doi: 10.1074/jbc.274.29.20259

Gruninger, R. J., Gong, X., Forster, R. J., and McAllister, T. A. (2014). Biochemical and kinetic characterization of the multifunctional β-glucosidase/β-xylosidase/α-arabinosidase, Bgxa1. *Appl. Microbiol. Biotechnol.* 98, 3003–3012.

Han, T., Zeng, F., Li, Z., Liu, L., Wei, M., Guan, Q., et al. (2013). Biochemical characterization of a recombinant pullulanase from *Thermococcus kodakarensis* KOD1. *Lett. Appl. Microbiol.* 57, 336–343. doi: 10.1111/lam.12118

Harris, P. J., and Stone, B. A. (2008). "Chemistry and molecular organization of plant cell walls," in *Biomass Recalcitrance*, ed. M. E. Himmel (Oxford: Blackwell Publishing Ltd), 61–93. doi: 10.1002/9781444305418.ch4

He, S., and Withers, S. G. (1997). Assignment of sweet almond beta-glucosidase as a family 1 glycosidase and identification of its active site nucleophile. *J. Biol. Chem.* 272, 24864–24867. doi: 10.1074/jbc.272.40.24864

Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J.-P. (1989). Cellulase families revealed by hydrophobic cluster analysi. *Gene* 81, 83–95. doi: 10.1016/0378-1119(89)90339-9

Heredia, A., Jiménez, A., and Guillén, R. (1995). Composition of plant cell walls. *Z. Lebensm. Unters. Forsch.* 200, 24–31.

Hsieh, Y. S. Y., and Harris, P. J. (2019). Xylans of red and green algae: what is known about their structures and how they are synthesised? *Polymers* 11:354. doi: 10.3390/polym11020354

Huber, R., Dyba, D., Huber, H., Burggraf, S., and Rachel, R. (1998). Sulfur-inhibited *Thermosphaera aggregans* sp. nov., a new genus of hyperthermophilic archaea isolated after its prediction from environmentally derived 16S rRNA sequences. *Int. J. Syst. Bacteriol.* 48, 31–38. doi: 10.1099/00207713-48-1-31

Husaini, A. M., Morimoto, K., Chandrasekar, B., Kelly, S., Kaschani, F., Palmero, D., et al. (2018). Multiplex fluorescent, activity-based protein profiling identifies active α-glycosidases and other hydrolases in plants. *Plant Physiol.* 177, 24–37. doi: 10.1104/pp.18.00250

Janecek, S., Lévêque, E., Belarbi, A., and Haye, B. (1999). Close evolutionary relatedness of alpha-amylases from Archaea and plants. *J. Mol. Evol.* 48, 421–426. doi: 10.1007/pl00006486

Jeon, B. S., Taguchi, H., Sakai, H., Ohshima, T., Wakagi, T., and Matsuzawa, H. (1997). 4-alpha-glucanotransferase from the hyperthermophilic archaeon *Thermococcus litoralis*-enzyme purification and characterization, and gene cloning, sequencing and expression in *Escherichia coli*. *Eur. J. Biochem.* 248, 171–178. doi: 10.1111/j.1432-1033.1997.00171.x

Jiang, J., Kuo, C.-L., Wu, L., Franke, C., Kallemeijn, W. W., Florea, B. I., et al. (2016). Detection of active mammalian GH31 α-glucosidases in health and disease using in-class, broad-spectrum activity-based probes. *ACS Cent. Sci.* 2, 351–358. doi: 10.1021/acscentsci.6b00057

Jung, J.-H., Seo, D.-H., Holden, J. F., and Park, C.-S. (2014). Maltose-forming α-amylase from the hyperthermophilic archaeon *Pyrococcus* sp. ST04. *Appl. Microbiol. Biotechnol.* 98, 2121–2131. doi: 10.1007/s00253-013-5068-6

Juturu, V., and Wu, J. C. (2012). Microbial xylanases: engineering, production and industrial applications. *Biotechnol. Adv.* 30, 1219–1227. doi: 10.1016/j.biotechadv.2011.11.006

Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi: 10.1093/nar/gkm256

Kallemeijn, W. W., Li, K.-Y., Witte, M. D., Marques, A. R. A., Aten, J., Scheij, S., et al. (2012). Novel activity-based probes for broad-spectrum profiling of retaining β-exoglucosidases in situ and in vivo. *Angew. Chem. Int. Ed Engl.* 51, 12529–12533. doi: 10.1002/anie.201207771

Kallnik, V., Bunescu, A., Sayer, C., Bräsen, C., Wohlgemuth, R., Littlechild, J., et al. (2014). Characterization of a phosphotriesterase-like lactonase from the hyperthermoacidophilic crenarchaeon *Vulcanisaeta moutnovskia*. *J. Biotechnol.* 190, 11–17. doi: 10.1016/j.jbiotec.2014.04.026

Keller, L. J., Babin, B. M., Lakemeyer, M., and Bogyo, M. (2020). Activity-based protein profiling in bacteria: applications for identification of therapeutic targets and characterization of microbial communities. *Curr.Opin. Chem. Biol.* 54, 45–53. doi: 10.1016/j.cbpa.2019.10.007

Kevbrin, V. V., and Zavarzin, G. A. (1992). The effect of sulfur compounds on growth of halophilic homoacetic bacterium *Acetohalobium arabaticum*. Mikrobiologiâ (Moskva, 1932), 61, 812–817.

Khandeparker, R., and Numan, M. T. (2008). Bifunctional xylanases and their potential use in biotechnology. *J. Ind. Microbiol. Biotechnol.* 35, 635–644. doi: 10.1007/s10295-008-0342-9

Krogh, A., Larsson, B., Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Laderman, K. A., Davis, B. R., Krutzsch, H. C., Lewis, M. S., Griko, Y. V., Privalov, P. L., et al. (1993). The purification and characterization of an extremely thermostable alpha-amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *J. Biol. Chem.* 268, 24394–24401.

Liang, P.-H., Lin, W.-L., Hsieh, H.-Y., Lin, T.-Y., Chen, C.-H., Tewary, S. K., et al. (2018). A flexible loop for mannan recognition and activity enhancement in a bifunctional glycoside hydrolase family 5. *Biochim. Biophys. Acta Gen. Subj.* 1862, 513–521. doi: 10.1016/j.bbagen.2017.11.004

Liu, C., Zou, G., Yan, X., and Zhou, X. (2018). Screening of multimeric β-xylosidases from the gut microbiome of a higher termite, *Globitermes brachycerastes*. *Int. J. Biol. Sci.* 14, 608–615. doi: 10.7150/ijbs.22763

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Matsui, I., Sakai, Y., Matsui, E., Kikuchi, H., Kawarabayasi, Y., and Honda, K. (2000). Novel substrate specificity of a membrane-bound β-glycosidase from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *FEBS Lett.* 467, 195–200. doi: 10.1016/s0014-5793(00)01156-x

McGinnis, S., and Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 32, W20–W25. doi: 10.1093/nar/gkh435

McIlvaine, T. C. (1921). A buffer solution for colorimetric comparison. *J. Biol. Chem.* 49, 183–186. doi: 10.1016/s0021-9258(18)86000-8

Menon, G., and Datta, S. (2017). "Xylanases: from paper to fuel," in *Microbial Applications*, Vol. 1, eds V. C. Kalia and P. Kumar (Cham: Springer International Publishing), 153–164.

Meyer, F., Goesmann, A., McHardy, A. C., Bartels, D., Bekel, T., Clausen, J., et al. (2003). GenDB–an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 31, 2187–2195. doi: 10.1093/nar/gkg312

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* 31, 426–428.

Morimoto, K., and van der Hoorn, R. A. L. (2016). The increasing impact of activity-based protein profiling in plant science. *Plant Cell Physiol.* 57, 446–461. doi: 10.1093/pcp/pcw003

Motta, F. L., Andrade, C. C. P., and Santana, M. H. A. (2013). "A review of xylanase production by the fermentation of xylan: classification, characterization and applications," in *Sustainable Degradation of Lignocellulosic Biomass–Techniques, Applications and Commercialization*, ed. A. Chandel (Rijeka: InTech).

Niehaus, F., Bertoldo, C., Kaehler, M., and Antranikian, G. (1999). Extremophiles as a source of novel enzymes for industrial application. *Appl. Microbiol. Biotechnol.* 51, 711–729. doi: 10.1007/s002530051456

Niphakis, M. J., and Cravatt, B. F. (2014). Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* 83, 341–377. doi: 10.1146/annurev-biochem-060713-035708

Olsen, J. V., de Godoy, L. M. F., Li, G., Macek, B., Mortensen, P., Pesch, R., et al. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell Proteomics* 4, 2010–2021. doi: 10.1074/mcp.T500030-MCP200

Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B., and Gough, J. (2019). The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res.* 47, D490–D494. doi: 10.1093/nar/gky1130

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Patel, H., Kumar, A. K., and Shah, A. (2018). Purification and characterization of novel bi-functional GH3 family β-xylosidase/β-glucosidase from *Aspergillus niger* ADH-11. *Int. J. Biol. Macromol.* 109, 1260–1269.

Peña, M. J., Kulkarni, A. R., Backe, J., Boyd, M., O'Neill, M. A., and York, W. S. (2016). Structural diversity of xylans in the cell walls of monocots. *Planta* 244, 589–606. doi: 10.1007/s00425-016-2527-1

Perevalova, A. A., Svetlichny, V. A., Kublanov, I. V., Chernyh, N. A., Kostrikina, N. A., Tourova, T. P., et al. (2005). *Desulfurococcus fermentans* sp. nov., a novel hyperthermophilic archaeon from a Kamchatka hot spring, and emended description of the genus *Desulfurococcus*. *Int. J. Syst. Evol. Microbiol.* 55, 995–999. doi: 10.1099/ijs.0.63378-0

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448

Prade, R. A. (1996). Xylanases: from biology to biotechnology. *Biotechnol. Genet. Eng. Rev.* 13, 101–131. doi: 10.1080/02648725.1996.10647925

Prokofeva, M. I., Kostrikina, N. A., Kolganova, T. V., Tourova, T. P., Lysenko, A. M., Lebedinsky, A. V., et al. (2009). Isolation of the anaerobic thermoacidophilic crenarchaeote *Acidilobus saccharovorans* sp. nov. and proposal of Acidilobales ord. nov., including Acidilobaceae fam. nov. and Caldisphaeraceae fam. nov. *Int. J. Syst. Evol. Microbiol.* 59, 3116–3122. doi: 10.1099/ijs.0.010355-0

Puls, J. (1997). Chemistry and biochemistry of hemicelluloses: relationship between hemicellulose structure and enzymes required for hydrolysis. *Macromol. Symp.* 120, 183–196. doi: 10.1002/masy.19971200119

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906. doi: 10.1038/nprot.2007.261

Rodrigues Mota, T., Matias de Oliveira, D., Marchiosi, R., Ferrarese-Filho, O., and Dantas dos Santos, W. (2018). Plant cell wall composition and enzymatic deconstruction. *AIMS Bioeng.* 5, 63–77.

Rolland, J.-L., Gueguen, Y., Flament, D., Pouliquen, Y., Street, P. F. S., and Dietrich, J. (2002). Comment on "The first description of an archaeal hemicellulase: the xylanase from *Thermococcus zilligii* strain AN1": evidence that the unique N-terminal sequence proposed comes from a maltodextrin phosphorylase. *Extremophiles* 6, 349–350. doi: 10.1007/s00792-001-0258-z

Schocke, L., Bräsen, C., and Siebers, B. (2019). Thermoacidophilic *Sulfolobus* species as source for extremozymes and as novel archaeal platform organisms. *Curr. Opin. Biotechnol.* 59, 71–77. doi: 10.1016/j.copbio.2019.02.012

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shi, P., Tian, J., Yuan, T., Liu, X., Huang, H., Bai, Y., et al. (2010). Paenibacillus sp. strain E18 bifunctional xylanase-glucanase with a single catalytic domain. *Appl. Environ. Microbiol.* 76, 3620–3624. doi: 10.1128/AEM.00345-10

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., and Panda, A. K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb. Cell Fact.* 14:41. doi: 10.1186/s12934-015-0222-8

Sorokin, D. Y., Toshakov, S. V., Kolganova, T. V., and Kublanov, I. V. (2015). Halo(natrono)archaea isolated from hypersaline lakes utilize cellulose and chitin as growth substrates. *Front. Microbiol.* 6:942. doi: 10.3389/fmicb.2015.00942

Speers, E. A., and Cravatt, B. F. (2004). Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.* 11, 535–546. doi: 10.1016/j.chembiol.2004.03.012

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sun, Y., Lv, X., Li, Z., Wang, J., Jia, B., and Liu, J. (2015). Recombinant cyclodextrinase from *Thermococcus kodakarensis* KOD1: expression, purification, and enzymatic characterization. *Archaea* 2015:397924. doi: 10.1155/2015/397924

Szabo, Z., and Pohlschroder, M. (2012). Diversity and subcellular distribution of archaeal secreted proteins. *Front Microbiol.* 3:207. doi: 10.3389/fmicb.2012.00207

Tharanathan, R. N., Muralikrishna, G., Salimath, P. V., and Rao, M. R. (1987). Plant carbohydrates—an overview. *Proc. Plant Sci.* 97, 81–155. doi: 10.1007/bf03053322

Thomas, L., Joseph, A., Singhania, R. R., Patel, A. K., and Pandey, A. (2017). "Industrial enzymes," in *Current Developments in Biotechnology and Bioengineering. 6–Industrial Enzymes: Xylanases*, eds L. Thomas, A. Joseph, R. R. Singhania, A. K. Patel, and A. Pandey (Amsterdam: Elsevier), 127–148.

Tomazetto, G., Hahnke, S., Wibberg, D., Pühler, A., Klocke, M., and Schlüter, A. (2018). *Proteiniphilum saccharofermentans* str. M3/6T isolated from a laboratory biogas reactor is versatile in polysaccharide and oligopeptide utilization as deduced from genome-based metabolic reconstructions. *Biotechnol. Rep.* 18:e00254. doi: 10.1016/j.btre.2018.e00254

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319. doi: 10.1038/nprot.2016.136

Uhl, A. M., and Daniel, R. M. (1999). The first description of an archaeal hemicellulase: the xylanase from *Thermococcus zilligii* strain AN1. *Extremophiles* 3, 263–267. doi: 10.1007/s007920050126

Usov, A. I. (2011). Polysaccharides of the red algae. *Adv. Carbohydr. Chem. Biochem.* 65, 115–217. doi: 10.1016/b978-0-12-385520-6.00004-2

Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.-E., and Streit, W. R. (2003). Prospecting for novel biocatalysts in a soil metagenome. *Appl. Environ. Microbiol.* 69, 6235–6242. doi: 10.1128/AEM.69.10.6235-6242.2003

Walia, A., Guleria, S., Mehta, P., Chauhan, A., and Parkash, J. (2017). Microbial xylanases and their industrial application in pulp and paper biobleaching: a review. *3 Biotech* 7:11. doi: 10.1007/s13205-016-0584-6

Wessel, D., and Flügge, U. I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 138, 141–143. doi: 10.1016/0003-2697(84)90782-6

Wibberg, D., Andersson, L., Tzelepis, G., Rupp, O., Blom, J., Jelonek, L., et al. (2016). Genome analysis of the sugar beet pathogen *Rhizoctonia solani* AG2-2IIIB revealed high numbers in secreted proteins and cell wall degrading enzymes. *BMC Genomics* 17:245. doi: 10.1186/s12864-016-2561-1

Wibberg, D., Price-Carter, M., Rückert, C., Blom, J., and Möbius, P. (2020). Complete genome sequence of ovine *Mycobacterium avium* subsp. paratuberculosis strain JIII-386 (MAP-S/type III) and its comparison to MAP-S/type I, MAP-C, and M. avium complex genomes. *Microorganisms* 9:70. doi: 10.3390/microorganisms9010070

Willems, L. I., Overkleeft, H. S., and van Kasteren, S. I. (2014). Current developments in activity-based protein profiling. *Bioconjug. Chem.* 25, 1181–1191. doi: 10.1021/bc500208y

Woldesenbet, F., Virk, A. P., Gupta, N., and Sharma, P. (2012). Effect of microwave irradiation on xylanase production from wheat bran and biobleaching of eucalyptus kraft pulp. *Appl. Biochem. Biotechnol.* 167, 100–108. doi: 10.1007/s12010-012-9663-2

Wolin, E. A., Wolin, M. J., and Wolfe, R. S. (1963). Formation of methane by bacterial extracts. *J. Biol. Chem.* 238, 2882–2886. doi: 10.1016/S0016-0032(13)90081-8

Wu, L., Armstrong, Z., Schröder, S. P., de Boer, C., Artola, M., Aerts, J. M., et al. (2019). An overview of activity-based probes for glycosidases. *Curr. Opin. Chem. Biol.* 53, 25–36. doi: 10.1016/j.cbpa.2019.05.030

Xavier, K. B., Peist, R., Kossmann, M., Boos, W., and Santos, H. (1999). Maltose metabolism in the hyperthermophilic archaeon *Thermococcus litoralis*: purification and characterization of key enzymes. *J. Bacteriol.* 181, 3358–3367. doi: 10.1128/JB.181.11.3358-3367.1999

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47, W52–W58. doi: 10.1093/nar/gkz333

Yin, Y.-R., Sang, P., Xiao, M., Xian, W.-D., Dong, Z.-Y., Liu, L., et al. (2020). Expression and characterization of a cold-adapted, salt- and glucose-tolerant GH1 β-glucosidase obtained from *Thermobifida halotolerans* and its use in sugarcane bagasse hydrolysis. *Biomass Conv. Bioref.* 11:10.

Zhang, H., Yohe, T., Le, H., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418

Zhou, J., Bao, L., Chang, L., Liu, Z., You, C., and Lu, H. (2012). Beta-xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans. *Lett. Appl. Microbiol.* 54, 79–87. doi: 10.1111/j.1472-765X.2011.03175.x

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A Completely reimplemented MPI bioinformatics Toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007

Zweerink, S., Kallnik, V., Ninck, S., Nickel, S., Verheyen, J., Blum, M., et al. (2017). Activity-based protein profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nat. Commun.* 8:15352. doi: 10.1038/ncomms15352

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Klaus, Ninck, Albersmeier, Busche, Wibberg, Jiang, Elcheninov, Zayulina, Kaschani, Bräsen, Overkleeft, Kalinowski, Kublanov, Kaiser and Siebers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Activity-based protein profiling for the identification of novel carbohydrate-active enzymes involved in xylan degradation in the hyperthermophilic Euryarchaeon *Thermococcus* sp. strain 2319x1E

Thomas Klaus[1#], Sabrina Ninck[2#], Andreas Albersmeier[3] Tobias Busche[3], Daniel Wibberg[3], Jianbing Jiang[4§], Alexander G. Elcheninov[5], Kseniya S. Zayulina[5], Farnusch Kaschani[2], Christopher Bräsen[1], Herman S. Overkleeft[4], Jörn Kalinowski[3], Ilya V. Kublanov[5], Markus Kaiser[2], Bettina Siebers[1]

[1]Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Faculty of Chemistry, Centre for Water and Environmental Research (CWE), University of Duisburg-Essen, Essen, Germany

[2]Chemical Biology, Center of Medical Biotechnology, Faculty of Biology, University of Duisburg-Essen, Germany

[3]Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

[4]Bio-organic synthesis, Leiden Institute of Chemistry, University of Leiden, Leiden, Netherlands

[5]Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia

Corresponding authors: Bettina Siebers and Markus Kaiser

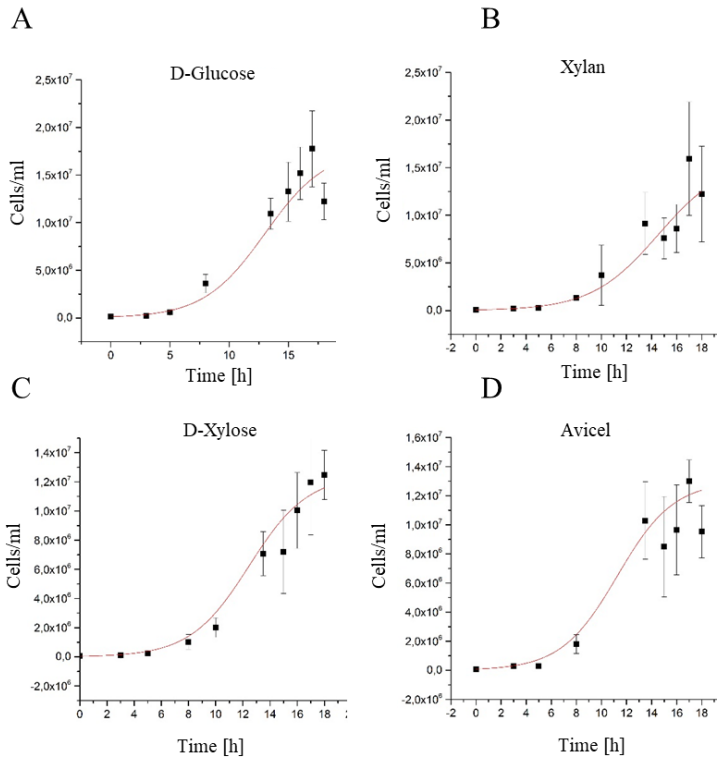
#both authors contributed equally

§current address: Health Science Center, School of Pharmacy, Shenzhen University, Shenzhen, China

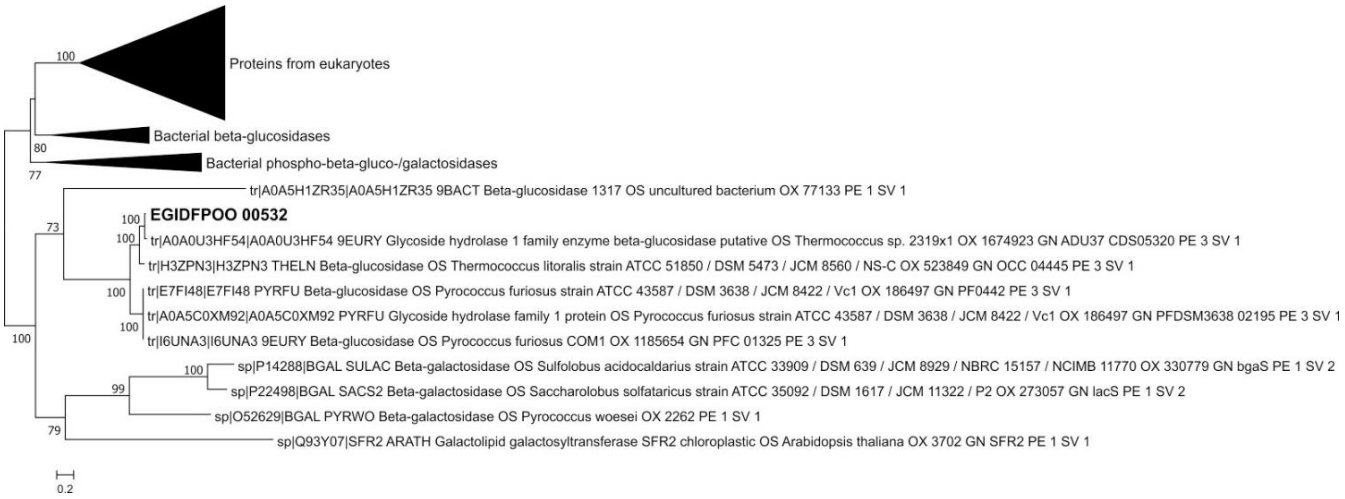
1 Supplementary Figures

Supplementary Figure 1 ‖ Phylogenetic position of the GH13 family enzymes EGIDFPOO_00018, EGIDFPOO_01849 and EGIDFPOO_01993 from *Thermococcus* sp. strain 2319x1E, top 5 BLAST homologs and all proteins from the Swiss-Prot database (evidence at protein level only) affiliated to the GH13 family.

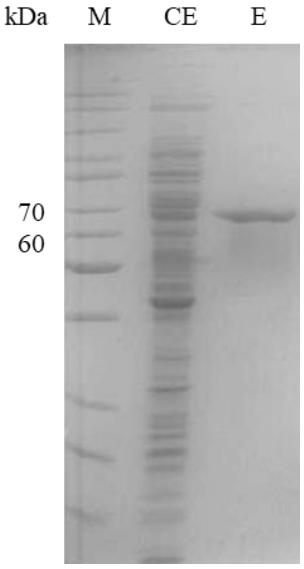
Supplementary Figure 2 ‖ Phylogenetic position of the GH57 family enzymes EGIDFPOO_00375, EGIDFPOO_00674 and EGIDFPOO_01845 from *Thermococcus* sp. strain 2319x1E, top 5 BLAST homologs and all proteins from the Swiss-Prot database (evidence at protein level only) affiliated to the GH57 family.



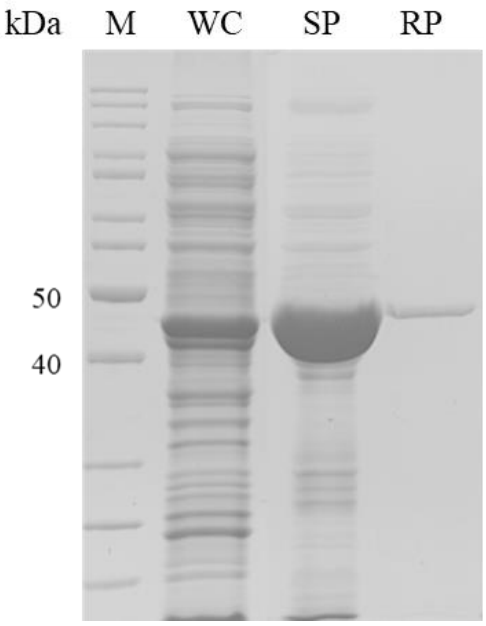
Supplementary Figure 3 ‖ Growth of *Thermococcus* sp. strain 2319x1E on D-glucose (A), xylan (B), D-xylose (C) or Avicel (D) as carbon source. Modified pfennig medium at pH 7.0 with 0.1 g l^{-1} yeast extract and 1 g l^{-1} of the respective sugar was used for growth. Cells were stained with DAPI, counted on a Zeiss Axioscope and the resulting cell numbers were fitted in Origin 2019 (OriginLab corporation, USA) using the sigmoidal fit function sfit1.



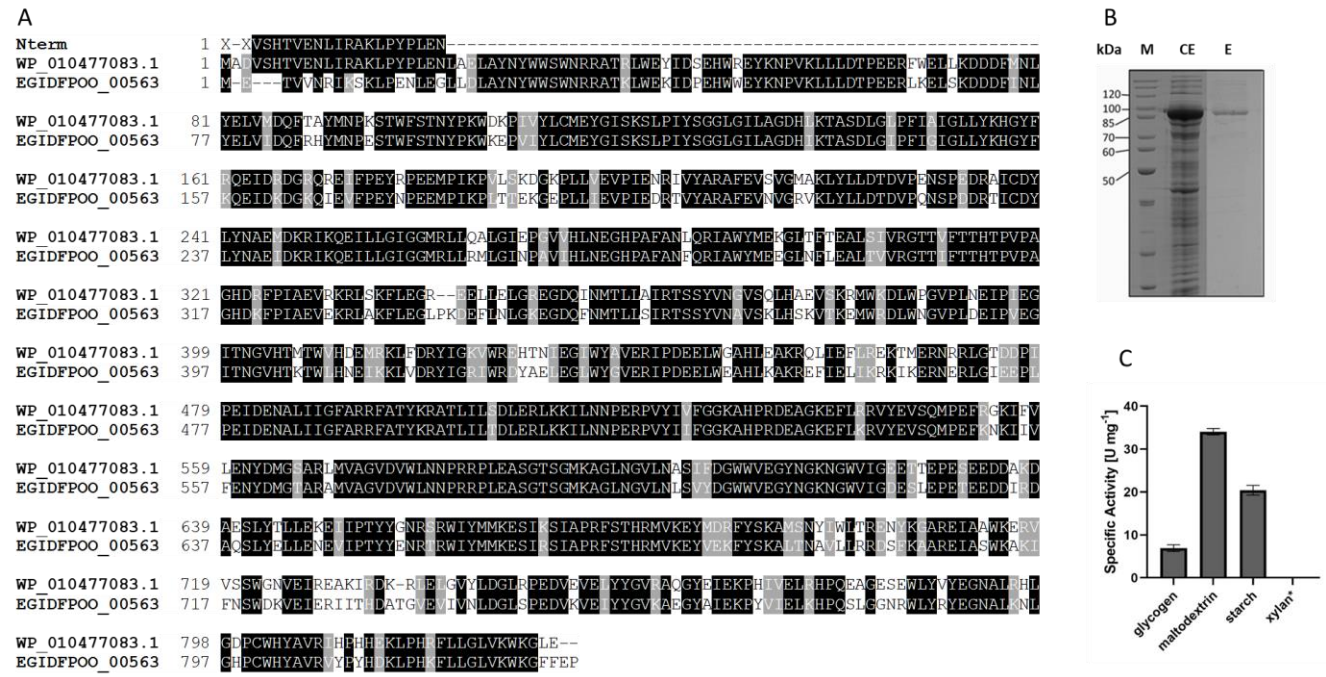
Supplementary Figure 4 ‖ Phylogenetic position of the GH1 family enzyme EGIDFPOO_00532 from *Thermococcus* sp. strain 2319x1E, top 5 BLAST homologs and all proteins from the Swiss-Prot database (evidence at protein level only) affiliated to the GH1 family.



Supplementary Figure 5 ‖ Purification of EGIDFPOO_00674 after heterologous expression in *E. coli*. The recombinant protein was purified from the soluble crude cell extract (CE) by immobilized metal affinity chromatography (IMAC). A protein band corresponding to a size of ~70 kDa is present in the elution fraction (E). (M) Marker, PageRuler^{TM} Unstained Protein Ladder.



Supplementary Figure 6 ‖ Purification of EGIDFPOO_00532 from inclusion bodies after heterologous expression in *E. coli*. The protein was solubilized from the pellet fraction obtained after sonication with a buffer containing 2 M urea at pH 12.5. The solubilized proteins (SP) were refolded by dilution into a pH 8.0 buffer containing 10% (w/v) sucrose and 2 M urea and were subjected to a heat precipitation at 80° C for 20 min. The refolded protein (RP) has a size of about ~45 kDa. (WC): whole cell extract; (M), Marker PageRuler^{TM} Unstained Protein Ladder.



Supplementary Figure 7 ‖ Analysis of the GT35 family maltodextrin phosphorylase EGIDFPOO_00563. **(A)** Multiple Sequence Alignment generated with the N-terminal protein Sequence of the first putative archaeal xylanase from *Thermococcus zilligii* (Uhl *et al.*, 1999), the most suitable BLAST hit from the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi, WP_010477083.1) and the proteinsequence of EGIDFPOO_00563. The alignment was constructed with T-Coffee (http://tcoffee.crg.cat/apps/tcoffee/do:regular) and Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Identical corresponding amino acids are marked with a black box, whereas similar amino acids are marked with a grey box. **(B)** Purification of EGIDFPOO_00563 after heterologous expression in *E. coli*. The recombinant protein was purified from the soluble crude cell extract (CE) by heat precipitation (80° C, 20 min) and immobilized metal affinity chromatography (IMAC). A protein band corresponding to a size of ~100 kDa is present in the elution fraction (E). (M) Marker, PageRuler^{TM} Unstained Protein Ladder. **(C)** Specific activities of EGIDFPOO_00563 for the substrates glycogen, maltodextrin, starch and xylan. The phosphorylase activity with glycogen, maltodextrin and starch was determined by quantification of glucose 1-phosphate using phosphoglucomutase and glucose-6-phosphate dehydrogenase as auxiliary enzymes. Endo-xylanase activity (xylan*) was determined using the DNSA assay, however, no formation of reducing sugars was observed.

2 Supplementary Tables

Supplementary Table 1 ‖ Comparison of the CAZymes present in the genomes of *Thermococcus* sp strain 2319x1 and strain 2319x1E.

CAZy domain	Function Swissprot	strain_2319x1	strain 2319x1E	Identity
GT55		lcl|CP012200.1_prot_ALV61890.1_191	EGIDFPOO_00184	100
		lcl|CP012200.1_prot_ALV63315.1_1616	EGIDFPOO_01570	99.808
GT35	Maltodextrin phosphorylase	lcl|CP012200.1_prot_ALV62263.1_564	EGIDFPOO_00563	99.639
GH57	amylopullulanase	lcl|CP012200.1_prot_ALV63593.1_1894	EGIDFPOO_01845	99.637
		lcl|CP012200.1_prot_ALV62017.1_318	EGIDFPOO_00304	99.553
GH13	neopullulanase	lcl|CP012200.1_prot_ALV63597.1_1898	EGIDFPOO_01849	99.542
CE1	putative carbohydrate esterase	lcl|CP012200.1_prot_ALV63041.1_1342	EGIDFPOO_01302	99.517
GH57		lcl|CP012200.1_prot_ALV62358.1_659	EGIDFPOO_00674	99.499
GT4		lcl|CP012200.1_prot_ALV63230.1_1531	EGIDFPOO_01440	99.43
GT39		lcl|CP012200.1_prot_ALV62042.1_343	EGIDFPOO_00332	99.417
GH57	alpha-amylase/ alpha-gluconotranspherase	lcl|CP012200.1_prot_ALV62084.1_385	EGIDFPOO_00375	99.393
GH57	1,4alpha-branching enzyme	lcl|CP012200.1_prot_ALV62996.1_1297	EGIDFPOO_01266	99.388
GH130	beta-1,4 mannooligosaccharide phosphorilase	lcl|CP012200.1_prot_ALV61989.1_290	EGIDFPOO_00274	99.286
GH13		lcl|CP012200.1_prot_ALV63752.1_2053	EGIDFPOO_01993	99.219
CE10		lcl|CP012200.1_prot_ALV62685.1_986	EGIDFPOO_00955	99.211
GT66		lcl|CP012200.1_prot_ALV61904.1_205	EGIDFPOO_00197	99.016
GH57		lcl|CP012200.1_prot_ALV62233.1_534	EGIDFPOO_00534	98.626
GT2		lcl|CP012200.1_prot_ALV62109.1_410	EGIDFPOO_00399	98.611
GT5		lcl|CP012200.1_prot_ALV63599.1_1900	EGIDFPOO_01851	98.462
GT2		lcl|CP012200.1_prot_ALV61710.1_11	EGIDFPOO_00012	98.295
GH13	alpha-amylase	lcl|CP012200.1_prot_ALV61716.1_17	EGIDFPOO_00018	98.044
GT2		lcl|CP012200.1_prot_ALV63648.1_1949	EGIDFPOO_01900	98.006
CE10		lcl|CP012200.1_prot_ALV62860.1_1161	EGIDFPOO_01219	97.953
GT4		lcl|CP012200.1_prot_ALV62898.1_1199	EGIDFPOO_01181	97.878
GT81		lcl|CP012200.1_prot_ALV63409.1_1710	EGIDFPOO_01674	97.817
GH1	putative beta-glucosidase	lcl|CP012200.1_prot_ALV62231.1_532	EGIDFPOO_00532	97.619
GH122	alpha-glucosidase	lcl|CP012200.1_prot_ALV62474.1_775	EGIDFPOO_00753	97.586
GT2		lcl|CP012200.1_prot_ALV62310.1_611	EGIDFPOO_00612	97.5
GT2		lcl|CP012200.1_prot_ALV63858.1_2159	EGIDFPOO_02198	95.076
GT4		lcl|CP012200.1_prot_ALV63856.1_2157	EGIDFPOO_02196	93.333
GT66		lcl|CP012200.1_prot_ALV62046.1_347	EGIDFPOO_00336	90.369
		lcl|CP012200.1_prot_ALV63429.1_1730	EGIDFPOO_01702	72.784
GT4		lcl|CP012200.1_prot_ALV63888.1_2189	EGIDFPOO_02119	71.525
GT66		lcl|CP012200.1_prot_ALV63874.1_2175	EGIDFPOO_02109	37.951
GT2		cl|CP012200.1_prot_ALV63703.1_2004		
	putative glycogen debranching enzyme	lcl|CP012200.1_prot_ALV61988.1_289		
GH35	exo-beta D-glucosaminidase	lcl|CP012200.1_prot_ALV62426.1_727		
GH1	beta-galactosidase	lcl|CP012200.1_prot_ALV62433.1_734		
GH1	beta-galactosidase	lcl|CP012200.1_prot_ALV62443.1_744		
CE14		lcl|CP012200.1_prot_ALV62446.1_747		
GH13		lcl|CP012200.1_prot_ALV63800.1_2101		
GT2		lcl|CP012200.1_prot_ALV63855.1_2156		
GT4		lcl|CP012200.1_prot_ALV63859.1_2160		
GT4		lcl|CP012200.1_prot_ALV63871.1_2172		
GT2		lcl|CP012200.1_prot_ALV63872.1_2173		
GT4		lcl|CP012200.1_prot_ALV63873.1_2174		
GT2		lcl|CP012200.1_prot_ALV63886.1_2187		
		lcl|CP012200.1_prot_ALV63887.1_2188		
GH5+GH12+H12+CBM2	endoglucanase/endoxylanase	lcl|CP012200.1_prot_ALV63957.1_2258		
GT39		lcl|CP012200.1_prot_ALV63958.1_2259		

CAZy domain	Function Swissprot	strain_2319x1	strain 2319x1E	Identity
	amylo- α -1,6-glucosidase		EGIDFPOO_01323	
GH1	beta-glucosidase		EGIDFPOO_01324	
GT2			EGIDFPOO_02112	
GT4			EGIDFPOO_02121	
GT4			EGIDFPOO_02122	
GT4			EGIDFPOO_02135	
GT4			EGIDFPOO_02136	
GT4			EGIDFPOO_02137	
GT2			EGIDFPOO_02138	
GT2			EGIDFPOO_02139	
GT55			EGIDFPOO_02209	

Supplementary Table 2 ‖ Glycoside hydrolases identified by ABPP and/or comparative proteomics in *Thermococcus* sp. strain 2319x1E with their respective family affiliation, predicted function and reported substrates of characterized homologs. The sequence identity (SI) of the respective homologs to the *Thermococcus* sp. strain 2319x1E proteins is indicated. The proteins which are marked with a grey background have been characterized in this study, revealing promiscuous β-glucosidase activity with PNPG, PNPX and ONPG (EGIDFPOO_00532), maltose-forming α-amylase and deacetylase activity with PNPA (EGIDFPOO_00674) or glycogen phosphorylase activity with glycogen, maltodextrin or starch (EGIDFPOO_00563).

Gene	Identified by		Family	Predicted function (PFAM)	Characterized homolog	Reported substrates
	Comparative proteomics	ABPP				
EGIDFPOO_00018	+	-	GH13	α-amylase		
EGIDFPOO_00375	+	+	GH57	α-amylase/ α-gluconotranspherase	gtpK, *T. kodakarensis*, 78% SI (Ahmad *et al.*, 2014)	Maltooligosaccharides, starch
EGIDFPOO_00532	+	+	GH1	β-glucosidase	BGPh, *T. horikoshii*, 81% SI	β-D-saccharides, β-D-glucosides
EGIDFPOO_00563	+	+	GT35	Glycogen phosphorylase	*T. zilligii* strain AN1, 76% SI (Uhl & Daniel, 1999)*	Larch/oat spelt/birch wood xylan, wheat arabinoxylan,
EGIDFPOO_00674	+	+	GH57	GH family 57 protein	Py04_0872, *P.* sp ST04, 69 % SI (Jung *et al.*, 2014)	6-O-α-maltosyl-β-cyclodextrin,
EGIDFPOO_00753	(+)**	+	GH122	alpha-glucosidase		
EGIDFPOO_01845	+	+	GH57	GH family 57 protein, putative amylopullulanase	Tk1770, *T. kodakarensis*, 55% SI (Han et al. 2013)	
EGIDFPOO_01849	+	-	GH13	GH13 family protein, putative neopullulanase		Pullulan, α-cyclodextrine
EGIDFPOO_01993	-	+	GH13	1,4-α-glucan branching enzyme GlgB	TK_RS04810, *T. kodakarensis*, 62% SI (Sun et al. 2015)	Soluble starch, pullulan

* In accordance with previous reports *(Rolland et al., 2002)* and our studies, the xylanase activity for the *T. zilligii* homolog (Uhl & Daniel, 1999) has erroneously been reported.
**EGIDFPOO_00753 was slightly upregulated in xylan grown cells compared to D-glucose and Avicel® cellulose, but slightly downregulated compared to D-xylose.

Supplementary Table 3 ‖ Structural homologs of EGIDFPOO_00674 predicted with HHpred (Zimmermann et al. 2018). Remote homologs with deacetylase activity are highlighted in grey.

Hit	Function	Origin	Probability [%]	E-value	Score
4CMR_A	glycosyl hydrolase/deacetylase family protein; GH 57 exo-type maltose-forming amylase	*Pyrococcus sp. strain* ST04	100	2.8e-119	1016.49
3N98_A	α-amylase, GH57 family; GH57 family member, branching enzyme, transferase	*Thermococcus kodakarensis*	100	2.4e-31	300.44
5WU7_A	amylase, glycogen branching enzyme	*Pyrococcus horikoshii*	100	3.7e-32	307.05
2B5D_X	α-Amylase; (beta/alpha)7 barrel,	*Thermotoga maritima*	99.97	1.4e-28	275.94
3P0B_A	Glycoside Hydrolase GH57, glycogen branching, transferase	*Thermus thermophilus*	99.97	1.6e-28	276.2
1K1X_B	4-α-glucanotransferase	*Thermococcus litoralis*	99.95	2e-25	256.64
5JM0_A	α-mannosidase	*Saccharomyces cerevisiae* S288C	99.26	2.2e-9	131.46
6B9O_A	α-mannosidase	*Canavalia ensiformis*	99.21	2.2e-9	129.14
2WYH_B	family GH38 α-mannosidase	*Streptococcus pyogenes*	99.08	4.5e-9	126.02
3RXZ_D	Polysaccharide deacetylase	*Mycobacterium smegmatis*	98.68	0.000003	85.65
3S6O_C	Polysaccharide deacetylase family protein	*Burkholderia pseudomallei*	98.51	0.000021	81.24
4LY4_A	peptidoglycan deacetylase	*Helicobacter pylori*	98.37	0.000055	78.65
2CC0_B	Acetyl-xylan esterase	*Streptomyces lividans*	97.88	0.0012	61.56
4M1B_A	Polysaccharide deacetylase; carbohydrate esterase	*Bacillus anthracis*	97.85	0.0018	63.85

Supplementary Table 4 ‖ Structural homologs of EGIDFPOO_00532 predicted with HHpred (Zimmermann et al. 2018) . Remote homologs with xylanase activity are highlighted in grey.

Hit	Function	Origin	Probability [%]	E-value	Score
1VFF_A	β-glucosidase	*Pyrococcus horikoshii*	100	1.1e-47	345.5
4HA4_A	β-galactosidase	*Acidilobus saccharovorans*	100	1e-43	324.26
3WDP_Q	β-glucosidase	*Pyrococcus furiosus*	100	2.2e-43	321.04
1QVB_B	β-glycosidase	*Thermosphaera aggregans*	100	3.3e-43	320.2
2J78_B	β-glucosidase	*Thermotoga maritima*	100	3.8e-43	320.06
5YIF_A	β-galactosidase	*Bacillus sp.*	100	7e-43	318.93
4RE2_A	β-mannosidase	*Oryza sativa Indica*	100	1e-42	319.49
3F5L_B	β-glucosidase	*Oryza sativa Japonica*	100	1e-42	318.17
1PBG_A	β-galactosidase	*Lactococcus lactis*	100	3.2e-42	314.09
6KDC_A	β-glucosidase/galactosidase	*Fervidobacterium pennivorans*	100	2.6e-41	307.38
5OKA_B	β-galactosidase	*Geobacillus stearothermophilus*	100	6.3e-41	306.98
4PMU_B	Endo-1,4-β-xylanase	*Xanthomonas axonopodis pv. citri*	99.94	4.4e-25	194.16
4PMD_A	Endo-1,4-β-xylanase	*Caldicellulosiruptor bescii*	99.94	2.8e-25	193.73
5XZO_A	β-xylanase	*Bispora sp. strain* MEY-1	99.94	5.8e-25	191.6
1UR1_A	β-xylanase	*Cellvibrio mixtus*	99.94	6.5e-25	194.79
3EMZ_A	Endo-1,4-β-xylanase	*Bacillus sp. strain* BP-23	99.94	6.7e-25	190.82
1R85_A	Endo-1,4-β-xylanase	*Geobacillus stearothermophilus*	99.94	1.2e-24	193.06
3NIY_A	Endo-1,4-β-xylanase	*Thermotoga petrophila* RKU-1	99.94	1.3e-24	189.97

Chapter 3.2

Function-based identification of novel enzymes from environmental microbial communities using activity-based protein profiling

Environmental activity-based protein profiling for function-driven enzyme discovery from natural communities

(Submitted manuscript – *bioRxiv*)

Sabrina Ninck[1,#], Thomas Klaus[2,#], Tatiana V. Kochetkova[3], Sarah P. Esser[4], Leonard Sewald[1], Farnusch Kaschani[1], Christopher Bräsen[2], Alexander J. Probst[4], Ilya V. Kublanov[3], Bettina Siebers[2,*], Markus Kaiser[1,*]

1 Chemical Biology, ZMB, Faculty of Biology, Universität Duisburg-Essen, Universitätsstr. 2, 45117 Essen, Germany.

2 Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Centre for Water and Environmental Research (CWE), Faculty of Chemistry, Universität Duisburg-Essen, Universitätsstr. 5, 45117 Essen, Germany.

3 Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Prospekt 60-Let Oktyabrya 7-2, 117312 Moscow, Russia

4 Group for Aquatic Microbial Ecology, UMB, Faculty of Chemistry, Universität Duisburg-Essen, Universitätsstr. 5, 45117 Essen, Germany.

* Corresponding Authors:

Email: bettina.siebers@uni-due.de

markus.kaiser@uni-due.de.

[#] Both authors contributed equally.

Abstract

Microbial communities are significant drivers of global biogeochemical cycles, yet accurate function prediction of their proteome and discerning their activity *in situ* for bioprospecting remains challenging. Here, we present environmental activity-based protein profiling (eABPP) as a novel proteomics-based approach bridging the gap between environmental genomics, correct function annotation and *in situ* enzyme activity. As a showcase, we report the successful identification of active thermostable serine hydrolases by combining genome-resolved metagenomics and mass spectrometry-based eABPP of natural microbial communities from two independent hot springs in Kamchatka, Russia. eABPP does not only advance current methodological approaches by providing evidence for enzyme and microbial activity *in situ* but also represents an alternative approach to sequence homology-guided biocatalyst discovery from environmental ecosystems.

Keywords

Activity-based protein profiling, click chemistry, chemical proteomics, environmental microbial communities, hot springs, metagenomics, metaproteomics, serine hydrolases, target identification

Introduction

Microbial organisms account for the vast majority of the unexplored natural biodiversity[1] and colonize all of the earth's conceivable ecological niches, thereby forming microbial communities of distinct complexity and fluctuating composition[2,3]. Many of the terrestrial and marine environmental microbial communities thriving under moderate to extreme conditions in diverse ecosystems such as soil, freshwater, seawater or hot springs harbor unique genes with (novel) associated functionalities[2]. Their set of enzymes equips them with diverse metabolic activities, some of which yet unknown, that turn these communities into a promising resource for environmental biotechnology[4]. Especially thermostable enzymes that can function as biocatalysts in diverse biotechnological processes are sought with intense demand[5] and have fostered systematic bioprospecting of hydrothermal microbial communities[6-9]. Their metagenomic-based investigation has turned into a feasible biocatalysts discovery approach, in particular for screening organisms or communities that are non-culturable using standard techniques[10,11]. The subsequent functional analysis is commonly done by sequence-driven bioinformatic prediction of protein function[12]. This approach, however, is hampered by a large quantity of proteins of unknown function (i.e hypotheticals) or misannotated enzymes as well as the presence of large protein superfamilies, for which functional predictions are still often insufficient[13]. 'Functional metagenomics' approaches can help to overcome these limitations by complementing sequence-based approaches with an activity-based screening after construction of a metagenomic library[14]. They enable the discovery of novel enzyme classes but require elaborate and challenging cloning and expression efforts with subsequent biochemical characterization[12,15]. Moreover, this approach delivers no information on the expression and, as applies to all 'omics'-based strategies, the activity state of an enzyme-of-interest in its natural environment.

Since its first application more than two decades ago, activity-based protein profiling (ABPP) has emerged into a powerful approach for studying enzyme activities. Within this period, a huge variety of activity-based probes (ABPs) that target different enzymes or even whole enzyme classes have been developed[16-18]. A 'classical' ABP is composed of a reactive warhead, a tag and, if present, a chemical or peptidic linker region. While the warhead is often an electrophile-containing inhibitor molecule that targets a nucleophile at the active site of an enzyme for covalent labeling, the tag is used for target detection or enrichment[19]. Thus, ABPP not only allows the visualization of labeled enzymes via the use of fluorescent reporter groups, but also enables target enzyme identification by mass spectrometry (MS) if a moiety for target

enrichment, such as biotin, is used as a reporter group[20]. Accordingly, an appropriate use of enzyme or enzyme class-specific ABPs enables the functional annotation of enzymes[21]. With the integration of 'click chemistry' into the ABPP workflow, two-step chemical probes have emerged, which facilitate a simple *in vivo* application of ABPP under physiological conditions[22-24].

In the last years, ABPP was mainly used in the context of biomarker or drug discovery as well as *in vivo* imaging[25]; beyond that, diverse applications in micro- or plant biology, including the study of pathogens or host-pathogen interactions, have been frequently reported[26,27]. Only recently, the use of ABPP in biocatalyst discovery for industrial applications, for example for elucidating lignocellulose-degrading enzymes, has emerged[28]. In addition, first reports on the use of ABPP for studying microbial communities, with a focus on host-associated microbial communities, have been made[29]. Mayers *et al*., for example, combined ABPP with a cysteine-reactive probe and quantitative metaproteomics to identify enzymes with an altered expression in inflammatory bowel disease in a mouse model[30]. For the identification of host as well as microbiome proteins from murine fecal samples, a Comprehensive Protein Identification Library (ComPIL) database constructed from publicly available genome data[31] was used. The exploration of environmental microbial communities by ABPP, in contrast, still relies mostly on profiling isolated strains of environmental microbes rather than on the direct profiling of complex communities[32-36]. Recently, an activity-based labeling of ammonia- and alkane-oxidizing bacteria from complex microbial communities using 1,7-octadiyne was reported[37]. This study additionally used metagenomic sequencing for further confirmation of the functional potential of the targeted microorganisms; a generic workflow for ABPP of microbial communities in the environment providing a function-based target identification of the labeled enzymes via downstream mass-spectrometry is however still lacking today.

We herein describe an ABPP approach that allows the functional identification of active enzymes from complex environmental microbial communities (hereafter referred to as environmental ABPP, eABPP). Our approach is based on a combination of ABPP with genome-resolved metagenomics, facilitating the assignment of dedicated activities to enzymes from microorganisms in the environment, even to those that belong to so far uncultured or unknown microorganisms. This has become possible by a tailored sample preparation and data analysis procedure that allows to detect single active enzymes within a complex metaproteome. As a 'proof-of-concept', we employed an alkyne-tagged version of the well-established fluorophosphonate-based (FP) chemical probe[38], which has been previously shown to allow

ABPP of serine hydrolases even under extreme experimental conditions such as elevated temperatures or low pH[36]. Moreover, members of the serine hydrolase superfamily are widely distributed across all domains of life[39] and are biocatalysts with broad application in biotechnological processes[40-45]. Accordingly, we profiled serine hydrolase activities of microbial communities from two different hot springs located at the Uzon Caldera (Kamchatka Peninsula, Russia) under native conditions directly at the site of sampling. For further sample processing, target enzyme enrichment and MS-based analysis, the *in vivo*-labeled sediments were transferred to the laboratory. By metagenome sequencing of an untreated reference sample, an *in silico* metaproteome was constructed that was used as a reference database to identify microbial enzymes with serine hydrolase activity.

Results

Workflow design

For establishing a function-based enzyme identification approach for the direct profiling of an environmental microbial community, we designed a general workflow based on a combination of ABPP with metagenome sequencing (Fig. 1). This workflow starts with the collection of a microbial community sample with sufficient biomass, here a hot spring sediment. This sample is then split into seven aliquots and six of them are used for *in vivo* labeling employing a specific two-step ABP aiming to target enzymes with the desired activity among the microbial enzyme repertoire directly at the site of sampling (eABPP). While three aliquots are treated with the respective probe, the other three replicates serve as solvent controls. The remaining sample aliquot, on the other hand, is used for extraction of eDNA. The eDNA is subsequently submitted to metagenome sequencing, enabling the generation of the microbial community-specific metaproteome database via sequence assembly, genome binning and gene prediction (Supplementary Fig. 1). In parallel, proteins from the eABPP samples are extracted and cleaned up by phenol extraction and ammonium acetate precipitation, followed by an *in vitro* ‘click’ reaction with downstream affinity enrichment of labeled enzymes. Enzyme identification is then achieved by on-bead digestion of captured enzymes with subsequent LC-MS/MS analysis using the self-assembled metaproteome database for identification of active target enzymes. Finally, enzyme function can be confirmed by recombinant expression of target enzymes and biochemical enzyme characterization.

Microbial composition of the sampled springs

To test this eABPP approach, we performed an *in vivo* ABPP of serine hydrolases from an environmental microbial community using a FP-based ABP. Accordingly, two hot springs located in the Uzon Caldera (Kamchatka Peninsula, Russia), the 'Arkashin shurf' (KAM3811) and the 'Helicopter spring' (KAM3808), were sampled (Fig. 2). These two sites were chosen due to their physicochemical properties (a temperature in the mesothermal range (40-75°C)[46] and a slightly acidic pH) favoring the development of an abundant microbial community. While KAM3811 is an originally artificial but for more than 30 years stable thermal pool[47], for which 16 metagenome-assembled genomes (MAGs) have been published recently[48], KAM3808 is a larger natural spring with no metagenome data available so far. After sampling both sites, metagenome sequencing of the epi-sedimentary microbial community was performed. We first analyzed the community composition of both sediments at the domain and class level based on the relative abundance of ribosomal protein S3 (rpS3) genes within the metagenomic datasets (relative abundance was determined via metagenomic read mapping; Fig. 3a). Both springs displayed considerable differences in terms of the percent distribution between Bacteria and Archaea and the number of different classes present within each domain. For KAM3811, 66% of the assigned species belonged to the bacterial domain, with the Aquificota being the dominating phylum, while the archaeal domain was mainly represented by the Thermoproteota (Supplementary Tab. 1). In the phenotypically and phylogenetically more diverse spring KAM3808, bacterial species account for 57 % of all organisms, with the Caldisericota making up the largest fraction, followed by the Desulfobacterota and the Aquificota. In contrast to KAM3811, archaeal species comprised mainly Euryarchaeota, while Archaea from the DPANN group made up the smallest fraction (Supplementary Tab. 2).

In a more detailed analysis of these metagenomic datasets, again by relying on genome coverages, all phyla identified within the respective spring were displayed in a domain-specific phylogenetic tree with their corresponding abundancies (Fig. 3b; Supplementary File 1). For KAM3808, which displayed more heteromorphic sediments, rpS3 genes of 49 microorganisms could be identified, with *Aciduliprofundum* sp. (relative abundance of 488.1), *Caldisericum exile* (408.2) and *Caldimicrobium thiodismutans* (139.9) as the most abundant species (Supplementary Tab. 2). KAM3811, a distinctly smaller thermal pool with homogenic brown to red sediments, comprised 19 different organisms as determined from rpS3 gene analysis.

Among them, *Sulfurihydrogenibium* sp. (2329.1) was identified as the dominating genus, followed by *Pyrobaculum ferrireducens* (411.3) and *Caldisphaera* sp. (389.6; Supplementary Tab. 1).

Term-based annotation of serine hydrolases in the metaproteome database is diverse

Based on protein prediction of the assembled metagenomes, metaproteome databases consisting of 45,649 (KAM3811) and 99,930 (KAM3808) protein coding sequences, respectively, were constructed. Annotation of the metaproteome databases against the UNIREF100 database resulted in 35,323 genes encoding proteins with a predicted function for the KAM3811 dataset while 8,575 proteins remained uncharacterized. Analogously, 85,393 genes encoded proteins with a predicted function for the KAM3808 dataset, whereas 17,923 genes encoded uncharacterized proteins. The discrepancy between the number of proteins with predicted function and hypothetical ones stemmed from the lack of represented similar genes within public databases or from fragmentation of genes. Within the proteome dataset of KAM3811 and KAM3808, only two and sixteen proteins, respectively, were annotated with the term ‘serine hydrolase’. Please note that the variety of term-based annotations of enzymes within both metaproteomes that potentially belong to the serine hydrolase superfamily is huge. Predicting the actual number of serine hydrolase sequences present in the two datasets thus remains intricate, since it would require elaborate manual curation of relevant search strings. This difficulty, atop of hypothetical or misannotated proteins, represents another drawback of simple sequence-driven bioinformatic annotation when searching for dedicated enzyme activities, especially from large superfamilies, which is also addressed by our function-based eABPP approach.

Identification of active serine hydrolases by eABPP

To identify active serine hydrolases from the sampled environmental communities, the self-constructed metaproteome databases were then employed as a reference database for the analysis of eABPP MS data resulting from an affinity enrichment of FP-alkyne-labeled target proteins via a biotin handle, on-bead digest and subsequent LC-MS/MS analysis (Supplementary File 2). Overall, 811 and 1489 protein groups were identified for the KAM3811 and KAM3808 datasets, respectively, excluding hits from the implemented contaminants database. After filtering of the initial data (see the methods section), a total of 413 protein groups for KAM3811 and 689 protein groups for KAM3808 remained for further analysis. For KAM3811, 221 protein groups displayed a positive log2-fold change when comparing the

group of FP-labeled samples against the DMSO-treated control group (Fig. 4a). Of them, 15 protein groups were predicted as serine hydrolases. Most notably, these protein groups were predominantly found among the top enriched hits, with 12 protein groups found among the top 30 enriched hits (Tab. 1). For KAM3808, which showed a larger microbial diversity, 348 protein groups were found to be enriched in the group of FP-labeled samples when compared against the DMSO-treated control group (Fig. 4b). These include 31 predicted serine hydrolases. As already observed for KAM3811, most of them were top hits, with 17 protein groups found among the top 30 enriched hits (Tab. 2). To confidently predict serine hydrolase activity across the proteins that were enriched with FP-alkyne, we here complemented the protein annotation from the UniRef100 database with sequence searches against the Pfam, NCBI CCD and InterPro databases. Where necessary, additional searches against the SWISS-MODEL Template Library or with HHpred were performed to identify structural homologs (Supplementary File 3).

The active microbial community serine hydrolases enriched by eABPP comprised enzymes from all enzyme classes belonging to the serine hydrolase superfamily, including proteases (or peptidases as synonymous term), lipases, amidases and esterases[39]. Across both springs, we found proteins predicted as serine-type peptidases from the (super-)families S1 (chymotrypsin family, subfamily S1C), S8/S53 (clan SB: S8 (subtilisin family, including subfamily S8A), S53 (type peptidase: sedolisin)), S9 (prolyl oligopeptidase family), S15 (type peptidase: Xaa-Pro dipeptidyl peptidase), S33 (type peptidase: prolyl aminopeptidase), S45 (type peptidase: penicillin G acylase precursor), S49 (protease IV family), and S66 (type peptidase: murein tetrapeptidase LD-carboxypeptidase)[49]. Moreover, we detected proteins predicted as esterases/lipases, including enzymes from the SGNH hydrolase superfamily[50] or enzymes containing a GDSL[51], a PNPLA (patatin-like phospholipases)[52] or a Lipase_bact_N[53] domain, and a DUF915 family enzyme with structural similarity to esterases/lipases as well as a putative esterase from the UPF0227 family. Furthermore, enzymes predicted as alpha/beta fold-1 hydrolases[54] without further classification but with esterases/lipases as structural homologs were identified from these datasets.

Function validation of a selected serine hydrolase

We further confirmed the applicability of eABPP for the function-based protein identification from an environmental microbial community through subsequent bioinformatical and biochemical characterization of a selected serine hydrolase. From the list of predicted serine

hydrolases identified by eABPP, we chose a putative esterase (ExploCarb_3811S_S4_483_length_13114_cov_941_5; Tab. 1) and retrieved and confined the corresponding gene sequence as described in the methods section. This enzyme was selected due to its relatively short and complete gene sequence as well as the availability of a straightforward enzyme kinetic assay.

The putative esterase is a 191 amino acid-containing protein possessing a UPF0227 domain (uncharacterized protein family 0227). A structural model of the putative esterase resembled secondary structural elements of characterized homologous esterases as well as its GTSLG sequence previously associated with thioesterase activity[55] that fits within the conserved GxSxG motif of serine hydrolases (Fig. 5a,b). The esterase sequence shows 100% identity to a predicted esterase (NCBI accession number: PMP76296.1) from a *Sulfurihydrogenibium* sp. MAG from a published Metagenome[48], originating from the same spring. BLAST analysis predicted YqiA esterases, alpha/beta-fold hydrolases, other unspecified esterases and hypothetical proteins with no evidence of function at protein level as closely related hits. Structural analysis with HHpred reveals confident homology to a broad variety of esterases, including methylesterases, strigolactone esterases, thioesterases and carbohydrate esterases. Besides, structural homologs with diverging activities such as lipases, epoxide hydrolases, haloperoxidases, halogenalkane dehalogenases, polyethylene terephthalate hydrolases (PETases), cutinases, biphenyl catabolic (bph) hydrolases and other meta-cleavage product (MCP) hydrolases have been identified (Supplementary Tab. 3). Based on these bioinformatic analyses, the UPF0227 domain-containing serine hydrolase is most likely an esterase of the alpha/beta-hydrolases superfamily with unknown substrate specificity.

In order to biochemically validate the function of the putative esterase, *E. coli* Rosetta (DE3) cells were transformed with the codon-optimized gene sequence cloned into a pET-28b(+) vector for heterologous protein expression. The purified enzyme was biochemically characterized using pNP-substrates. The putative esterase showed highest activity at a pH of 8.0 (Supplementary Fig. 2a) and a temperature of 70 °C (Supplementary Fig. 2b) using pNP-butyrate as a substrate. Moreover, enzyme kinetics were determined using different pNP-esters (Fig. 5c). Effective hydrolysis was observed for pNP-acetate (V_{max} = 8.93 U mg^{-1}, K_m = 0.12 mM), pNP-butyrate (V_{max} = 4.51 U mg^{-1}, K_m = 0.18 mM) and pNP-octanoate (V_{max} = 1.22 U mg^{-1}, K_m = 0.16 mM), confirming its function as an esterase, whereas no activity was measured for pNP-esters with longer chain lengths, such as pNP-decanoate or pNP-dodecanoate. Furthermore, *in vitro* ABPP of the purified esterase with FP-alkyne, analogously to the *in vivo*

ABPP experiment, revealed robust labeling over a range of esterase concentrations, which could be strongly diminished by pre-incubation with the generic serine esterase inhibitor paraoxon (Fig. 5d). Consequently, the esterase was proven a bonafide target of the covalent-acting FP probe. In line with this result, the FP-alkyne was furthermore shown to reduce the esterase activity of the enzyme in correlation with the applied probe concentration (Supplementary Fig. 2c).

Discussion

Extremal environments harbor microbial communities that provide myriad capabilities for identifying novel enzymatic activities and associated metabolic processes[56]. Due to their tremendous but mostly still hidden enzymatic potential, different strategies for their uncovering such as bioprospecting have emerged[11,12]. Despite recent technical advances, the study of the functional enzyme repertoire of an environmental microbial community and the reliable identification of the corresponding enzymes however remains a challenging task.

In this study, we established eABPP for the efficient function-based identification of active enzymes directly from the environment. The eABPP workflow relies on an activity-based labeling of a microbial community sample in combination with metagenome sequencing. In this way, the bioinformatic annotation of the metagenome can be directly confirmed through the activity-dependent reaction of the ABP-targeted enzyme, thereby providing direct experimental evidence for the bioinformatically predicted enzyme function along with its protein identification. So far, to the best of our knowledge, such a combination of metagenomics and MS-based ABPP has neither been established for protein identification from complex microbial communities and environmental samples, nor has it been applied directly in ecosystems.

As a showcase, we here reported the successful identification of active serine hydrolases from epi-sedimentary communities of two hot springs located in the Uzon Caldera region in Kamchatka (Russia), using the serine hydrolase target-specificity of FP-alkyne for functional enzyme annotation. Fluorophosphonate-based probes have been proven to show high specificity to serine hydrolases, even under harsh conditions such as high temperature[36,57], providing information about the activity state of the labeled target enzymes across the examined microbial community. A robust LC-MS/MS workflow furthermore allows the identification and quantification of the respective proteins. Thorough bioinformatic analysis of the eABPP-

enriched and -identified proteins from the sediments of the ’Arkashin shurf’ (KAM3811) and the ’Helicopter spring‘ (KAM3808) revealed that most of the top hits (i.e. proteins with the highest log2-fold changes compared to control samples treated with DMSO) were indeed serine hydrolases, demonstrating the applicability of the method. It is therefore most likely that these proteins represent the most active serine hydrolases under the *in situ* conditions of the hot springs, due to either their higher abundances or their elevated activities compared to other serine hydrolases in the sampled environment. Of note, the applicability of our eABPP approach relies on different preconditions such as the availability of ABPs with sufficient target specificity and stability (e.g. in hot springs). It also required the development of distinct sample preparation procedures such as a suitable workflow for clean-up of proteins from the complex sample matrix prior to analysis by MS, or data analysis methodologies for detecting the low abundant single proteins from a metaproteome sample as well as the accurate construction of a metaproteome sequence database[29].

To further corroborate the versatility of our eABPP approach, we heterologously expressed a putative esterase from the ‘uncharacterized protein family UPF0227’ which was found among the top 30 hits from KAM3811 enriched with FP-alkyne and demonstrated that this enzyme is efficiently hydrolyzing pNP-acetate (C2), pNP-butyrate (C4) and pNP-octanoate (C8), thereby confirming that the enzyme has esterase activity. Moreover, the esterase was efficiently labeled with FP-alkyne *in vitro*, thus supporting that the UPF0227 family protein indeed belongs to the serine hydrolases and overall confirming the eABPP-based identification (Fig. 5).

Overall, our study thus proved that the herein described workflow is suitable for the reliable identification of active serine hydrolases directly in the environment, by employing a well-studied ABP that has a broad spectrum of target enzymes. The proteins reported as hits for both springs include only sequences which could be readily identified as serine hydrolases, demonstrating the robustness of the applied approach. Please note that it is however likely that there are several more enzymes displaying serine hydrolase features and activities across the two sets of enriched proteins, which were not uncovered by this analysis.

The investigation of the functional potential of an environmental microbial community is, for example, used in the discovery of novel biocatalysts[12]. With the onset of the metagenomics era, the search for enzymes with prospect industrial use became far more effective[58]. Of note, most of the metagenomics-derived biocatalysts have been discovered by functional screening[14,59]. We thus believe that eABPP represents a novel approach for the efficient function-based

identification of enzymes from microbial communities that may find application in industrial processes, thereby representing an alternative to functional metagenomics, which is quite time- and labor-intensive[15], but frequently results in a rather low success rate[59]. Furthermore, our approach has the advantage that it allows to explore the biological diversity of active enzymes with dedicated functions from an ecosystem in their native state. With the majority of active serine hydrolases identified from our screen being annotated as subtilases (peptidase family S8) or penicillin acylases (peptidase family S15; Tab.1-2), we indeed detected thermostable representatives of the functionally highly diverse family of serine hydrolases that already find application for industrial purposes. In particular, proteases from the subtilisin family are utilized in the laundry industry[60] or for the production of valuable chemicals from proteinaceous substrates[61,62], whereas penicillin acylases are important enzymes for the production of 6-aminopenicillanic acid (6-APA), the β-lactam nucleus used in the manufacturing process of semi-synthetic antibiotics, through hydrolysis of natural penicillins[63].

As a large and constantly growing repertoire of probes with a wide variety of warheads for targeting different enzymes or enzyme classes is already available, we believe that eABPP can be further extended to profile additional enzyme activities of microbial communities from diverse ecosystems, especially with respect to finding enzyme activities that harbor potential for application in industrial processes. These include, for example, cysteine proteases or glycoside hydrolases for which a broad set of probes is currently at hand[64,65]. Glycoside hydrolases, for instance, are of particular interest with regard to biocatalyst discovery since they largely function in (lignocellulosic) biomass degradation[66,67]. In addition, it might be reasonable to design more specific probes with activity towards serine hydrolases that function in the degradation of environmental pollutants (e. g. PETases, and different bph- and MCP hydrolases capable of cleaving mono- and bi-cyclic aromatic compounds) as these enzymes are of interest for environmental and industrial applications.

Finally, we believe that our approach may also find application besides industrial biocatalyst discovery. For example, we anticipate that eABPP may also be adapted for ecological research, for instance to decipher ecological enzyme activity patterns, e.g. from synergistic or mutual interactions, in microbial communities.

References

1 Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. USA* **113**, 5970-5975 (2016).

2 Verstraete, W. Microbial ecology and environmental biotechnology. *ISME J.* **1**, 4-8 (2007).

3 Baquero, F., Coque, T. M., Galan, J. C. & Martinez, J. L. The Origin of Niches and Species in the Bacterial World. *Front. Microbiol.* **12**, 657986 (2021).

4 Wohlgemuth, R. Biocatalysis - Key enabling tools from biocatalytic one-step and multi-step reactions to biocatalytic total synthesis. *N. Biotechnol.* **60**, 113-123 (2021).

5 Elleuche, S., Schroder, C., Sahm, K. & Antranikian, G. Extremozymes - biocatalysts with unique properties from extremophilic microorganisms. *Curr. Opin. Biotech.* **29**, 116-123 (2014).

6 Koskinen, P. E. P. *et al.* Bioprospecting thermophilic microorganisms from Icelandic hot springs for hydrogen and ethanol production. *Energy Fuels* **22**, 134-140 (2008).

7 Sahoo, R. K., Kumar, M., Sukla, L. B. & Subudhi, E. Bioprospecting hot spring metagenome: lipase for the production of biodiesel. *Environ. Sci. Pollut. Res.* **24**, 3802-3809 (2017).

8 Thankappan, S. *et al.* Bioprospecting thermophilic glycosyl hydrolases, from hot springs of Himachal Pradesh, for biomass valorization. *AMB Expr.* **8**, 168 (2018).

9 Özdemir, S. C. & Uzel, A. Bioprospecting of hot springs and compost in West Anatolia regarding phytase producing thermophilic fungi. *Sydowia* **72**, 1-11 (2020).

10 Prayogo, F. A. *et al.* Metagenomic applications in exploration and development of novel enzymes from nature: a review. *J. Genet. Eng. Biotechnol.* **18**, 39 (2020).

11 Kennedy, J., Marchesi, J. R. & Dobson, A. D. W. Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microb. Cell Fact.* **7**, 27 (2008).

12 DeCastro, M. E., Rodriguez-Belmonte, E. & Gonzalez-Siso, M. I. Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Front. Microbiol.* **7**, 1521 (2016).

13 Harrington, E. D. *et al.* Quantitative assessment of protein function prediction from metagenomics shotgun sequences. *Proc. Natl. Acad. Sci. USA* **104**, 13913-13918 (2007).

14 Berini, F., Casciello, C., Marcone, G. L. & Marinelli, F. Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* **364**, fnx211 (2017).

15 Lam, K. N., Cheng, J. J., Engel, K., Neufeld, J. D. & Charles, T. C. Current and future resources for functional metagenomics. *Front. Microbiol.* **6**, 1196 (2015).

16 Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **77**, 383-414 (2008).

17 Willems, L. I., Overkleeft, H. S. & van Kasteren, S. I. Current Developments in Activity-Based Protein Profiling. *Bioconjugate Chem.* **25**, 1181-1191 (2014).

18 Yang, P. Y. & Liu, K. Activity-Based Protein Profiling: Recent Advances in Probe Development and Applications. *Chembiochem* **16**, 712-724 (2015).

19 Fonović, M. & Bogyo, M. Activity-based probes as a tool for functional proteomic analysis of proteases. *Expert Rev. Proteomics* **5**, 721-730 (2008).

20 Wright, M. H. & Sieber, S. A. Chemical proteomics approaches for identifying the cellular targets of natural products. *Nat. Prod. Rep.* **33**, 681-708 (2016).

21 Barglow, K. T. & Cravatt, B. F. Activity-based protein profiling for the functional annotation of enzymes. *Nat. Methods* **4**, 822-827 (2007).

22 Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper(I)-catalyzed azide-alkyne [3+2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686-4687 (2003).

23 Speers, A. E. & Cravatt, B. F. Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.* **11**, 535-546 (2004).

24 Ovaa, H. *et al.* Chemistry in living cells: Detection of active proteasomes by a two-step labeling strategy. *Angew. Chem. Int. Ed.* **42**, 3626-3629 (2003).

25 Berger, A. B., Vitorino, P. M. & Bogyo, M. Activity-based protein profiling: applications to biomarker discovery, in vivo imaging and drug discovery. *Am. J. Pharmacogenomics* **4**, 371-381 (2004).

26 Keller, L. J., Babin, B. M., Lakemeyer, M. & Bogyo, M. Activity-based protein profiling in bacteria: Applications for identification of therapeutic targets and characterization of microbial communities. *Curr. Opin. Chem. Biol.* **54**, 45-53 (2020).

27 Morimoto, K. & van der Hoorn, R. A. The Increasing Impact of Activity-Based Protein Profiling in Plant Science. *Plant Cell Physiol.* **57**, 446-461 (2016).

28 Klaus, T. *et al.* Activity-Based Protein Profiling for the Identification of Novel Carbohydrate-Active Enzymes Involved in Xylan Degradation in the Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1E. *Front. Microbiol.* **12**, 734039 (2021).

29 Whidbey, C. & Wright, A. T. Activity-Based Protein Profiling-Enabling Multimodal Functional Studies of Microbial Communities. *Curr. Top. Microbiol. Immunol.* **420**, 1-21 (2019).

30 Mayers, M. D., Moon, C., Stupp, G. S., Su, A. I. & Wolan, D. W. Quantitative Metaproteomics and Activity-Based Probe Enrichment Reveals Significant Alterations in Protein Expression from a Mouse Model of Inflammatory Bowel Disease. *J. Proteome Res.* **16**, 1014-1026 (2017).

31 Chatterjee, S. *et al.* A comprehensive and scalable database search system for metaproteomics. *BMC Genomics* **17**, 642 (2016).

32 Chauvigne-Hines, L. M. *et al.* Suite of activity-based probes for cellulose-degrading enzymes. *J. Am. Chem. Soc.* **134**, 20521-20532 (2012).

33 Ansong, C. *et al.* Characterization of protein redox dynamics induced during light-to-dark transitions and nutrient limitation in cyanobacteria. *Front. Microbiol.* **5**, 325 (2014).

34 Sadler, N. C. *et al.* Live cell chemical profiling of temporal redox dynamics in a photoautotrophic cyanobacterium. *ACS Chem. Biol.* **9**, 291-300 (2014).

35 Bennett, K., Sadler, N. C., Wright, A. T., Yeager, C. & Hyman, M. R. Activity-Based Protein Profiling of Ammonia Monooxygenase in Nitrosomonas europaea. *Appl. Environ. Microbiol.* **82**, 2270-2279 (2016).

36 Zweerink, S. *et al.* Activity-based protein profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nat. Commun.* **8** (2017).

37 Sakoula, D. *et al.* Universal activity-based labeling method for ammonia- and alkane-oxidizing bacteria. *ISME J.* **16**, 958-971 (2022).

38 Liu, Y. S., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: The serine hydrolases. *Proc. Natl. Acad. Sci. USA* **96**, 14694-14699 (1999).

39 Simon, G. M. & Cravatt, B. F. Activity-based Proteomics of Enzyme Superfamilies: Serine Hydrolases as a Case Study. *J. Biol. Chem.* **285**, 11051-11055 (2010).

40 Chandra, P., Enespa, Singh, R. & Arora, P. K. Microbial lipases and their industrial applications: a comprehensive review. *Microb. Cell Fact.* **19**, 169 (2020).

41 Panda, T. & Gowrishankar, B. S. Production and applications of esterases. *Appl. Microbiol. Biotechnol.* **67**, 160-169 (2005).

42 Barzkar, N. *et al.* Marine Bacterial Esterases: Emerging Biocatalysts for Industrial Applications. *Appl. Biochem. Biotechnol.* **193**, 1187-1214 (2021).

43 Anobom, C. D. *et al.* From Structure to Catalysis: Recent Developments in the Biotechnological Applications of Lipases. *Biomed. Res. Int.* **2014**, 684506 (2014).

44 Romano, D. *et al.* Esterases as stereoselective biocatalysts. *Biotechnol. Adv.* **33**, 547-565 (2015).

45 Wu, Z. M., Liu, C. F., Zhang, Z. Y., Zheng, R. C. & Zheng, Y. G. Amidase as a versatile tool in amide-bond cleavage: From molecular features to biotechnological applications. *Biotechnol. Adv.* **43** (2020).

46 Pentecost, A., Jones, B. & Renaut, R. W. What is a hot spring? *Can. J. Earth Sci.* **40**, 1443-1446 (2003).

47 Burgess, E. A., Unrine, J. M., Mills, G. L., Romanek, C. S. & Wiegel, J. Comparative geochemical and microbiological characterization of two thermal pools in the Uzon Caldera, Kamchatka, Russia. *Microb. Ecol.* **63**, 471-489 (2012).

48 Wilkins, L. G. E., Ettinger, C. L., Jospin, G. & Eisen, J. A. Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci. Rep.* **9**, 3059 (2019).

49 Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624-D632 (2018).

50 Mølgaard, A., Kauppinen, S. & Larsen, S. Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases. *Structure* **8**, 373-383 (2000).

51 Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H. & Shaw, J. F. GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43**, 534-552 (2004).

52 Baulande, S. & Langlois, C. [Proteins sharing PNPLA domain, a new family of enzymes regulating lipid metabolism]. *Med. Sci.* **26**, 177-184 (2010).

53 Chuang, Y. C., Chiou, S. F., Su, J. H., Wu, M. L. & Chang, M. C. Molecular analysis and expression of the extracellular lipase of Aeromonas hydrophila MCC-2. *Microbiol.* **143**, 803-812 (1997).

54 Ollis, D. L. *et al.* The Alpha/Beta-Hydrolase Fold. *Protein Eng.* **5**, 197-211 (1992).

55 Thankachan, D. *et al.* A trans-Acting Cyclase Offloading Strategy for Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **14**, 845-849 (2019).

56 Shu, W. S. & Huang, L. N. Microbial diversity in extreme environments. *Nat. Rev. Microbiol.* **20**, 219-235 (2022).

57 Kidd, D., Liu, Y. & Cravatt, B. F. Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**, 4005-4015 (2001).

58 Ferrer, M. *et al.* Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol.* **9**, 22-34 (2016).

59 Zhang, L. *et al.* Advances in Metagenomics and Its Application in Environmental Microorganisms. *Front. Microbiol.* **12**, 766364 (2021).

60 Niehaus, F. *et al.* Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases. *Microb. Biotechnol.* **4**, 767-776 (2011).

61 De Oliveira Martinez, J. P. *et al.* Challenges and Opportunities in Identifying and Characterising Keratinases for Value-Added Peptide Production. *Catalysts* **10**, 184 (2020).

62 Li, J. *et al.* A Novel Gelatinase from Marine Flocculibacter collagenilyticus SM1988: Characterization and Potential Application in Collagen Oligopeptide-Rich Hydrolysate Preparation. *Mar. Drugs* **20**, 48 (2022).

63 Arroyo, M., de la Mata, I., Acebal, C. & Castillon, M. P. Biotechnological applications of penicillin acylases: state-of-the-art. *Appl. Microbiol. Biotechnol.* **60**, 507-514 (2003).

64 Wu, L. *et al.* An overview of activity-based probes for glycosidases. *Curr. Opin. Chem. Biol.* **53**, 25-36 (2019).

65 Serim, S., Haedke, U. & Verhelst, S. H. Activity-based probes for the study of proteases: recent advances and developments. *ChemMedChem* **7**, 1146-1159 (2012).

66 Suleiman, M., Kruger, A. & Antranikian, G. Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnol. Biofuels* **13** (2020).

67 Prieto, A. *et al.* Fungal glycosyl hydrolases for sustainable plant biomass valorization: Talaromyces amestolkiae as a model fungus. *Int. Microbiol.* **24**, 545-558 (2021).

Acknowledgements

BS, MK and IK acknowledge funding within the DFG-RSF Cooperation for joint German-Russian projects by the DFG (SI 642/12-1, KA 2894/6-1) and RSF (18-44-04024). This work was also supported by the DFG grant INST 20876/322-1 FUGG (to M. K. and F.K.).

Author contributions

T.K. and T.V.K. performed in-field chemical labeling experiments, T.V.K. and I.V.K. performed DNA extraction and metagenomic sequencing, S.P.E. and A.J.P. performed refinement of metagenome data and metaproteome assembly, S.N., T.K. and L.S. performed protein extraction and affinity enrichment experiments, S.N. and F.K. performed MS analyses, T.K. performed protein expression and enzyme assays, B.S., M.K. and I.V.K. designed the workflow for the field ABPP, C.B, B.S. and M.K. supervised the study and S.N., T.K. B.S. and M.K. wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Figures

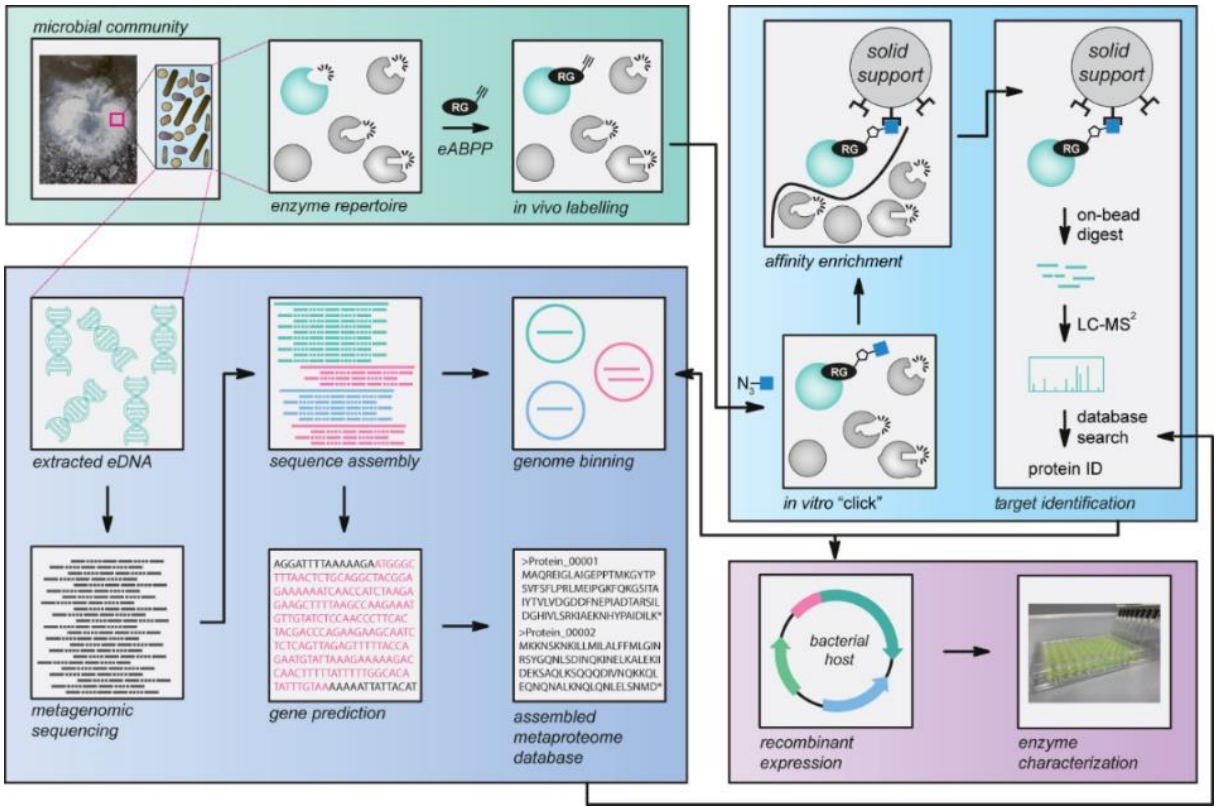


Figure 1. Environmental ABPP workflow

Workflow of the established environmental ABPP (eABPP) approach for the function-based identification of serine hydrolases. This approach can be divided into four different blocks, i.e. sampling and *in vivo* labeling of an environmental microbial community, metagenomics, target protein identification by LC-MS/MS and enzyme characterization of a protein of interest.

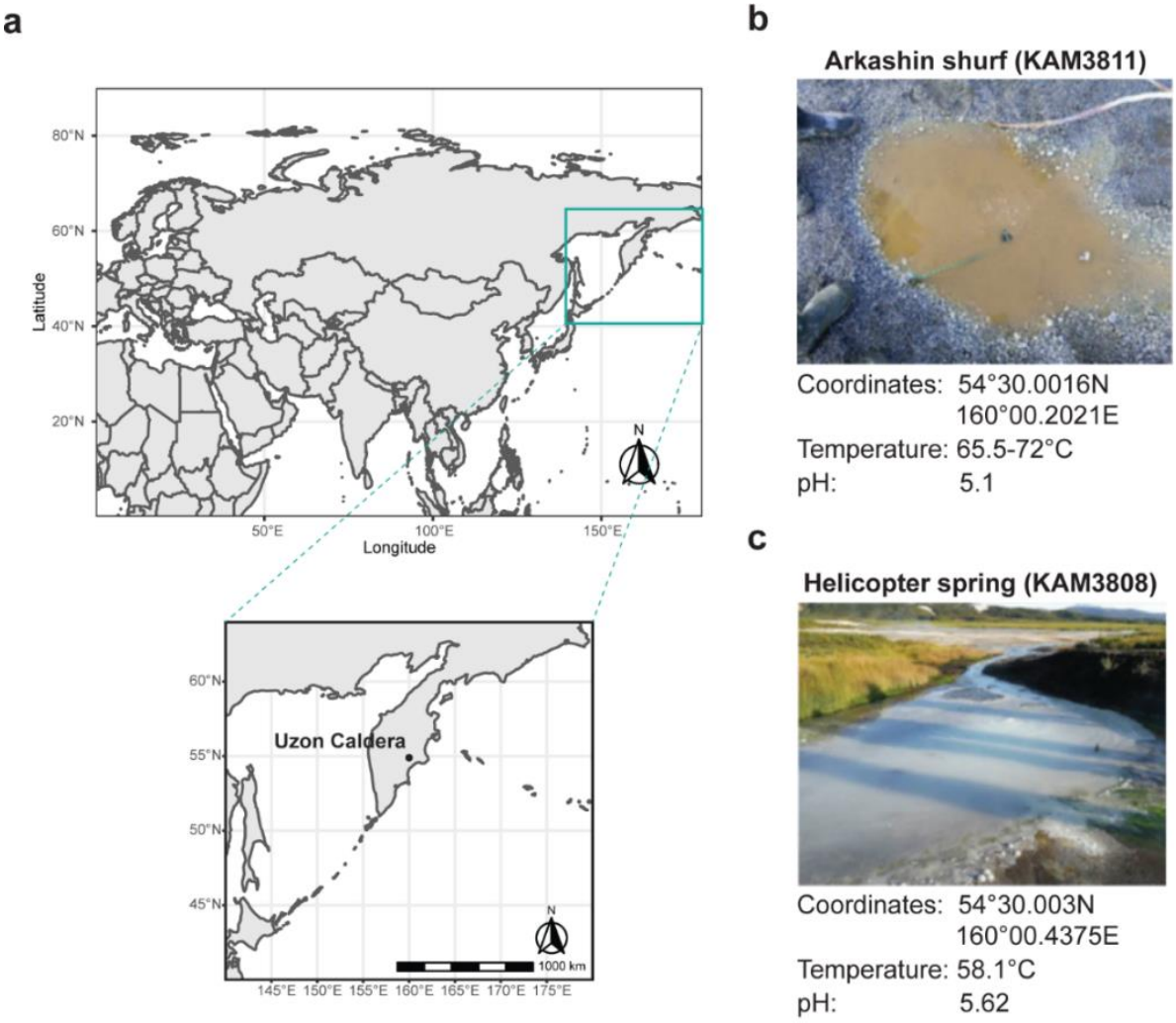


Figure 2. Location and community composition of the sampled springs

(**a**) The map shows the location of the two sampled springs ‘Arkashin shurf’ (KAM3811) and ‘Helicopter spring’ (KAM3808) at the Uzon Caldera region, Kamchatka Peninsula, Russia. (**b,c**) A representative picture is shown along with the exact coordinates and physicochemical properties of both springs.

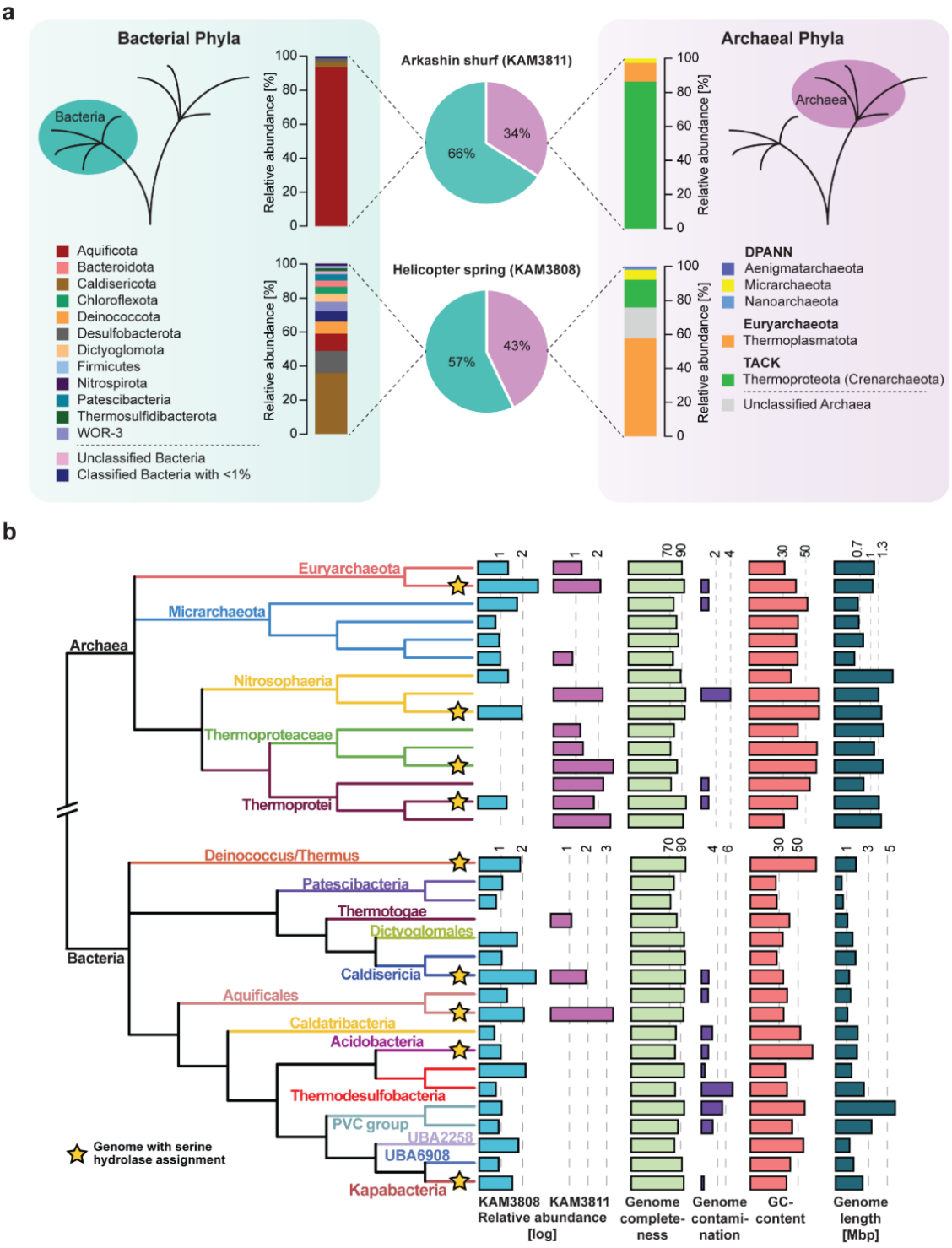


Figure 3. Distribution of microorganisms across KAM3811 and KAM3808

(**a**) The proportion of Bacteria (cyan) and Archaea (magenta) within the sediments sampled for eABPP is depicted as pie diagrams for KAM3811 and KAM3808. An overview of the relative distribution of representative phyla from these domains is given respectively based on the GTDB taxonomy. (**b**) Phylogenetic tree displaying the relationship between the microorganisms found across both springs as calculated with GTDB-Tk based on the

dereplicated, binned and curated metagenomes from KAM3808 and KAM3811. Relative abundances of microorganisms based on the coverage of genomes are given for KAM3808 (light blue) and KAM3811 (magenta), respectively, along with their genome completeness (light green), contamination (purple), GC content (light red) and genome length (dark cyan) as calculated via checkM. The yellow stars indicate from which genomes predicted serine hydrolases were confidently identified with the applied eABPP approach.

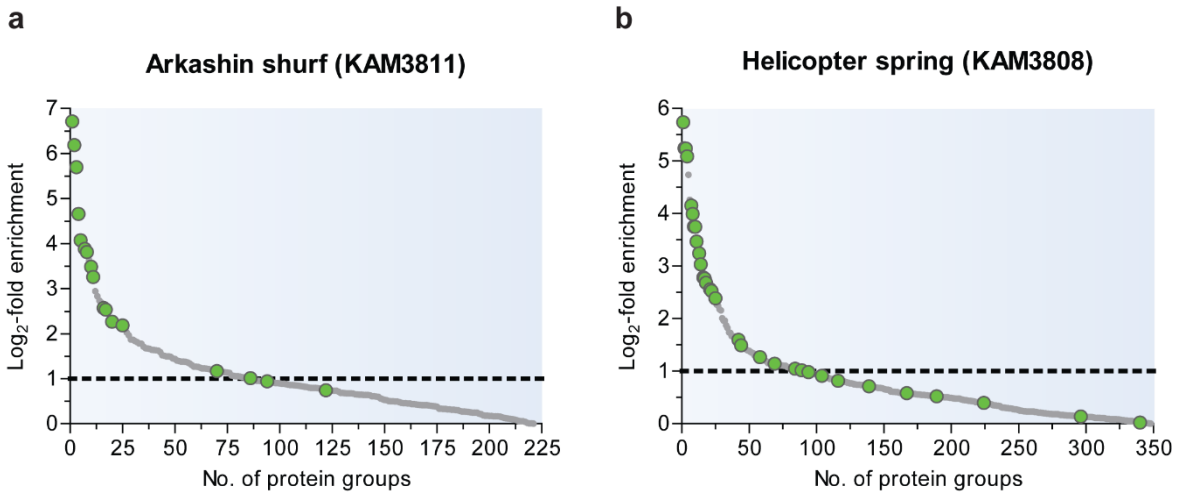
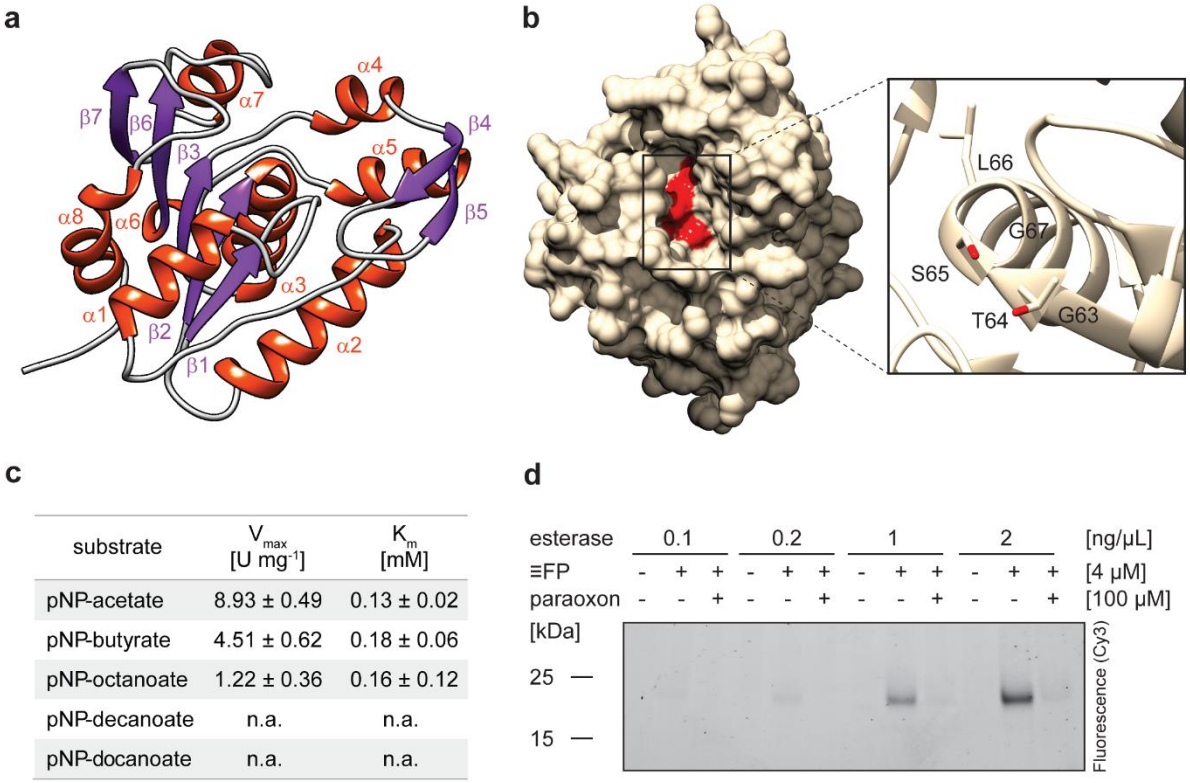


Figure 4. Predicted serine hydrolases identified from the sampled hot springs

Log2-fold enrichment of identified proteins labeled with FP-alkyne compared to the DMSO control for the sediments sampled from KAM3808 (**a**) and KAM3811 (**b**). Proteins predicted as serine hydrolases are displayed as green dots. Hits lying above the dotted line were more than two-fold enriched with the probe and were therefore considered primary hits. Each treatment group comprised three biological replicates.



substrate	V_{max} [U mg^{-1}]	K_m [mM]
pNP-acetate	8.93 ± 0.49	0.13 ± 0.02
pNP-butyrate	4.51 ± 0.62	0.18 ± 0.06
pNP-octanoate	1.22 ± 0.36	0.16 ± 0.12
pNP-decanoate	n.a.	n.a.
pNP-docanoate	n.a.	n.a.

Figure 5. Characterization of the putative esterase

(**a**) Predicted structure of the UPF0227 protein with the secondary structures visualized in red (alpha helices) and purple (beta strands). The five-stranded parallel beta sheet consists of the strands β1, β2, β3, β6 and β7. (**b**) Surface-displaying structure of the putative esterase. The close-up displays the conserved serine hydrolase motif GxSxG, comprising the residues G63, T64, S65, L66 and G67, which is located in a substrate pocket (depicted in red). Structure prediction was performed with AlphaFold and the output was processed in Chimera. (**c**) Kinetic characterization of the esterase using the pNP-substrates pNP-acetate, -butyrate, -octanoate, -decanoate and -dodecanoate at concentrations up to 0.7 mM at pH 8.0 and 70°C and calculation of V_{max} and K_m. Values represent the mean of three technical replicates ± SD. (**d**) *In vitro* labeling of varying amounts of the esterase with FP-alkyne in presence or absence of paraoxon.

Tables

Table 1. Predicted serine hydrolases identified for KAM3811.

No.	Log2-fold change	Identifier	Protein annotation[#]
1	6.711	ExploCarb_3811S_S4_2994_length_2665_cov_56_1	Peptidase S8, subtilisin-related
2	6.188	ExploCarb_3811S_S4_103_length_35923_cov_48_7	Penicillin/GL-7-ACA/AHL acylase
3	5.695	ExploCarb_3811S_S4_37_length_59942_cov_23_2	Peptidase S8, subtilisin-related
4	4.656	ExploCarb_3811S_S4_179_length_24712_cov_181_27	Penicillin/GL-7-ACA/AHL acylase
5	4.074	ExploCarb_3811S_S4_782_length_9153_cov_146_4	Penicillin/GL-7-ACA/AHL acylase
7	3.881	ExploCarb_3811S_S4_477_length_13181_cov_9_10	Peptidase S8, subtilisin-related
8	3.810	ExploCarb_3811S_S4_412_length_14764_cov_159_5	Peptidase S8, subtilisin-related
10	3.483	ExploCarb_3811S_S4_5916_length_1495_cov_4_1	Peptidase_S8/S53_dom
11	3.258	ExploCarb_3811S_S4_7165_length_1270_cov_6_1	Peptidase S8, subtilisin-related
17	2.533	ExploCarb_3811S_S4_1591_length_4725_cov_42_1	Protein of unknown function DUF915, hydrolase-like
20	2.264	ExploCarb_3811S_S4_483_length_13114_cov_941_5	Uncharacterised protein family UPF0227/Esterase YqiA
25	2.184	ExploCarb_3811S_S4_1380_length_5439_cov_180_2	Peptidase_S9
70	1.172	ExploCarb_3811S_S4_259_length_20079_cov_121_4	Peptidase_S49
94	0.943	ExploCarb_3811S_S4_1740_length_4311_cov_157_3	Xaa-Pro-like_dom
122	0.745	ExploCarb_3811S_S4_1428_length_5265_cov_777_6	Peptidase_S49

InterProScan annotation on either protein family or domain level

Table 2. Predicted serine hydrolases identified for KAM3808.

No.	Log2-fold change	Identifier	Protein annotation[#]
1	5.739	ExploCarb_3808S_S2_15464_length_1283_cov_2_1	GDSL lipase/esterase
2	5.244	ExploCarb_3808S_S2_5917_length_3034_cov_1_3	Peptidase S8, subtilisin-related
3	5.238	ExploCarb_3808S_S2_19633_length_1026_cov_55_1	Putative S8A family peptidase[†]
4	5.086	ExploCarb_3808S_S2_470_length_27453_cov_4_11	SGNH_hydro
7	4.152	ExploCarb_3808S_S2_593_length_22580_cov_176_21	Esterase/lipase
8	3.990	ExploCarb_3808S_S2_12785_length_1520_cov_45_2	Peptidase S8, subtilisin-related
9	3.754	ExploCarb_3808S_S2_10394_length_1832_cov_227_1	Peptidase_S8/S53_dom_sf
10	3.748	ExploCarb_3808S_S2_1348_length_11661_cov_169_4	PNPLA_dom
11	3.467	ExploCarb_3808S_S2_277_length_40468_cov_33_33	Peptidase S8A, fervidolysin-like
13	3.243	ExploCarb_3808S_S2_1901_length_8595_cov_3_5	Penicillin/GL-7-ACA/AHL acylase
14	3.030	ExploCarb_3808S_S2_12072_length_1604_cov_127_1	Peptidase_S8/S53_dom_sf
16	2.780	ExploCarb_3808S_S2_5180_length_3427_cov_2_3	Peptidase S8, subtilisin-related
17	2.761	ExploCarb_3808S_S2_12016_length_1611_cov_2_1	AB_hydrolase_1
18	2.684	ExploCarb_3808S_S2_2184_length_7533_cov_3_4	Penicillin/GL-7-ACA/AHL/aculeacin-A acylase
21	2.559	ExploCarb_3808S_S2_1_length_789533_cov_14_559	Peptidase S8, subtilisin-related
22	2.527	ExploCarb_3808S_S2_84_length_75002_cov_30_5	PNPLA_dom
25	2.390	ExploCarb_3808S_S2_14352_length_1370_cov_1_2	AB_hydrolase_1
42	1.597	ExploCarb_3808S_S2_5211_length_3405_cov_171_3	Peptidase family S66
44	1.491	ExploCarb_3808S_S2_7708_length_2386_cov_3_1	Lipase_bact_N

58	1.269	ExploCarb_3808S_S2_45_length_96604_cov_169_9	PNPLA_dom
69	1.142	ExploCarb_3808S_S2_1180_length_13195_cov_3_10	SGNH_hydro_sf
84	1.047	ExploCarb_3808S_S2_2714_length_6206_cov_211_1	Penicillin/GL-7-ACA/AHL acylase
89	1.016	ExploCarb_3808S_S2_1272_length_12299_cov_60_1	AB_hydrolase_1
104	0.911	ExploCarb_3808S_S2_17847_length_1126_cov_1_2	AB_hydrolase_1
116	0.819	ExploCarb_3808S_S2_5424_length_3285_cov_7_1	Peptidase S8, subtilisin-related
139	0.710	ExploCarb_3808S_S2_12631_length_1536_cov_2_2	Peptidase S1C, Do
167	0.583	ExploCarb_3808S_S2_2577_length_6506_cov_183_2	Peptidase S8, subtilisin-related
189	0.521	ExploCarb_3808S_S2_593_length_22580_cov_176_11	Peptidase S8, subtilisin-related
224	0.398	ExploCarb_3808S_S2_4160_length_4192_cov_19_2	Penicillin/GL-7-ACA/AHL acylase
296	0.139	ExploCarb_3808S_S2_21_length_132706_cov_167_47	Peptidase S1C
340	0.023	ExploCarb_3808S_S2_858_length_17095_cov_20_14	Hydrolase_4 (Serine aminopeptidase, S33)

InterProScan annotation on either protein family, homologous superfamily or domain level unless stated otherwise

† UniRef100 annotation

Supplementary Information

Environmental activity-based protein profiling for function-driven enzyme discovery from natural communities

(Submitted manuscript – *bioRxiv*)

Sabrina Ninck[1,#], Thomas Klaus[2,#], Tatiana V Kochetkova[3], Sarah P. Esser[4], Leonard Sewald[1], Farnusch Kaschani[1], Christopher Bräsen[2], Alexander J. Probst[4], Ilya V. Kublanov[3], Bettina Siebers[2,*], Markus Kaiser[1,*]

1 Chemical Biology, ZMB, Faculty of Biology, Universität Duisburg-Essen, Universitätsstr. 2, 45117 Essen, Germany.

2 Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Centre for Water and Environmental Research (CWE), Faculty of Chemistry, Universität Duisburg-Essen, Universitätsstr. 5, 45117 Essen, Germany.

3 Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Prospekt 60-Let Oktyabrya 7-2, 117312 Moscow, Russia

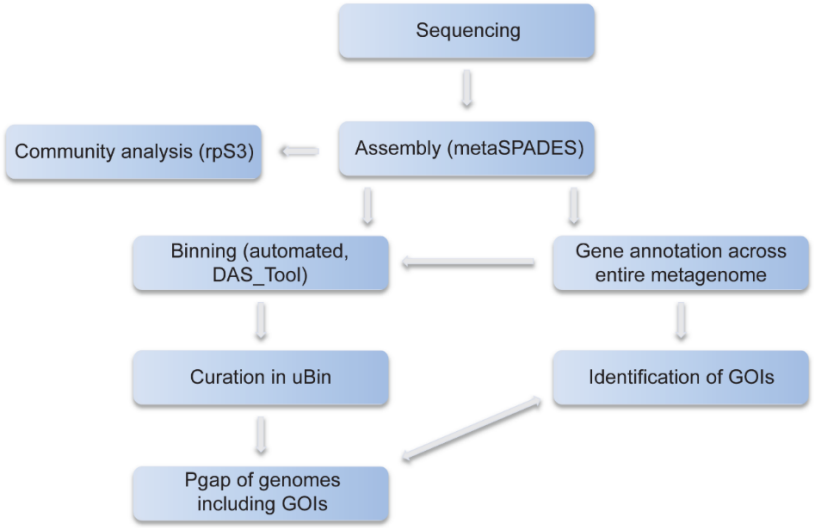
4 Group for Aquatic Microbial Ecology, UMB, Faculty of Chemistry, Universität Duisburg-Essen, Universitätsstr. 5, 45117 Essen, Germany.

* Corresponding Authors:

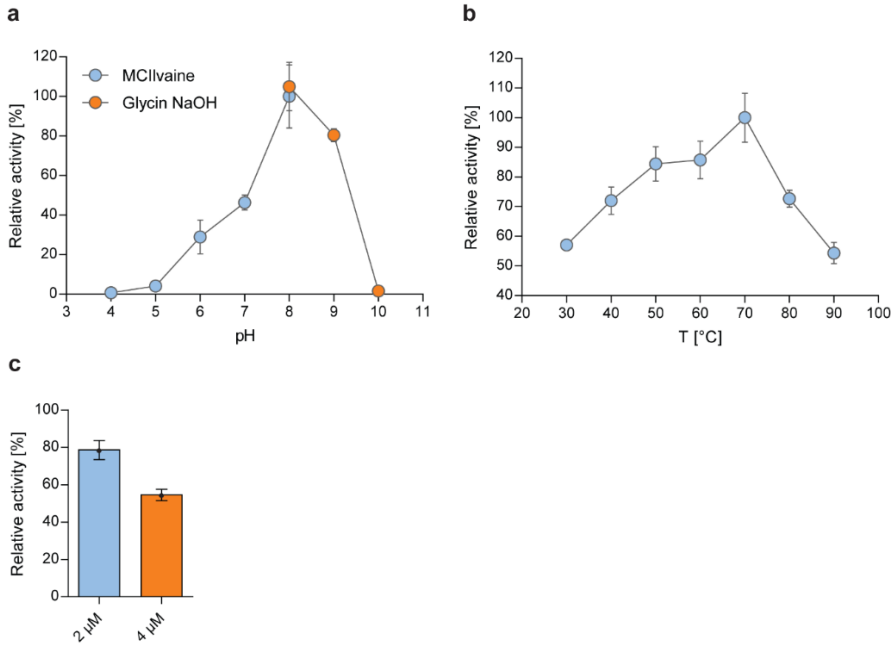
Email: bettina.siebers@uni-due.de
markus.kaiser@uni-due.de.

[#] Both authors contributed equally.

Supplementary Figures



Supplementary Figure 1. Bioinformatics pipeline. Visualization of the bioinformatics pipeline described in the Method section. The pipeline starts with sequencing, followed by quality-filtering and assembly, to building the metaproteome database and the identification of genes of interest.



Supplementary Figure 2. Biochemical characterization of the heterologously expressed esterase. Effect of pH (**a**) and temperature (**b**) on the relative activity of the esterase with pNP-butyrate. (**c**) Activity inhibition of the esterase upon preincubation with 2 µM or 4 µM FP-alkyne at pH 7.5 and 70° C for 10 min. The remaining activity was measured using pNP-acetate as the substrate. Error bars represent SD of three technical replicates.

Supplementary Tables

Supplementary Table 1. Relative abundance of microorganisms in KAM3811 based on the coverage of scaffolds containing the respective rpS3 gene.

Nearest predicted relative based on BLASTp	Phylum	Domain	Abundance (rpS3)
Sulfurihydrogenibium sp.	Aquificota	Bacteria	2329.1
Pyrobaculum ferrireducens	Thermoproteota	Archaea	411.3
Caldisphaera sp.	Thermoproteota	Archaea	389.6
Hydrogenobaculum sp.	Aquificota	Bacteria	281.7
Nitrososphaera sp.	Thermoproteota	Archaea	160.8
Sulfolobales (SCGC AB-777 J03)	Thermoproteota	Archaea	158.5
Aciduliprofundum sp.	Thermoplasmatota	Archaea	139.0
Caldisericum exile	Caldisericota	Bacteria	78.2
Fervidicoccus sp.	Thermoproteota	Archaea	59.7
Caldimicrobium thiodismutans	Desulfobacterota	Bacteria	58.9
Candidatus Nanobsidianus stetteri	Nanoarchaeota	Archaea	38.5
Thermoproteus sp. (JCHS 4)	Thermoproteota	Archaea	26.9
Aciduliprofundum sp.	Thermoplasmatota	Archaea	19.1
Caldivirga maquilingensis (IC-167)	Thermoproteota	Archaea	17.1
Mesoaciditoga sp.	Thermotogota	Bacteria	13.9
Caldivirga maquilingensis (IC-167)	Thermoproteota	Archaea	11.5
Thiomonas sp. (CB3)	Proteobacteria	Bacteria	11.1
Thiomonas sp. (20-64-5)	Proteobacteria	Bacteria	10.9
Acidilobus saccharovorans (345-15)	Thermoproteota	Archaea	7.7

Supplementary Table 2. Relative abundance of microorganisms in KAM3808 based on the coverage of the scaffold containing the respective rpS3 gene.

Nearest predicted relative based on BLASTp	Phylum	Domain	Abundance (rpS3)
Aciduliprofundum sp.	Thermoplasmatota	Archaea	488.1
Caldisericum exile	Caldisericota	Bacteria	408.2
Caldimicrobium thiodismutans	Desulfobacterota	Bacteria	139.9
Archaeon, unclassified	not predicted	Archaea	124.6
Sulfurihydrogenibium sp.	Aquificota	Bacteria	119.9
Nitrososphaera sp.	Thermoproteota	Archaea	93.7
Thermus arciformis	Deinococcota	Bacteria	81.6
Candidate division WOR-3 JGI Cruoil (03_51_56)	WOR-3	Bacteria	64.5
Dictyoglomus turgidum (DSM 6724)	Dictyoglomota	Bacteria	54.9
Candidatus Micrarchaeota (CG1_02_51_15)	Micrarchaeota	Archaea	52.9
Thermoflexia	Chloroflexota	Bacteria	46.3
Chlorobi (MS-B_bin-24)	Bacteroidota	Bacteria	37.1
Parcubacteria (ADurb.Bin305)	Patescibacteria	Bacteria	31.5
Candidatus Bathyarchaeota (ex4484_205)	Thermoproteota	Archaea	25.4
Aciduliprofundum sp.	Thermoplasmatota	Archaea	20.1
Thermosulfidibacter takaii (ABI70S6)	Thermosulfidibacterota	Bacteria	18.7
Fervidicoccus sp.	Thermoproteota	Archaea	18.1
Thermodesulfovibrio aggregans	Nitrospirota	Bacteria	12.8
Bacterium, unclassified	not predicted	Bacteria	11.7
Candidatus Cryosericum odellii	Caldisericota	Bacteria	11.7
Archaeon, unclassified	not predicted	Archaea	11.6
Thermogutta terrifontis	Planctomycetota	Bacteria	10.7
Carboxydocella sp.	Firmicutes	Bacteria	10.5
Candidatus Woesearchaeota (ex4484_78)	Nanoarchaeota	Archaea	9.7
Bacterium, unclassified	not predicted	Bacteria	9.5
Candidatus Roizmanbacteria	Patescibacteria	Bacteria	9.1
Elusimicrobia (CG08_land_8_20_14_0_20_51_18)	Elusimicrobiota	Bacteria	8.1
Archaeon, unclassified	not predicted	Archaea	7.8
Verrucomicrobia	Verrucomicrobiota	Bacteria	7.8
Candidate division Zixibacteria (RBG_16_43_9)	Zixibacteria	Bacteria	7.6
Candidatus Omnitrophica	Omnitrophota	Bacteria	7.4
Candidatus Aenigmarchaeota	Aenigmatarchaeota	Archaea	7.2
Geobacter sp.	Desulfobacterota	Bacteria	6.8
Bacterium HR16	Armatimonadota	Bacteria	6.7
Heliorestis acidaminivorans	Firmicutes	Bacteria	6.5
Archaeon, unclassified	not predicted	Archaea	6
Thermofilum uzonense	Thermoproteota	Archaea	5.8
Fervidobacterium nodosum (Rt17-B1)	Thermotogota	Bacteria	5.4
Archaeon, unclassified	not predicted	Archaea	5.3
Thermodesulforhabdus norvegica	Desulfobacterota	Bacteria	5.2
Archaeon, unclassified	not predicted	Archaea	5.1
Candidatus Atribacteria (ADurb.Bin276)	Atribacterota	Bacteria	5.1
Desulfurella sp.	Campylobacterota	Bacteria	5.1
Ignavibacteriales (CG07_land_8_20_14_0_80_59_12)	Bacteroidota	Bacteria	4.3
Parcubacteria (ADurb.Bin305)	Patescibacteria	Bacteria	4
Candidatus Aminicenantes	Acidobacteriota	Bacteria	3.4
Candidatus Aminicenantes	Acidobacteriota	Bacteria	3.4

Nearest predicted relative based on BLASTp	Phylum	Domain	Abundance (rpS3)
Lentisphaeria	Verrucomicrobiota	Bacteria	3.4
Ignavibacteria (RBG_13_36_8)	Bacteroidota	Bacteria	2.7

Supplementary Table 3. Selected structural homologs of the putative esterase determined with HHpred.

Hit	Function	Origin	Probability [%]	E-value	ScoreOrigin
4FLE_A	Ancestral haloalkane dehalogenase AncHLD3;	Synthetic construct	99.96	7.8e-26	140.97
413F_A	serine hydrolase CCSP0084; MCP cleaving	*Cycloplasticus* sp. P1	99.93	4.3e-24	141.6
3N98_A	Chloroperoxidase F; Haloperoxidase, Oxidoreductase	*Pseudomonas fluorescens*	99.93	3.5e-24	138.8
3G9X_A	Haloalkane dehalogenase; alpha/beta hydrolase, helical cap domain	*Rhodococcus rhodochrous*	99.93	1.6e-23	137.84
4UHC_A	Esterase; alpha/beta hydrolase, pnp-ester cleaving	*Thermogutta terrifontis*	99.93	5.4e-24	138.98
3RM3_A	Thermostable monoacylglycerol lipase; alpha/beta hydrolase fold	*Bacillus* sp. H257	99.92	1.5e-23	136.84
3PFB_B	Cinnamoyl esterase; alpha/beta hydrolase fold, esterase, hydrolase, cinnamoyl/feruloyl esterase	*Lactobacillus johnsonii*	99.92	7.7e-23	133.82
1C4X_A	2-hydroy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (BPHD); PCB degradation	*Rhodococcus* sp. Strain Rha1	99.92	7.9e-24	138.27
5OLU_A	carboxyl esterase, 1,2-O-isopropylideneglycerol hydrolyzing, lipase, alpha/beta hydrolase	*Bacillus coagulans*	99.92	1.3e-22	136.85
4LXH_A	MCP Hydrolase; carbon-carbon bond hydrolase, Rossmann Fold, alpha/beta hydrolase fold	*Sphingomonas wittichii* RW1	99.92	2.2e-23	135.76
5Y6Y_B	Epoxide hydrolase	*Vigna radiata*	99.92	2.1e-23	138.9
2WTM_C	Promiscuous Feruloyl Esterase (Est1E)	*Butyrivibrio Proteoclasticus*	99.92	1.2e-22	130.56
4C6H_A	Haloalkane dehalogenase	Rhodobacteraceae	99.92	5.1e-23	135.33
2XTO_A	Haloalkane Dehalogenase	*Plesiocystis pacifica* SIR-I	99.92	3.5e-23	136.21
1Q0R_A	aclacinomycin methylesterase; Anthracycline, methylesterase, polyketide hydrolase	*Serratia marcescens*	99.92	9.6e-24	139.12
6F9O_A	haloalkane dehalogenase DpcA	*Psychrobacter cryohalolentis* K5	99.92	5.3e-23	136.58
6Y9G_B	Ancestral haloalkane dehalogenase AncHLD5	Synthetic construct	99.91	1.5e-22	133.85

Hit	Function	Origin	Probability [%]	E-value	ScoreOrigin
2E3J_A	epoxide hydrolase B (Rv1938)	*Mycobacterium tuberculosis*	99.91	1.3e-22	137.43
5XKS_F	Thermostable monoacylglycerol lipase	*Geobacillus* sp. 12AMOR	99.91	4.6e-22	128.73
2WFL_A	polyneuridine aldehyde esterase (PNAE)	*Rauvolfia serpentina*	99.91	3.3e-22	128.71
3BF7_B	Esterase YbfF; esterase, thioesterase	*Escherichia coli*	99.91	2.4e-22	128.91
6THS_A	Serine esterase, cutinase S165A	Uncultured bacterium	99.9	6.5e-22	129
5XWZ_B	Alpha/beta-hydrolase, lactonase, zearalenone hydrolase	*Cladophialophora bantiana*	99.9	3.8e-22	130.52
1IUP_A	meta-Cleavage product hydrolase; aromatic compounds, cumene, isopropylbenzene, meta-cleavage compound hydrolase	*Pseudomonas fluorescens* IP01 (CumD)	99.9	1.6e-21	127.29
6BA9_A	yersiniabactin synthesis enzyme, YbtT; Thioesterase, non-ribosomal peptide synthesis	*Escherichia coli*	99.9	3e-22	129.51
4CCY_A	Carboxylesterase YBFK, naproxenmethylester hydrolase	*Bacillus subtilis*	99.9	6.4e-23	135.19
2WUE_B	2-hydroy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (BPHD	*Mycobacterium tuberculosis*	99.89	2e-21	127.37
2RHW_A	BphD, C-C Bond Hydrolase Involved in Polychlorinated Biphenyls Degradation	*Burkholderia xenovorans* LB400	99.89	1.1e-21	127.92
3FCY_A	Xylan esterase 1; alpha/beta hydrolase, carbohydrate esterase, CE7	*Thermoanaerobacterium* sp. JW/SL YS485	99.89	8.9e-22	133.37
6AGQ_A	acetyl xylan esterase	*Paenibacillus* sp. R4	99.89	1.8e-21	130.01
1MJ5_A	1,3,4,6-tetrachloro-1,4-cyclohexadiene hydrolase; LINB, hydrolase, Haloalkane dehalogenase	*Sphingomonas paucimobilis* UT26	99.88	1e-20	125.53
5XH2_A	Poly(ethylene terephthalate) hydrolase	*Ideonella sakaiensis* 201-F6	99.88	5.6e-21	124.45
4CG1_A	Cutinase, PET degrading hydrolase	*Thermobifida fusca*	99.86	1.4e-19	119.52
7NEI_B	Polyester Hydrolase Leipzig 7 (PHL-7); PETase, Cutinase	unidentified	99.86	7.4e-20	119.78

Material and Methods

Sample collection and chemical labeling

Sediments for chemical labeling experiments have been sampled from the two hot springs 'Arkashin shurf' (54°30.0016N 160°00.2021E, 65.5-72° C, pH = 5.01, internal number #3811) and 'Helicopter spring' (54°30.003N 160°00.4375E, 58.1° C, pH = 5.62, internal number #3808), both located at the Uzon volcanic caldera (Kamchatka Peninsula, Russia). FP-alkyne was dissolved in DMSO. A slurry of sediments in spring water was collected and after gentle mixing, 10 ml of the slurry was dispensed in a reaction tube and incubated with 4 µM FP-alkyne for 2 h with occasional shaking while being placed back in the spring. An equal volume of DMSO was added to the negative controls. All samples were prepared in triplicates. For metagenomic sequencing, an unlabeled aliquot of the slurry was prepared. For downstream processing, the samples were transported to the laboratory on dry ice, the slurry was centrifuged to remove the spring water (12,000 × *g*, room temperature, 10 min) and the sediments were stored at -20° C until further processing.

DNA extraction and metagenomic sequencing

Total DNA was isolated from the sediments using phenol-chloroform extraction as described in Gavrilov *et al*.[1]. Prior to isolation, the cells were disrupted using a series of freezing-thawing cycles. Concentration of DNA was measured on a Qubit 2.0 fluorometer (Invitrogen, USA). Shotgun metagenome library preparation and sequencing were done at BioSpark Ltd., Moscow, Russia. The KAPA HyperPlus Library Preparation Kit (KAPA Biosystems, USA) was used for library construction according to the manufacturer's protocol and sequencing was perfromed on a NovaSeq 6000 platform (Illumina, USA) with the NovaSeq 6000 S2 Reagent Kit, which can read 100 nucleotides from each end (200 cycles).

Genome-resolved metagenomics

Raw reads were quality filtered and cleaned with BBDuk (https://jgi.doe.gov/data-and-tools/bbtools) and sickle[2] and followed by assembly with metaSPAdes[3] (version 3.14). Assembled sequences below 1 kbp in length were discarded and gene prediction was carried out using Prodigal in meta mode[4] followed by annotation against the FunTaxDB[5], which is based on the UniRef100[6] database. Community composition of samples was determined based on ribosomal protein S3 (rpS3) and its respective coverage on scaffolds in the individual metagenomes. The coverage was determine via mapping[7] of metagenomic reads and taken as the relative abundance of rpS3 genes and their respective microbes in the community. Binning

of genomes was performed with MaxBin2[8], ABAWACA[9] and emergent self-organizing maps[10]. High quality genomes were identified using DAS Tool[11] and further curated with uBin[5]. A phylogenetic tree was calculated with GTDB-Tk[12] based on the dereplicated genomes. Genome completeness, contamination, GC content and genome length was calculated via CheckM[13]. The target genome that contained the serine hydrolase chosen for biochemical characterization (see Heterologous protein expression and purification) was analyzed and re-annotated with NCBI's PGAP[14] (Prokaryotic Genomes Annotation Pipeline) to improve start and end of gene prediction.

Protein extraction and clean-up

For protein extraction from labeled sediments, the collected organic matter was thawed on ice and taken up in 5 mL of extraction buffer (100 mM Tris-HCl pH 8.8, 0.1 M DTT, 50 mM EDTA, 1.5% SDS, 30% sucrose). The cells were lysed in a three-step sonication procedure: seven iterations of sonication in an ultrasonic bath (BANDELIN electronic, Germany) for 1 min followed by vigorous mixing, 10 cycles with high power in a Bioruptor UCD-200 (Diagenode, Belgium) device with the following conditions: 1 min pulse and 30 sec pause and another ten iterations of sonication in an ultrasonic bath as described before. The extracts were then cleared by centrifugation (100 × *g*, room temperature, 5 min) and the supernatant was collected in a fresh tube. The debris were centrifuged again (15,000 × *g*, room temperature, 20 min) and the supernatant was combined with the supernatant in the fresh tube. The combined supernatant was subjected to another centrifugation step (15,000 × *g*, room temperature, 15 min) to separate the soluble protein containing fraction from any remaining debris. Protein clean-up was done by performing a phenol extraction[15] with downstream ammonium acetate precipitation[16] according to the literature with few modifications. In brief, the protein containing fractions were mixed and incubated (15 min, room temperature, shaking) with an equal volume of TE-buffered liquid Phenol (Carl Roth, Germany; #0038). To achieve phase separation, the samples were centrifuged (12,000 × *g*, room temperature, 10 min). The upper phenol phase was collected and re-extracted with extraction buffer thrice. Thereto, the phenol phase was mixed with an equal volume of extraction buffer and the phases were separated by centrifugation (12,000 × *g*, room temperature, 10 min). The proteins in the phenol phase were precipitated with a five-fold volume of 0.1 M ammonium acetate in methanol (-20 °C, overnight). The proteins were collected by centrifugation (12,000 × *g*, 4 °C, 30 min) and the pellet was successively washed twice with 0.1 M ammonium acetate in methanol, twice with 80% (v/v) acetone, and once with 70% (v/v) ethanol, respectively. Each washing step included an

incubation step at -20 °C for 20-30 min prior to centrifugation (12,000 × *g*, 4 °C, 30 min). The precipitated proteins were dissolved in 100 µL 8 M urea in 50 mM HNa_2PO_4, pH 8.0 and further diluted with 50 mM HNa_2PO_4 pH 8.0 to a final concentration of 2 M urea. The protein concentration of the resulting protein solutions was determined by a modified Bradford assay with Roti-Nanoquant (Carl Roth, Germany).

Click reaction and affinity purification

400-650 µg of total protein were subjected to a click reaction with 10 µM 5/6-TAMRA-biotin-N_3 (Jena Bioscience, Germany; #CLK-1048), 100 µM TBTA, 2 mM TCEP and 2 mM $CuSO_4$ (all purchased from Sigma-Aldrich, USA) in a total reaction volume of 500 µL (1 h, room temperature, in the dark). Prior to affinity purification, unbound reporter and salts from the click reaction were removed by methanol-chloroform precipitation[17]. The resulting protein pellet was air-dried and subsequently dissolved (37 °C, ~1 h) in 850 µL 2% (w/v) SDS in 1× PBS (155 mM NaCl, 3 mM Na_2HPO_4, 1.06 mM KH_2PO_4, pH 7.4). Insoluble particles were removed by centrifugation (21,000 × *g*, 37 °C, 5 min) and the cleared protein solution was diluted with 1× PBS to a final concentration of 0.2% (w/v) SDS. The obtained protein mixture was incubated with 100 µl of pre-equilibrated avidin beads slurry (Thermo Scientific, USA; #20219) while gently tumbling (~1h, room temperature, in the dark). Subsequently, the beads were washed five times with 10 mL 1% (w/v) SDS (10 min, room temperature, gently rotating) and collected by centrifugation (400 × *g*, 5 min). To remove SDS from the samples, the beads were then washed four times with 1 mL of ultrapure water (VWR Chemicals, USA; 5 min, room temperature, vigorously shaking) and collected by centrifugation (3,000 × *g*, 1 min).

On-bead digestion of captured proteins

After affinity enrichment, the beads were taken up in 100 µL 0.8 M urea in 50 mM ammonium bicarbonate (ABC). Disulfide bonds were reduced by adding 10 mM DTT (Sigma-Aldrich, USA) in 50 mM ABC (1 h, room temperature, vigorous shaking) and the generated cysteine mercapto groups were masked by alkylation with 25 mM Iodoacetamide (IAM; Sigma-Aldrich, USA) in 50 mM ABC (1 h, room temperature, in the dark, vigorous shaking). Excess IAM was then quenched by adding DTT (final concentration of 35 mM, 10 min, room temperature, vigorous shaking). Protein on-bead digestion was started by adding 1 µg Trypsin (Thermo Scientific, USA; #90057) dissolved in 50 mM acetic acid (37 °C, ~16 h, vigorous shaking). After digestion, the reaction solution was cleared by centrifugation (3,000 × *g*, room temperature, 5 min) and the supernatant (contains the digestion products (peptides)) was transferred to a fresh reaction vessel and the digestion reaction stopped by adding formic acid

(FA) to a final concentration of 5% (v/v). Next, the beads were washed with 50 µL 1% (v/v) FA and the supernatant was combined with the recovered digestion mix. To remove residual beads from the peptide solution, the mix was passed over a home-made two-disc glass microfiber membrane (GE Healthcare, USA; poresize 1.2 µm, thickness 0.26 mm) tip.

Sample clean-up for LC–MS

Peptides were desalted on home-made C_{18} StageTips[18] containing two layers of an octadecyl silica membrane (3M, USA). All centrifugation steps were carried out at room temperature. The StageTips were first activated and equilibrated by passing 50 µL of methanol (600 × *g*, 2 min), 80% (v/v) acetonitrile (ACN) with 0.5% (v/v) FA (600 × *g*, 2 min) and 0.5% (v/v) FA (800 × *g*, 3 min) over the tips. Next, the tryptic digests were passed over the tips (800 × *g*, 3-4 min). The flow-through was collected and applied a second time (same settings). The immobilized peptides were then washed with 50 µL and 25 µL 0.5% (v/v) FA (800 × *g*, 3 min). Bound peptides were eluted from the StageTips by application of two rounds of 25 µL 80% (v/v) ACN with 0.5% (v/v) FA (600 × *g*, 2 min). After elution from the StageTips, the peptide samples were dried using a vacuum concentrator (Eppendorf, Germany) and the peptides were dissolved in 15 µl 0.1% (v/v) FA prior to analysis by MS.

LC-MS/MS analysis

LC–MS/MS experiments were performed on an Orbitrap Fusion Lumos Tribrid instrument (Thermo Scientific, USA) that was coupled to an EASY-nLC 1200 liquid chromatography (LC) system (Thermo Scientific, USA). The LC was operated in the one-column mode. The analytical column was a fused silica capillary (75 µm × 46 cm) with an integrated PicoFrit emitter (New Objective, USA) packed in-house with Reprosil-Pur 120 C18-AQ 1.9 µm resin (Dr. Maisch, Germany). The analytical column was encased by a PRSO-V2 column oven (Sonation, Germany) and attached to a nanospray flex ion source (Thermo Scientific, USA). The column oven temperature was adjusted to 50 °C during data acquisition. The LC was equipped with two mobile phases: solvent A (0.1% (v/v) FA in water) and solvent B (0.1% (v/v) FA in 80% (v/v) ACN). All solvents were of UPLC grade (Honeywell, USA). Peptides were directly loaded onto the analytical column with a maximum flow rate that would not exceed the set pressure limit of 980 bar (usually around 0.5–0.8 µl min-1). Peptides were subsequently separated on the analytical column by running a 200 min gradient of solvent A and solvent B at a flow rate of 300 nl min-1 (gradient: start with 9% solvent B; gradient 9–40% solvent B for 180 min; gradient 40–100% solvent B for 15 min and 100% solvent B for 5 min).

The mass spectrometer was operated using Xcalibur software (version 4.3.73.11; Thermo Fischer Scientific) and was set in the positive ion mode. The ionization potential (spray voltage) was set to 2.3 kV. A top-speed data-dependent method with a cycle time of 3 seconds was selected for data acquisition. Precursor ion scanning (MS^1) was performed in the Orbitrap analyzer (FTMS; Fourier Transform Mass Spectrometry) at a resolution of 120,000 FWHM (full width at half maximum @ 200 m/z) in the scan range of *m/z* 375–1,500 with the internal lock mass option turned on (lock mass was *m/z* 445.12002, polysiloxane)[19]. The automatic gain control (AGC) was set to "standard" and the maximum injection time was machine determined ("auto"). Product ion spectra were recorded in the Orbitrap at a resolution of 15,000 FWHM. The scan range for MS^2 was set to "auto". Ions for fragmentation were selected in the quadrupole (isolation window of *m/z* 1.6) based on their intensity (threshold 5×10^4 ions) and charge state (only charge state of 2-7) in the full survey scan. The AGC target was set to "standard" and the maximum injection time to "auto". Selected precursor ions were fragmented by Higher-energy C-trap dissociation (HCD) with normalized collision energy (NCE) set to 30%. Monoisotopic precursor selection was enabled. During MS^2 data acquisition, dynamic ion exclusion was set to 60 s with a repeat count of 1 and a mass tolerance of ± 10 ppm.

Peptide and Protein identification using MaxQuant and Perseus

RAW spectra were submitted to an Andromeda[20] search in MaxQuant (version 1.6.17.0) using the default settings[21]. Label-free quantification was activated[22]. MS/MS spectra data were searched against the self-assembled metaproteome databases of the 'arkashin shurf' (45 649 entries) or the 'helicopter spring' (99 930 entries), accordingly. All searches included a contaminants database (as implemented in MaxQuant, 246 sequences). The contaminants database contains known MS contaminants and was included to estimate the level of contamination. Andromeda searches allowed for oxidation of methionine residues (16 Da) and acetylation of the protein N-terminus (42 Da) as dynamic modifications while carbamidomethylation of cysteine residues (57 Da, alkylation with IAM) was selected as static modification. Enzyme specificity was set to "Trypsin/P". The instrument type in Andromeda searches was set to Orbitrap and the precursor mass tolerance was set to ±20 ppm (first search) and ±4.5 ppm (main search). The MS/MS match tolerance was set to ±20 ppm. The peptide spectrum match FDR and the protein FDR were set to 0.01 (based on target-decoy approach). Minimum peptide length was 7 amino acids. For protein quantification, unique and razor peptides were allowed. In addition to unmodified peptides, modified peptides with dynamic modifications were allowed for quantification. The minimum score for modified peptides was set to 40.

Further data analysis and filtering of the MaxQuant output was done in Perseus[23] (version 1.6.14.0). Label-free quantification (LFQ) intensities were loaded into the matrix from the proteinGroups.txt file and potential contaminants as well as reverse hits from the reverse database and hits only identified based on peptides with a modification site were removed.

Biological replicates of the unlabeled controls and the FP-alkyne labeled samples were combined into two categorical groups to allow comparison of the samples. The data were transformed to the log2-scale and only those protein groups with a minimum of 2 identified unique peptides were kept in the matrix. Furthermore, only hits with a valid LFQ intensity for at least one of the probe-labeled sample replicates were selected for further analysis. Prior to quantification, missing values were imputed from a normal distribution (width 0.3, down shift 1.8). Comparison of normalized protein group quantities (relative quantification) between different MS runs was solely based on the LFQ intensities as calculated by MaxQuant (MaxLFQ algorithm)[22]. Briefly, label-free protein quantification was switched on and unique and razor peptides were considered for quantification with a minimum ratio count of 2. Retention times were recalibrated based on the built-in nonlinear time-rescaling algorithm. MS/MS identifications were transferred between LC-MS/MS runs with the "Match between runs" option in which the match time window was set to 0.7 min and the alignment time window to 20 min. The quantification was based on the "value at maximum" of the extracted ion current. At least two quantitation events were required for a quantifiable protein. The log2-fold enrichment of protein groups with FP-alkyne was calculated based on the mean LFQ intensity compared to the DMSO control. Protein groups with a negative fold enrichment were excluded from further analysis. The remaining protein groups were reported in the respective figure (Fig. 4).

Bioinformatic analyses of enriched proteins

In order to confidently predict potential serine hydrolases among the group of proteins that was enriched with FP-alkyne, the protein sequences of the respective proteins were analyzed using various tools and databases. Sequence similarity to deposited serine hydrolases including the presence of characteristic serine hydrolase domains was analyzed using UniRef100[6], PFAM[24], NCBI CDD[25] and InterProScan[26]. Structural homology to known serine hydrolases was assessed using the SWISS-MODEL template library[27] and HHpred[28].

Heterologous protein expression and purification

The putative esterase selected for biochemical characterization (identifier: ExploCarb_3811S_S4_483_length_13114_cov_941_5) was heterologously expressed in *E. coli*. Thereto, *E. coli* Rosetta cells (Novagen, USA) were transformed with the commercially obtained construct (BioCat, Germany) of the codon-optimized gene cloned into a pET-28b(+) vector encoding a C-terminal 6× His-tag. For recombinant expression of the UPF0227 gene, a freshly inoculated 1 L culture in LB medium supplemented with 50 µg mL^{-1} kanamycin and 50 µg mL^{-1} chloramphenicol was grown to an OD_{600} of 0.4 at 37 °C with constant shaking (180 rpm) until subsequent induction of the protein expression with 500 µM isopropyl-β-D-thiogalactopyranoside (IPTG). Upon further incubation at 18 °C for 16 h, the cells were harvested by centrifugation (8,000 × *g*, 4 °C, 20 min) and the resuspended in 5 mL 50 mM Tris-HCl pH 7.0 per gram wet weight of the pellet. Cell lysis was performed by sonication in three cycles for 5 min (cycle 0.5, amplitude 50) with a UP 200S sonicator (Hielscher Ultrasonics, Germany). The crude extract was cleared by centrifugation (12,000 × *g*, 45 min, 4 °C) and the lysate was passed through a 0.45 µm filter. Protein affinity purification was done using a Protino™ Ni-TED 1000-packed column (Macherey-Nagel, Germany) according to the manufacturer's instructions. Prior to further clean-up of the recombinant protein by size exclusion chromatography (SEC), the elution buffer was exchanged with size exclusion buffer (50 mM Tris-HCl pH 7.5, 20 mM NaCl) by centrifugation (6,000 × *g*, room temperature, 40 min) using Amicon® centrifugal filter devices (10 kDa cutoff, Merck, Germany) and the solution was concentrated to 1 mL. SEC was performed on a HiLoad® 16/600 Superdex® 200 pg column (GE Healthcare, USA) connected to an ÄKTA™ FPLC system (GE Healthcare, USA) at a flow rate of 1 mL min^{-1}. Fractions containing the recombinant protein (detection at 280 nm) were pooled and concentrated as described above. For long-time storage at -80 °C, 50% (v/v) glycerol was added to the protein solution that was flash-frozen in liquid nitrogen.

Structural and bioinformatics analysis of the putative esterase

Retrieval of homologous sequences, structures and domains was conducted with BLAST (blastp tool)[29], HHpred[28] and HMMER (phmmer tool)[30], respectively. For further structural analysis and structure comparison, a model of the UPF0227 protein was constructed with AlphaFold (version 2.0) using default settings[31]. The resulting PDB file was used for visualization and processing of the protein structure with UCSF Chimera[32] (version 1.14).

Biochemical characterization of the heterologous protein

The activity of the heterologously expressed UPF0227 protein was determined using a continuous assay with the chromogenic *para*-nitrophenyl (pNP) substrates pNP-acetate, pNP-butyrate, pNP-octanoate, pNP-decanoate and pNP-dodecanoate (all obtained from Megazyme, Ireland). The increase in absorbance at 384 nm was measured using a Specord 210® photometer (Analytik Jena, Germany) and the enzymatic activity was calculated from a calibration curve with pNP. The pH and temperature optimum of the enzyme was determined with pNP-butyrate prior to assessing the enzyme kinetics with the different pNP-esters at the respective pH and temperature using substrate concentrations up to 0.7 mM. To test the inhibition effect of the FP-alkyne probe used for ABPP, 6 µg/ml protein in 50 mM Tris-HCl pH 7.5, 20 mM NaCl were incubated with 2 or 4 µM of the probe for 10 min at 70 °C. The protein solutions were then submitted to activity measurement with pNP-acetate to determine the residual esterase activity compared to a DMSO-treated control.

***In vitro* labeling**

In vitro ABPP with FP-alkyne was performed by incubation of indicated amounts of the enzyme with 4 µM of the probe (1 h, 70 °C) in a final reaction volume of 50 µL. An equal volume of DMSO was added to the negative controls. Preincubation with paraoxon-ethyl (Sigma-Aldrich, USA), if applicable, was done at a final concentration of 100 µM (15 min, 70 °C). The subsequent click reaction was performed as described above (Click reaction and affinity purification) using Cy3-N_3 (synthesized in house) as click tag. For gel-based analysis, the samples were mixed with 1 equivalent 4× LDS gel loading dye (423 mM Tris HCl, 563 mM Tris base, 8% (w/v) lithium dodecyl sulfate (LDS), 40% (w/v) glycerol, 2 mM EDTA, 0.075% (w/v) SERVA Blue G250; supplemented with 100 mM DTT) and incubated at 70 °C for 15 min. Separation of proteins (1/5 of the initial amount) by gel electrophoresis was done on a 11% Bis-Tris resolving gel, followed by visualization of labeled proteins using a Typhoon FLA 9000 laser scanner (GE Healthcare, USA).

Supplementary References

1 Gavrilov, S. N. *et al.* Isolation and Characterization of the First Xylanolytic Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1 and Its Unusual Multidomain Glycosidase. *Front. Microbiol.* **7**, 552 (2016).

2 Joshi, N. & Fass, J. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33) [Software]. (2011).

3 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824-834 (2017).

4 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

5 Bornemann, T. L. V., Esser, S. P., Stach, T. L., Burg, T. & Probst, A. J. uBin – a manual refining tool for metagenomic bins designed for educational purposes. *BioRxiv* 2020.2007.2015.204776 (2020).

6 Suzek, B. E. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926-932 (2015).

7 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).

8 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).

9 Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208-211 (2015).

10 Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).

11 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836-843 (2018).

12 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927 (2020).

13 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043-1055 (2015).

14 Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614-6624 (2016).

15 Wang, W. *et al.* Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis* **24**, 2369-2375 (2003).

16 Benndorf, D. *et al.* Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. *Biodegradation* **20**, 737-750 (2009).

17 Wessel, D. & Flügge, U. I. A Method for the Quantitative Recovery of Protein in Dilute-Solution in the Presence of Detergents and Lipids. *Anal. Biochem.* **138**, 141-143 (1984).

18 Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896-1906 (2007).

19 Olsen, J. V. *et al.* Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell Proteomics* **4**, 2010-2021 (2005).

20 Cox, J. *et al.* Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10**, 1794-1805 (2011).

21 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

22 Cox, J. *et al.* Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell Proteomics* **13**, 2513-2526 (2014).

23 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731-740 (2016).

24 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412-D419 (2021).

25 Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265-D268 (2020).

26 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).

27 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252-W258 (2014).

28 Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237-2243 (2018).

29 McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**, W20-W25 (2004).

30 Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200-W204 (2018).

31 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

32 Pettersen, E. F. *et al.* UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).

Chapter 3.3

Identification and characterization of a prevalent thermophilic β-glucosidase from a hot spring enrichment metagenome in Kamchatka identified with Activity Based Protein Profiling

Identification and characterization of a prevalent thermophilic β-glucosidase from a hot spring enrichment metagenome in Kamschatka identified with ABPP

(preliminary manuscript for submission)

Thomas Klaus[1#], Sabrina Ninck[2#], Sarah P. Esser[3], Tatiana Kochetkova[4], Alexander Elcheninov[4], Michaela Bojara[1], Farnusch Kaschani[2], Christopher Bräsen[1], Ilya Kublanov[4], Markus Kaiser[2], Alexander J. Probst[3], Bettina Siebers[1]

[1]Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Faculty of Chemistry, Centre for Water and Environmental Research (CWE), University of Duisburg-Essen, Essen, Germany,

[2]Chemical Biology, Center of Medical Biotechnology, Faculty of Biology, University of Duisburg-Essen, Germany

[3]Group for Aquatic Microbial Ecology (GAME), Environmental Microbiology and Biotechnology (EMB), centre for Water and Environmental Research (CWE), University of Duisburg-Essen, Essen, Germany

[4]Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia

Corresponding authors: Bettina Siebers and Markus Kaiser

#both authors contributed equally

Keywords: Glycosid Hydrolases, Cellulose degradation, Activity-Based Protein Profiling, Archaea, *Caldivirga maquilingensis*, Functional metagenomics

Abstract

An amorphous cellulose (AMC) enrichment culture in a hot spring in the Uzon caldera has been analyzed with regards to its microbial composition and the encoded carbohydrate active enzymes (CAZymes) using metagenomics, gene abundance ranking and activity based protein profiling (ABPP). The creanarchaeon *Caldivirga maquilingensis* has been identified as a mayor source of genes encoding active glycosidases within the constructed metagenome assembled genomes (MAGs) from the surveyed enrichment culture. To identify enzymes that are active during growth on cellulose, activity based protein profiling (ABPP) was conducted *in situ*. From the enzymes labelled with the applied activity based β-glucosidase probe JJB111, the β-glucosidase AMC3811_001245 has been selected for heterologous expression and thorough characterization, revealing an extremely thermo- and solvent stable glycosidase with exo-glucosidase activity on various glucose disacharides and polymeric carboxyl methyl cellulose (CMC) over a broad range of pH values with a temperature optimum at 100 °C.

Introduction

For the last decades, hot springs have been thoroughly investigated with respect to their potential to harbour extremophilic microorganisms which are able to thrive under high temperatures [1–3]. Microbes and microbial communities from hot springs have atracted scientific interest, since their physiology and ecology differs from mesophiles and understanding the molecular mechanisms needed to adapt to extreme environments is crucial to shed light upon ecological and evolutionary issues of microbial life [4,5]. In contrast to mesothermal environments, hot spring ecosystems usually exhibit a wide range of extreme conditions with pH values ranging from below 1 to above 10, temperatures ranging from moderate high temperatures up to more than 100 °C and unique geochemical features (metal composition, reducing agents), thus challenging the microbes thriving under these conditions [6–8]. For industrial purposes, enzymes from (hyper)thermophiles are especially important, since they are valuable and efficient biocatalysts which can withstand harsh conditions such as high temperatures above 100 °C, organic solvents and a broad range of pH values. Processes at high temperatures are oftenly neccesary due to pretreatment of subtrates, or exothermic reactions [9–11]. Moreover it is notable, that reactions at high temperatures are usually desired due to better substrate solubility and lower risk of contamination [12,13]. A variety of enzyme classes is currently being applied in industrial processes, including for example oxido-reductases, esterases, proteases, lipases, isomerases and different glycoside hydrolases, whereas especially the latter are getting more important for green chemistry application and for the production of valuable products such as biofuels from renewable raw materials [14,15]. Enzymes capable of hydrolyzing (ligno)cellulosic biomass are highly desired, since cellulose is considered the most abundant biopolymer on earth [16]. With novel techniques and progress in the field of (meta)genomics, the polymer degrading potential of hot springs gets more accesible, revealing a large resoirvoire of cultured and uncultured microorganisms [17,18]. Thus, many novel polymer degrading (hyper)thermophilic strains with interesting features such as *Caldivirga maquilingensis* [19], *Paenibacillus strain Y4.12MC10* [20], *Thermococcus sp. strain 2319x1* [21], *Caldanaerobacter sp. strain 1523vc* [22], *Thermofilum adornatum* [23] *and Thermosphaera sp. Strain 3507* [24] have been described over the last years and most likely many more will follow. Moreover, metagenomic analysis as well as other culture independend techniques have shed light upon the high number and variability of CAZymes present in hot ecosystems [25]. The Uzon caldera in Kamchatka, far eastern Russia, represents an active hydrothermal system, containing several different thermal fields, featuring various hydrothermal characteristics and an extensive reservoir of microbial diversity and is thus a promising sampling location for ecological and

biotechnological research on (hyper)thermophilic microbes [26–30]. Despite several studies have been published about the ecology of hot springs and the respective contribution of single microbial species to the hydrolytic activity within the sampled systems, sufficient functional analyses are still needed to fully decipher how breakdown of complex carbohydrates is achieved in individual hot springs [18,25,31]. Therefore, functional metagenomics, classical cultivation and characterization of carbohydrate active microorganisms, enrichment cultures and classic protein biochemistry have been proven as appropiate tools [25,32–34]. Still, these approaches are oftently missing valid information about the actual activity state *in vivo* of individual enzymes involved in metabolic processes and breakdown of polymers. For pure cultures or microbial communities with low complexity, Activity Based Protein Profiling (ABPP) is a well established technique, which can be used to lable class specific active enzymes *in vivo* under the surveyed conditions, with subsequent LC- MS/MS based quantification and identification of labelled proteins [35–37]. Thus, ABPP is a valuable tool to close the information gap between *omics* based data and functional analysis of enzymes present within an ecosystem. A broadrange of activity based probes (ABPs) for different enzyme functionalities has been developed over the last two decades, particularly for hydrolases, such as proteases and glycoside hydrolases [38–40]. Common probes consist of an electrophilic substrate analogous warhead, which covalently binds to the nucleophile of the targeted enzymes active centre. Via a linker region, the warhead is connected to a reporter- or affinity tag, which can be used to visualize or enrich the labelled protein [41]. Yet so far, ABPP has mostly been applied in relatively homogenous samples, such as eukaryotic cell culture and pure prokaryotic cultures, whereas surveys on heterogenic microbial communities and even environmental samples are sparse due to the complexity of these samples [42–45]. Moreover, most ABPP studies have been performed in mesophilic, moderate environments, whereas studies with extremophiles are scarce. Up to date, there are only two examples for succesfull ABPP in thermophilic archaea [46,47]. In this work, we analyzed a thermophilic enrichment culture with amorphous cellulose (AMC) from the Arkashin shurf thermal pool in the Uzon caldera region (Kamschatka, Russia) with respect to its hydrolytic capabilities. Besides analysis of microbial community composition and identification of CAZymes from the constructed metagenome, we applied ABPP to label and identify active β-glucosidases directly in the enrichment culture and subsequently heterologously expressed the most confidently labelled β-glucosidase and characterized it as a highly thermo- and solvent stable glycoside hydrolase.

Abbreviations

AA	Auxiliary activities
ABP	Activity-based probe
ABPP	Activity-based protein profiling
AMC	Amorphous cellulose
CAZymes	Carbohydrate-active enzymes
CBM	Carbohydrate binding module
CE	Carbohydrate esterase
CMC	Carboxymethyl cellulose
DMSO	Dimethylsulfoxide
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
GH	Glycosyl hydrolase
GT	Glycosyl transferase
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Lysogeny broth
LC	liquid chromatography
LFQ	Label-free quantification
MAG	Metagenome assembled genome
MS	Mass spectrometry
SLH	s-layer homology domain protein
ONPG	Ortho-nitrophenyl-β-D-galactopyranoside
PBS	Phosphate buffered saline
PL	polysaccharide lyase
pNP	Para-nitrophenol
pNPG	Para-nitrophenyl- β-D-glucopyranoside
pNPGal	Para-nitrophenyl-ß-D-galactopyranoside
pNPX	Para-nitrophenyl-β-D-xylopyranoside

Material and Methods

Preparation of enrichment culture for metagenome analysis and ABPP experiments

An enrichment culture with amorphous cellulose (AMC) was prepared in a hot spring located in the Uzon volcanic caldera (Kamchatka Peninsula, Russia). The 'Arkashin shurf' (54°30.0016N 160°00.2021E, 65.5-72 °C, pH = 5.01, internal number #3811) has been chosen for this approach based on its physical and chemical parameters, as well as previous studies on its microbial communitites. A 1L glass bottle was completely filled with spring slurry (5-10 ml of sediment with spring water) and 10 ml AMC solution (1 g/l) was injected. Then the bottle was closed with two needles remaining in the septum to enable water and gas exchange. After 7 days incubation in the spring, the culture was harvested and concentrated by filtration to 200 ml with a 0.22 µm filter.

For every ABPP sample (samples and negative controls in triplicates) 10 ml of the concentrated culture were incubated with 2 µl JJB111 (final concentration 4µM) or 2 µl DMSO (negative controls) and then incubated for 2 h inside the spring with occasional shaking. For metagenomic sequencing, the filter which was used for concentration of the culture was furtherly processed. For downstream processing, the samples were centrifuged (12 000 × g, room temperature, 10 min) and the pellets stored at dry ice until further use.

DNA extraction and sequencing

Total DNA was isolated from the half ot the filter containing biomass of the *in situ* enrichment culture using phenol-chloroform extraction (Gavrilov et al., 2016). Prior to isolation, the cells were disrupted using a series of freezing-thawing cycles. Concentration of DNA was measured by a Qubit 2.0 fluorimeter (Thermo Fischer Scientific, USA). The DNA sample was used for shotgun metagenome library construction with a KAPA HyperPlus Kit (KAPAbiosystems, UK). Sequencing was perfromed using the NovaSeq 6000 platform (Illumina, USA) with NovaSeq 6000 S2 Reagent Kit (200 cycles).

Metagenome analysis and CAZyme profiling

Metagenomic raw reads were quality controlled using bbduk (https://jgi.doe.gov/data-and-tools/bbtools/) and sickle (Joshi and Fass 2011); high quality reads were assembled using metaSPAdes (version 3.14) [48]. After discarding scaffolds <1 kbp, metagenome assembled genomes (MAGs) were binned using MaxBin [49], tetranucleotide-based emergent self-organizing maps (ESOM) [50] (window 5 kb to 10 kb), abawaca [51] and consolidated using dasTool [52]. uBin [53] was used to manually curate the MAGs followed by quality control with

checkM [54]. All genomes with completeness >70% and contamination <5% were used in downstream analyses. Proteins within the metagenome and the MAGs were predicted using prodigal [55] in meta and default mode, respectively. Annotation of genes was performed against UniRef100 [56,57]. Genomes were classified using gtdb-tk [58,59]. The enzymatic profile of the metagenomic assembled genomes (MAGs) and of the entire, assembled metagenome were determined via HMM (hidden markov model) search (version 3.2, June 2018) [60] against the dbCAN2 CAZyme database [61]. Relative abundancies of individual species within the microbial community in the enrichment culture were derived from coverage of respective rpS3 genes.

Extraction, processing and analysis of Proteins

The enrichment culture samples obtained as described previously in the methods section were furtherly processed. The proteins were extracted, affinity purified and on-bead digested for subsequent LC-MS/MS analysis and protein identification as described by Ninck et al (Ninck et al., 2022).

Bioinformatical analysis of selected glycosidase

Gene and protein sequences of AMC3811_001245 have been obtained from the metagenome data as described above and were analyzed using BLAST [62] and PHHMMER [63]. Presence of secretion signal peptides and transmembrane domains was predicted using SignalP5 [64], Phobius [65] and TMHMM [66]. A structural model of AMC3811_001245 was constructed using Swiss-Model [67] with the crystall structure of the β-glucosidase 3B from *Thermotoga neapolitana* in complex with alpha-D-glucose (2X42) as template and was visualized with UCSF Chimera [68].

Expression of selecteted glycosidase gene

From the glycosidase gene sequences derived from the metagenome (Supplementary file 1), AMC3811_001245 has been choosen for gene synthesis with codon optimization for expression in *E. coli*, using a and additional restriction sites (NdeI, XhoI) and was then cloned with standard restriction-ligation cloning methods into pET28a(+) vectors by BioCat (BioCat GmbH, Germany) for heterologous expression, providing the recombinant protein with a N-terminal 6× His-tag. Heterologous expression of the glycoside hydrolases was performed in *E. coli* Rosetta cells (Novagene, USA) upon transformation with the respective plasmids. A freshly inoculated 1 l culture in LB medium, supplemented with 50 µg ml^{-1} kanamycin and 50 µg ml^{-1} chloramphenicol, was grown to an OD_{600} of 0.6 at 37 °C with constant shaking (180 rpm), followed by induction of protein expression with 500 µM isopropyl-β-D-thiogalactopyranoside

(IPTG). The cells were incubated at 18 °C for further 16 h, harvested by centrifugation (8,000 × *g* for 20 min at 4°C) and resuspended in 3-5 mL of buffer (50 mM Tris-HCl pH 7.0) per gram wet weight for further protein purification

Purification of heterologously expressed *AMC3811_001245*

The cell suspension was subjected to sonification with a UP 200S sonicator (Hielscher Ultrasonics GmbH, Germany) for 3 × 5 min (cycle 0.5, amplitude 50) followed by centrifugation at 8,000 × *g* for 45 min at 4 °C to remove intact cells and cell debris. For further purification of the GH3 protein *AMC3811_001245*, the cleared lysate was passed through a 0.45 µm filter, followed by an affinity purification using a Protino® Ni-TED 1000 column as described in the manufacturers manual (Macherey-Nagel, Germany) The elution buffer was exchanged for storage buffer (50 mM Tris-HCl pH 7.0, 20 mM KCl, 5 mM $MgCl_2$) and the protein solution was concentrated using Amicon® centrifugal filter devices (30 or 50 kDa cutoff, Merck, Germany). For further purification via size exclusion chromatography (SEC), 1 ml of the concentrated protein solution (2.6 mg ml^{-1}) was applied on a HiLoad® 16/600 Superdex® 200 pg column (Sigma Aldrich, USA), connected to an ÄKTA™ Purifier (GE Healthcare, USA) with a flow of 1 ml/min. Collected fractions containing the protein (detection at 280 nm) were pooled and concentrated as described above and either used directly or stored at 4 °C.

Enzyme characterization

Temperature and pH dependent activity, enzyme kinetics, as well as product inhibition of AMC3811_001245 was determined photometrically in a continuous assay, using the pNP substrates p-Nitrophenyl β-D-glucopyranoside (pNPG), *p*-Nitrophenyl β-D-xylopyranoside (pNPX) and *o*-Nitrophenyl β-D-galactopyranoside (oNPGal); The respective substrates were incubated at concentrations up to 4 mM with 8 µg/ml AMC3811_001245 and hydrolysis was determined from the increase of absorption at 384 nm. For determination of thermostability, the purified protein was incubated in McIlvaine buffer pH 5 [69] at 60 °C, 80 °C or 100 °C for up to 24 h. Residual activity was then measured with pNPG as described above at 100 °C and pH 5. Kinetics were determined in McIlvaine Buffer at the optimal pH (5) and temperature (100 °C). To determine the effect of solvents and cations on the glucosidase activity, freshly purified protein was preincubated 10 min at 100° C with 10-30% of the respective solvent or 1 mM of the respective ion and subsequently the residual activity was measured as described above. Hydrolysis of natural substrates was determined with cellobiose, carboxymethylcellulose (CMC), Avicel® cellulose, curdlan, lichenan, laminaribiose and

gentiobiose in a discontinous assay; 0.5% (w/v) substrates were incubated with 8 µg/ml AMC3811_001245 in McIlvaine buffer pH 5 at 100 °C for 5 min. After incubation, the amount of glucose inside the samples was determined using a discontinuous enzymatic assay with hexokinase and glucose-6-phosphate dehydrogenase as auxiliary enzymes ; 1 mM NAD^+, 1 mM ATP, 1 U/ml hexokinase from rabbit muscle (Sigma Aldrich, USA), 1 U/ml glucose-6-phosphate dehydrogenase (Sigma Aldrich, USA) in 50 mM Tris pH 7 were incubated with 100 µl of the pre-incubated samples for up to 10 min at 40 °C and the resulting increase of absorbance at 340 nm was used to calculate the amount of glucose formed within the pre-incubated sample. For kinetic characterization, time dependent formation of glucose was followed at different substrate concentrations (0-20 mM). One unit (U) is defined as 1 µmol product per min per mg protein.

ABP labelling of the heterologously expressed glycosidase

To determine the inhibitory effect of the covalently bound ABP JJB111 to the heterologously expressed β-glucosidase, 0.25 mg/ml AMC3811_001245 in McIlvaine buffer pH 5 was incubated with JJB11 concentrations between 0 to 8 µM at 100° C for 10 min. Subsequently, a kinetic characterization was performed in presence of the varying ABP concentrations using pNPG as substrate.

Results

Community composition of AMC enrichment

The Arkashin Shurf (65.5 -71° C, pH 5.01) has been choosen for an *in situ* enrichment culture with AMC. After seven days incubation, samples for metagenomic analysis and further processing were taken (Figure 1A). The rank abundance of organisms identified in the hot spring enrichment was constructed using the coverage of rps3 genes from the metagenome data. Thus, the majority of rps3 genes quantified in the enrichment was assigned to mainly crenarchael species with *Caldisphaera lagunensis* (26%) and *Pyrobaculum ferrireducens* (13.28%) as the most dominant archaeal species. Other archaeal phyla such as Euryarchaeota (6.75% *Aciduliprofundum* species) and Nanoarchaeota (4.52% *Nanopusillus* sp. Nst1) were distinctly less abundant. However, bacterial species account to 33.06% of the community composition with *Sulfurihydrogenibium azorense* (29.86%) beeing the most abundant species overall (Figure 1B).

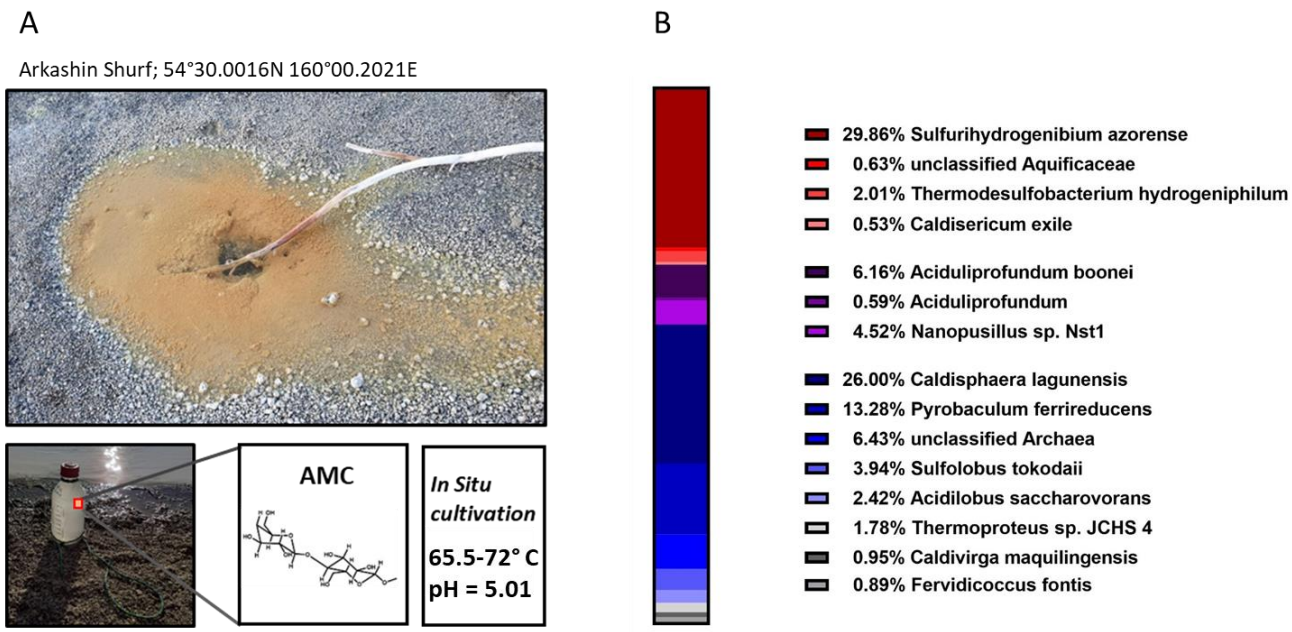


Figure 1: Sampling site and composition of microbial community. (A) *The enrichment culture with AMC was prepared in the Arkashin shurf with slurry and spring water and was incubated in situ for 7 days.* ***(B)*** The composition of the microbial community within the AMC enrichment culture according to rps3 abundance ranking.

Abundance of CAZymes identified in the AMC enrichment metagenome

The constructed MAGs, as well as the entire assembled metagenome of the AMC enrichment culture were analyzed by HMM search against the dbCAN2 CAZyme database to construct a CAZyme profile, resulting in 3878 genes identified as potential CAZymes of which 1990 could be assigned to individual MAGs. In total, 196 genes were identified as auxiliary activites (AA), 38 as carbohydrate binding modules (CBM), 120 as carbohydrate esterases (CE), 2329 glycosyl transferases (GT) and 1024 glycoside hydrolases (GH) (Figure 2A). Of the 50 different GH families identified, the families GH1, GH2, GH3, GH5, GH8, GH10, GH12, GH74, GH94 and GH116 are reported to contain activities involved in cellulose degradation, such as cellulases (EC number: 3.2.1.4), endo-1,3(4)-β-glucanases (EC number: 3.2.1.6), (exo) β-glucosidases (EC number: 3.2.1.21, 3.2.1.74) and cellulose phosphorylase (EC number: 2.4.1.20) and overall, 200 genes belonging to these families were identified in the metagenome. Genes of the GH families GH57, GH5 and GH13 were quantified the most, with 296, 104 and 93 hits respectively. Interestingly, the majority of GH genes which could be assigned to MAGs belonging to crenarchaeal species (46.98% of all GH genes), of which *Caldivirga maquilingensis* genes contribute the most (15.43% of all GH genes). Bacterial GH genes on the other hand, contribute only 11.71% of the identified GH genes (Figure 2B). However, a major part of GH genes (37.4%) could not be assigned to a binned MAG.

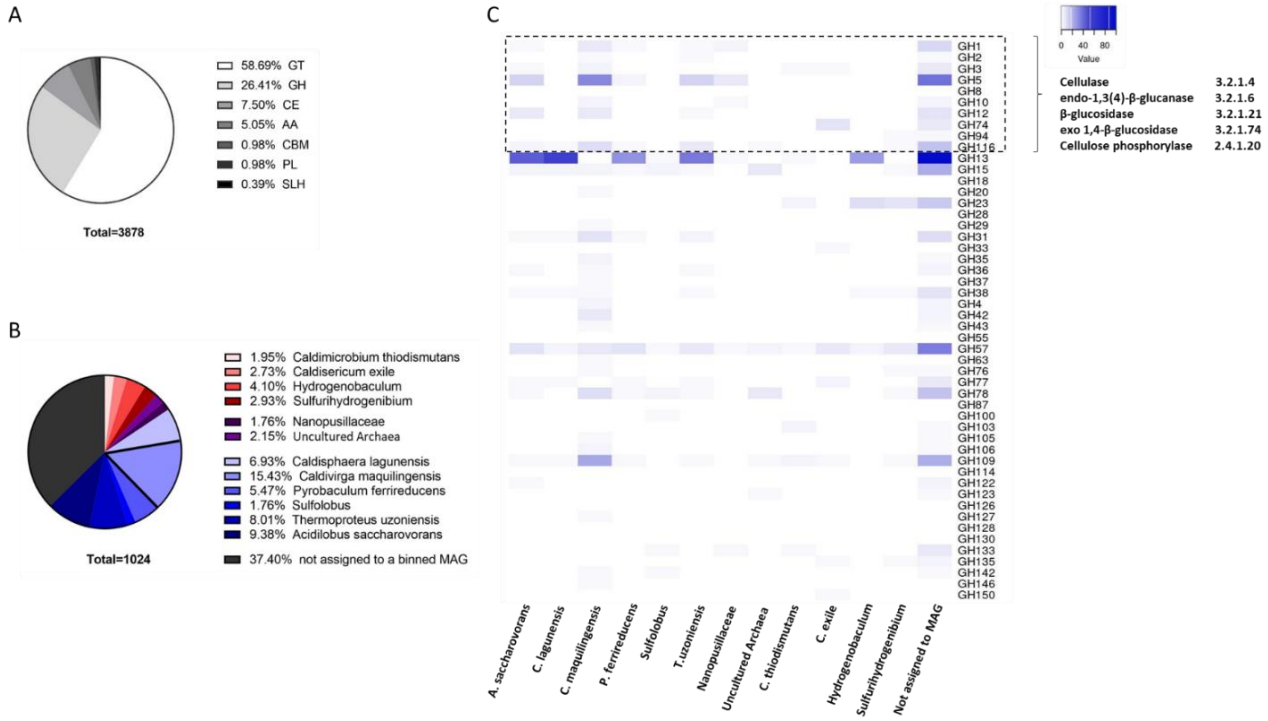


Figure 2: Abundance and distribution of identified CAZymes in the enrichment metagenom. (A) *Relative quantities of identified CAZyme classes within the identified 3878 genes encoding CAZymes. The mayority of CAZymes were identified as glycosyl transferases (GT) or glycosyl hydrolases (GH). Distinctly less carbohydrate esterases (CE), auxiliary activities (AA), carbohydrate binding modules (CBM), polysaccharide lyases (PL) and s-layer homology domain proteins (SLH) were identified.* ***(B)*** *Assignment of identified GH genes to binned MAGs.* ***(C)*** *Heatmap of identified GH families. Deep blue fields indicate higher numbers of identified genes corresponding to the respective GH family and the respective organism.*

CAZymes identified with ABPP

From the vast number of identified glycosidase genes within the enrichment culture, ABPP was applied to identify active β-glucosidases with the ABP JJB111, which is a biotinylated, N-alkylated aziridine analogue of the β-glucosidase inhibitor cyclophellitol. It structurally and configurationally resembles β-glucopyranosides and therefore preferentially reacts with β-glucosidases [70]. This probe was applied on samples of the enrichment culture with subsequent *in situ* incubation for 1h to lable active β-glucosidases. eight β-glycosidases have been identified within the samples labelled with the activity based β-glucosidase probe but not within the control samples without probe, indicating that these proteins are active under the *in situ* conditions. Interestingly, six of them belong to *Caldivirga maquilingensis*, what is beeing in accordance with the CAZyme profile obtained from the metagenomic analysis. The β-glucosidase with the highest MS/MS counts (identifier: ExploCarb_3811AMC_S5_48_length_58541_cov_8_56) was choosen for heterologous

expression and biochemical characterization to elucidate its contribution to the microbial degradation of AMC in the spring #3811 and its biotechnological potential (protein identifier: AMC3811_01245).

Table 1: Glycosidases identified with the β -glucosidase ABP JJB111 within the AMC enrichment culture prepared in the Arkashin shurf (65-72° C, pH 5.01).

Identifier	Annotation	Organism	MS/MS counts
ExploCarb_3811AMC_S5_264_length_20463_cov_9_19	Cellulase_domain-containing_protein	Caldivirga maquilingensis IC-167	21
ExploCarb_3811AMC_S5_21_length_87524_cov_25_7	β-galactosidase	Acidilobus saccharovorans 345-15	17
ExploCarb_3811AMC_S5_48_length_58541_cov_8_56	**Glycoside_hydrolase_family_3_ protein**	**Caldivirga maquilingensis IC-167**	**11**
ExploCarb_3811AMC_S5_645_length_9897_cov_11_3	β-galactosidase	Thermoproteus uzoniensis 768-20	10
ExploCarb_3811AMC_S5_409_length_15234_cov_9_7	Glycoside_hydrolase_family_3_ protein	Caldivirga maquilingensis IC-167	9
ExploCarb_3811AMC_S5_136_length_34245_cov_9_32	Glycoside_hydrolase_family_1	Caldivirga maquilingensis IC-167	3
ExploCarb_3811AMC_S5_42_length_61860_cov_10_5	Glycoside_hydrolase_family_1	Caldivirga maquilingensis IC-167	3
ExploCarb_3811AMC_S5_543_length_11672_cov_10_7	β-galactosidase	Caldivirga maquilingensis IC-167	2

Heterologous expression and characterization of the β-glucosidase AMC3811_01245

Quantification of proteins labelled with the β-glucosidase ABP JJB111 revealed AMC3811_01245 as the most abundant activity labelled β-glucosidase, suggesting that this GH is a major contributor of cellulose degradation in the analyzed enrichment. AMC3811_001245 belongs to the GH family 3, has a calculated size of 78.78 kDa and according to PHMMER prediction, the 705 amino acid (aa) containing glycosidase consists of three functional domains: a N-terminal glycosyl hydrolase family 3 domain (aa 25-282), consisting of a $(\alpha/\beta)_8$ TIM barrel structure, a C-terminal five-stranded alpha/β sandwich glycosyl hydrolase family 3 domain (aa 316-583) and a Fibronectin type III-like (FnIII) domain at the C- terminus (aa 615-686). A model displaying the three-dimensional arrangement of these domains was constructed with Swiss-Model, using the structure of the β-glucosidase 3B from *Thermotoga neapolitana* (PDB identifier 2x42, 58.48% sequence identity, global QMEANDisCo 0.85) as template and visualized with Chimera 1.15 (Figure 3A). According to a structural alignment with the crystal structure 2x42, the substrate is putatively bound within a pocket at the interface of the N- and

C-terminal GH3 domains (Figure 3B, C), whereas the FnIII domain is not directly involved. However, for other FnIII domains in different GHs, function in substrate binding [71] or cellulose-fiber loosening [72] has been proposed. BLAST analysis with the obtained sequence revealed 87.68% identity to a glycosyl family 3 domain containing protein from *C. maquilignensis* IC- 167, 78.73% identity to a glycosyl family 3 domain containing protein from *Caldivirga* sp. MU80 and 78.59% identity to a glycosyl hydrolase from *Caldivirga sp. JCHS 4.*

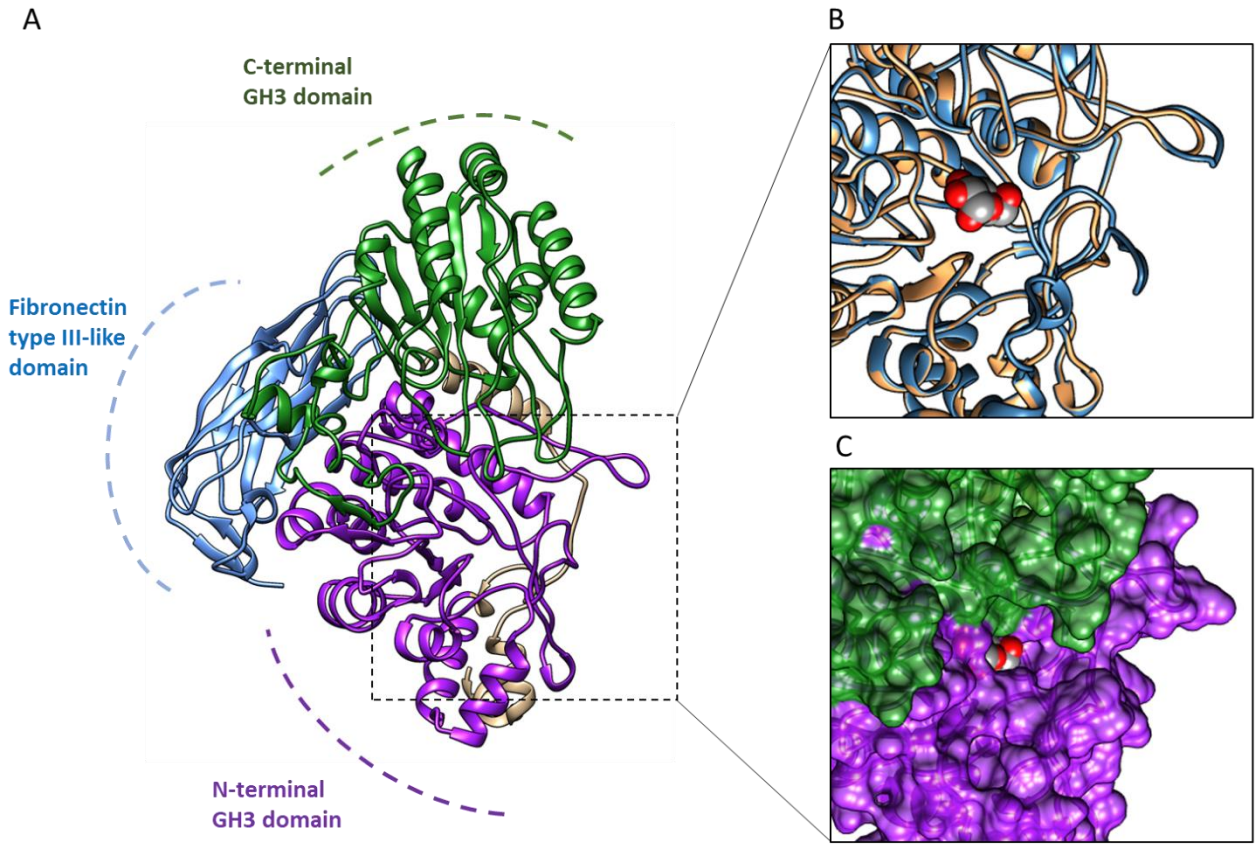


Figure 3: Predicted structure model of the β-glucosidase AMC3811_001245. *The model was constructed with Swissmodel using the crystal structure of the β-glucosidase 3B from Thermotoga neapolitana (PDB identifier 2x42).* ***(A)*** *Domain structure of the AMC3811_001245 model. The N-terminal GH3 family domain is coloured in purple, the C-terminal GH3 family domain is coloured in green and the fibronectin type III-like domain is coloured in blue. The putative substrate binding pocket is marked with a dashed square.* ***(B)*** *structural alignment of the constructed model of AMC3811_001245 and the crystal structure 2x42, showing the location of a bound alpha-D-glucopyranose within the 2x42 structure. 2x42 is depicted in blue, whereas the model structure of AMC3811_001245 is depicted in beige. The ligand is displayed in spherical mode (carbon atoms in grey, oxygen atoms in red).* ***(C)*** Visualization of the binding pocket in transparent surface mode. The *alpha-D-glucopyranose (position obtained from the structural alignment with 2x42) is located within a cleft between the N-terminal GH3 domain (purple) and the C-terminal GH3 domain (green).*

Thermostability, temperature and pH optimum of AMC3811_001245

The influence of pH, and temperature on the β-glucosidase activity of AMC3811_001245 was measured by determining the hydrolysis of pNPG at different temperatures and pH values, revealing the highest activity at pH 5, although the enzyme showed considerable activity over a broad pH range from 3-9 (Fig. 4A). As reported for other glycosidases from thermophiles, the β-glucosidase activity of AMC3811_001245 is significantly enhanced at higher temperatures with its highest measured activity at 100 °C. At temperatures lower than 50 °C, no decent activity was observed (Figure 4C). Consequently, high thermostability upon incubation of the purified enzyme at 60 °C and 80 °C was determined, resulting in only a slight decrease of glucosidase activity to 67.4% after incubation for 4h at 60 °C or 64.7% at 80 °C. Even after long time incubation over 24h, residual activity of 56.7% (60 °C) or 31.5% (80 °C) was determined. Upon incubation at 100 °C, less protein stability was oberseved, resulting in a significant decrease down to 12.2% residual activity after 2h (Figure 4C).

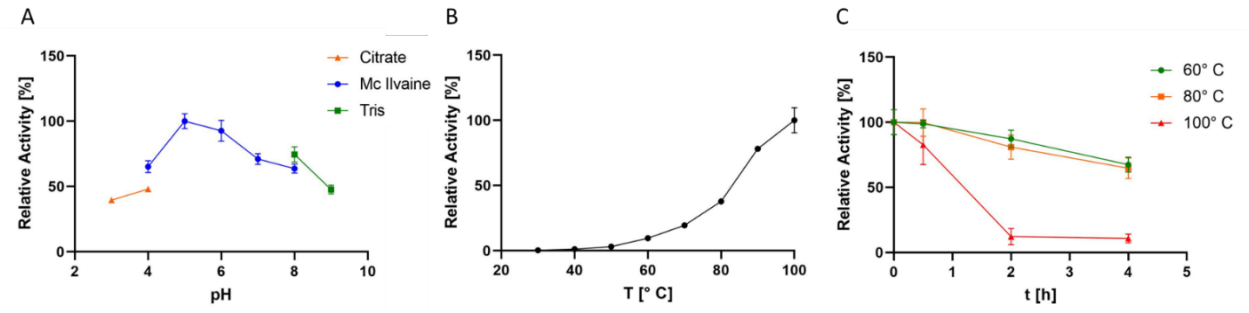


Figure 4: Characterization of the β-glucosidase AMC3811_001245. *The Influence of pH* ***(A)****, reaction temperature* ***(B)*** *and long time incubation at high temperatures* ***(C)*** *on the β-glucosidase activity of AMC3811_001245 are shown.*

Substrate specifity, kinetic characterization and substrate inhibition of AMC3811_001245

Glycosidase activity of the heterologously expressed β-glucosidase AMC3811_001245 was tested by quantification of glucose production via a discontinous hexokinase/G6PDH assay with a variaty of substrates, inlcuding crystalline cellulose (Avicel®), the modified β-1,4-linked glucose polymer carboxymethyl cellulose (CMC), curdlan (β-1,3-glucan), starch (α-1,4/6-glucan), the glucose disaccharides cellobiose (β-1,4), gentiobiose (β-1,6), laminaribiose (β-1,3), maltotriose (α-1,4) and trehalose (α-1,1). Moreover, the chromogenic substrates pNPG, pNPX and oNPGal were tested in a continous assay. From the natural substrates, highest activity was observed with the β-linked glucose disaccharides. Less, yet significant activity was determined using the soluble substrate CMC. No significant activity was measured for any α-linked substrate or for curdlan and Avicel® cellulose (Figure 5A). Exo-1,4-β-glucosidase activity was determined by following the D-glucose liberation with 0.5% (w/v) CMC (Figure 5B), giving a specific activity of 4.27 ± 1.42 U mg^{-1}.

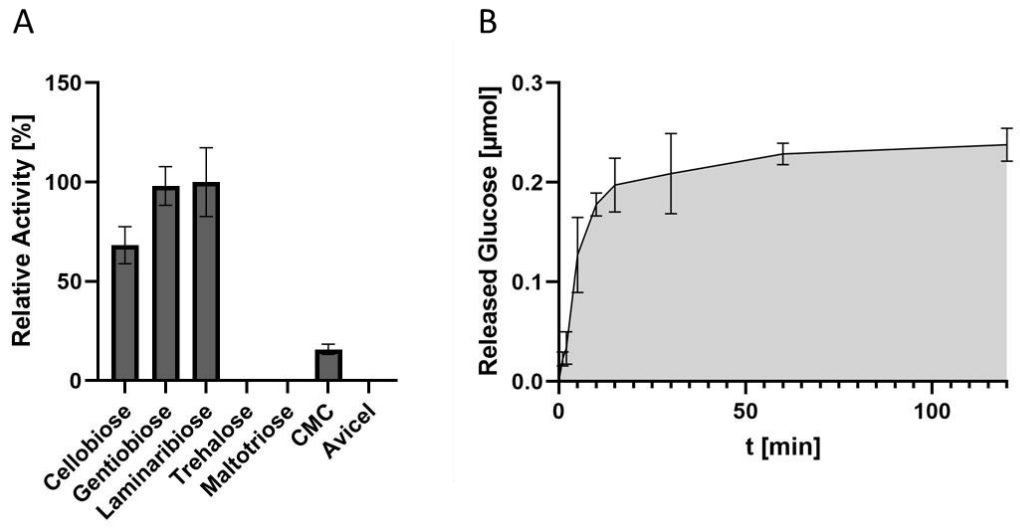


Figure 5: Substrate specifity of AMC3811_001245. (A) *Relative activity on different substrates (0.5% w/v) was determined at 100° C with 8 µg/ml AMC3811_001245 in McIlvaine buffer pH 5 using a discontinous hexokinase-G6PDH assay as described in the methods section. The activity with laminaribiose was set to 100%.* ***(B)*** *Exo-1,4-β-glucosidase activity was determined by glucose release from CMC during incubation at pH 5 and 100 °C with AMC3811_001245 over two hours.*

Under optimal conditions (pH 5 and 100 °C) classical Michaelis-Menten kinetics were determined for AMC3811_001245 with selected substrates, showing highest activity and affinity with pNPG (V_{max}= 29.89 U mg^{-1}, K_m= 0.12 mM), less activity with pNPX (V_{max}= 5.69 U mg^{-1}, K_m= 0.52 mM and no activity with oNPGal. Hydrolysis of cellobiose was observed at a V_{max} similar to the hydrolysis of pNPG (V_{max} = 31.17 U mg^{-1}), yet with a significantly higher K_m of 1.66 mM (Figure 6A). Inhibition of β-glucosidase activity with pNPG was observed with increasing concentrations of glucose, thus resulting in an activity reduction to roughly 50% when 5 mM glucose were added to the reaction. At 50 mM glucose, 12.8% residual activity was determined (Figure 6B).

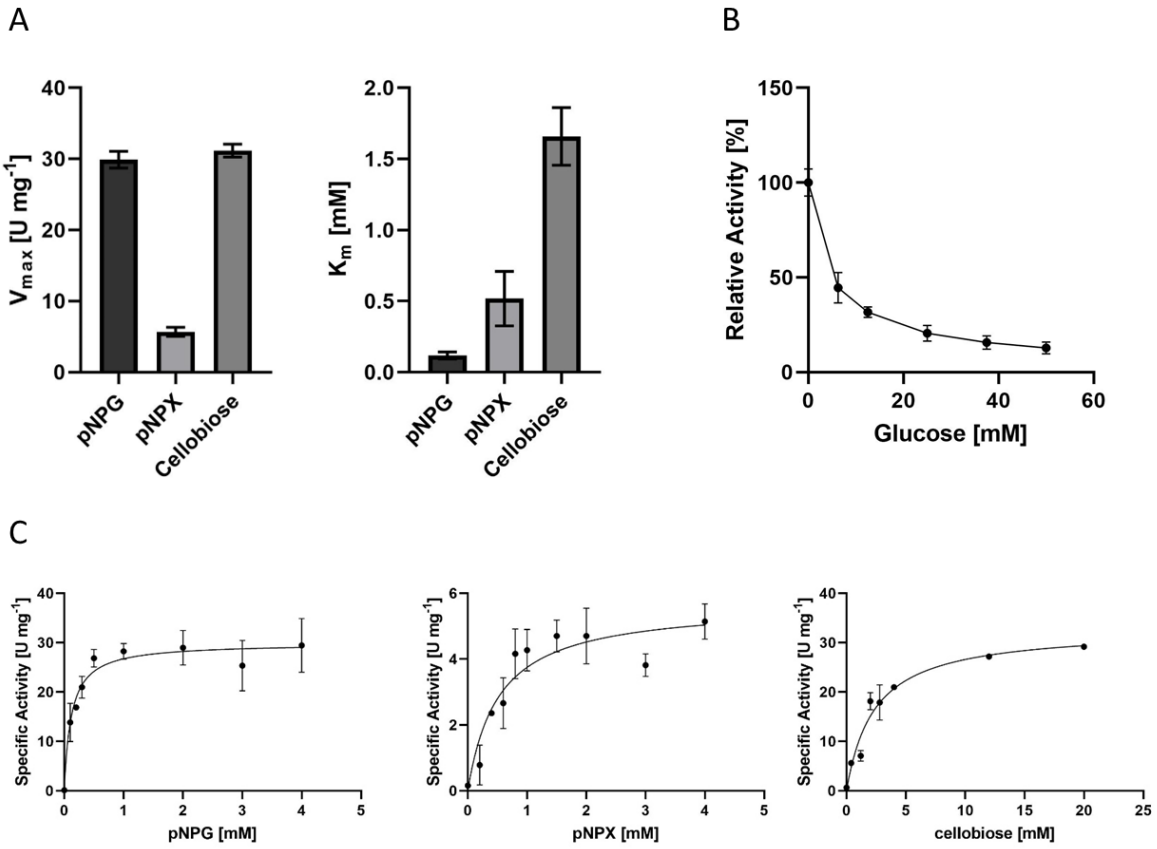


Figure 6: Kinetic characterization of the recombinant β-glycosidase AMC3811_001245 using the substrates pNPG, pNPX and cellobiose. (A) *Vmax and Km values determined for pNPG, pNPX and cellobiose. The kinetics for the pNP substrates were determined photometrically in a continous assay, whereas the activity on cellobiose was done by glucose determination via a discontinous discontinous hexokinase-G6PDH assay as described in the methods section.* ***(C)*** *Michaelis-Menten kinetics were determined for the chromogenig substrates substrates pNPG, pNPX and for cellobiose.*

Labelling and Activity Inhibition with the ABP JJB111

Since data on ABPP at high temperatures are scarce, the applicability of this approach with the β- glucosidase ABP JJB111 was investigated at different temperatures up to 100° C in *in vitro* studies (*in situ* ABPP was performed at 65-72° C). Using the established *in situ* ABPP workflow (as described above), labelling and subsequent visualization via immunoblot of the heterologously expressed and purified protein was succesfull at temperatures up to 80 °C. However, at higher temperatures, no distinct bands could be observed in the immunoblot, indicating loss of activity or even precipitation of proteins or hydroylsis of the probe upon 1 h incubation at these temperatures. Also, the band intensity at 30 °C was distinctly less and is increasing with increasing temperature (Figure 7A). Succesfull covalent binding of the ABP to the catalytic site of AMC3811_001245 consequently leads to inhibition of β-glucosidase activity, as depicted in figure 7B. Only 6% residual activity was determined upon pre-incubation of the protein (3 µM) with 8 µM JJB111.

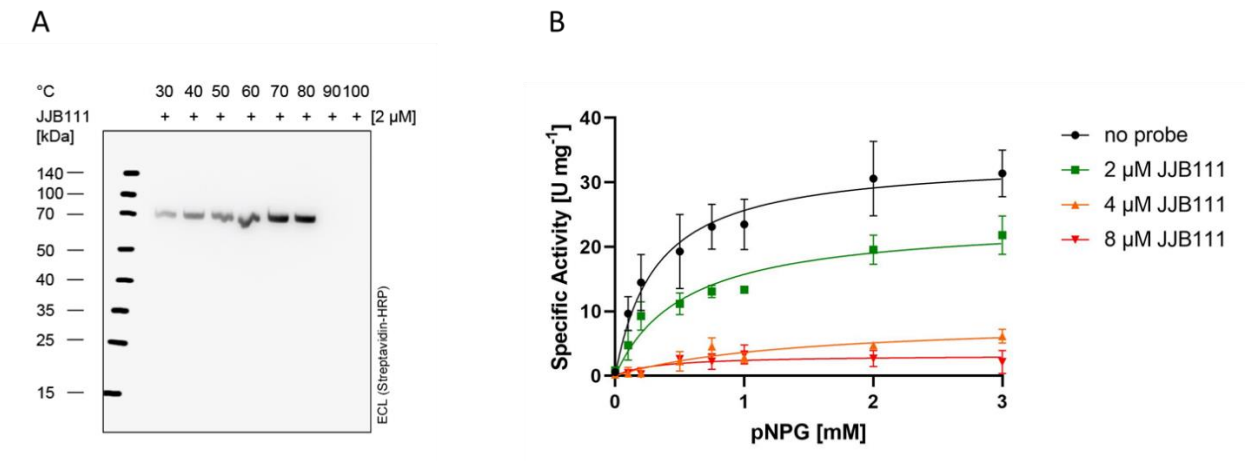


Figure 7: In vitro labelling of AMC3811_001245 with the ABP JJB111. (A) *Immunoblot of purified AMC3811_001245 which was incubated with 2 µM JJB111 for 1h at pH 5 and temperatures between 30-100 °C.* ***(B)*** *Kinetic characterization of the β-glucosidase activity of AMC3811_001245 after preincubation with different concentrations of the ABP JJB111 for 5 min.*

Influence of solvents and different ions

Application of glycosidases in industrial processes oftenly involves reactions at elevated temperatures but moreover, compatibility with other harsh conditions, such as the occurrence of different organic solvents or metal cations at high concentrations, is required. Therefore, the influence of different concentrations of organic solvents and different cations was examined for AMC3811_01245. As already reported for other enzymes [73,74] smaller amounts of ethanol, methanol and DMSO enhances the β-glucosidase activity of AMC3811_01245. Thus, preincubation at 100° C for 10 min with 20% (v/v) ethanol led to an increase of activity on pNPG to 151%. Moreover, the β-glucosidase activity was only slightly inhibited by small concentrations of a broad range of other organic solvents, such as acetone, acetonitrile, 2- butanol, chloroform, ethylacetate, 2-propanol and tetrahydrofuran. Even at concentrations up to 30% (v/v) of the applied solvents, considerable residual β-glucosidas activity was determined, except for 30% (v/v) tetrahydrofuran (Figure 8A). Glycosidases usually don't neccesarrily need metal ions as cofactors, although there are some exceptions, mainly from the family GH4, whose members also need NAD^+ as cofactor for proper function [75,76]. Consequently, pre-incubation with 1 mM EDTA did not have a significant effect on the β- glucosidase activity of AMC3811_01245, similarly as with 1 mM Ba^{2+} and Mg^{2+}. At the same time, addition of the same concentrations of Ca^{2+}, Co^{2+} or Mn^{2+} could significantly enhance the activity, whereas Cu^{2+}, Zn^{2+}, Al^{3+} and Fe^{3+} mostly or totally inhibits the β-glucosidase activity (Figure 8B).

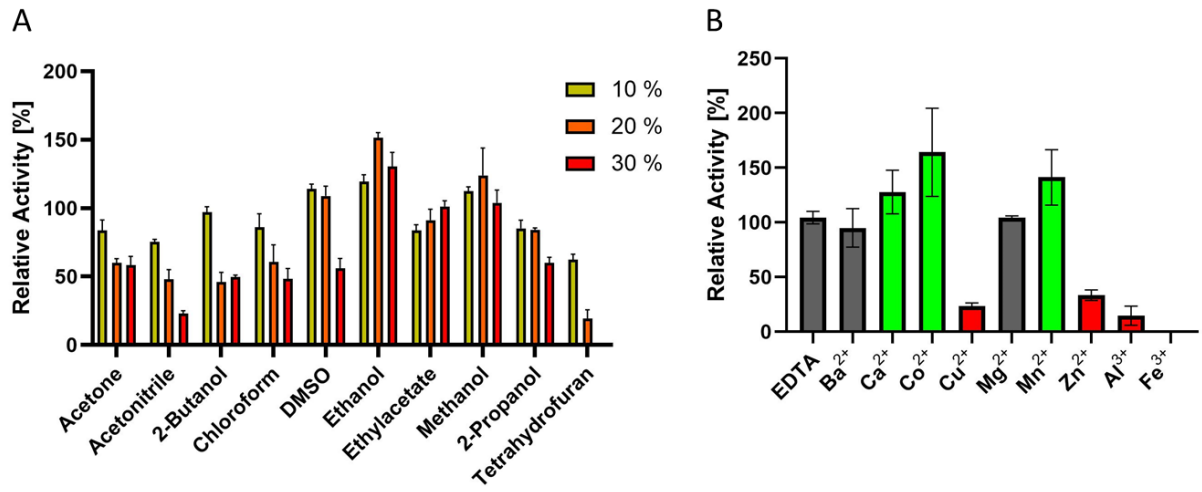


Figure 8: Influence of organic solvents and cations on the β-glucosidase acitivity of AMC3811_01245. (A) *Residual β- glucosidase activity at 100° C after 10 min preincubation with 10-30% (v/v) solvent compared to protein sample without added solvent.* ***(B)*** *Residual glucosidase activity at 100° C after 10 min preincubation with 1 mM of EDTA or different cations.*

Discussion

Due to its high abundance and availability, (ligno)cellulosic biomass represents a bulk renewable substrate for different industrial green chemistry processes, especially for the efficient production of biofuels and various platform chemicals [77,78]. Thus, the identification and characterization of enzymes and microbial strains capable of hydrolyzing cellulose polymers and its degradation products is crucial, especially when they feature high stability against high temperatures and other factors which could decrease the biocatalytic efficiency within the desired processes. For the identification of novel enzymes and strains with desired features, the use of enrichment cultures in combination with -omic approaches have been proven as a suitable strategy [21,79–81]. The Uzon caldera with its many hydrothermal systems and highly diverse microbial and geochemical characteristics has been intensively studied with different perspectives [7,27,85,86]. Past studies with focus on the biodiversity of thermophilic prokaryotes with hydrolytic activities on different polymers identified cultivated and uncultivated bacteria, belonging to the genera *Fervidobacterium*, *Dictyoglomus* and *Caldicellulosiruptor as well as* several uncultivated crenarchaeal species putatively belong to the *Thermoprotei* [18]. Consistent to previous studies on the microbiology of the Arkashin shurf and other thermophilic terrestrial systems, chemoautotrophic species of the phylum *Aquificae* are the predominant bacterial species in the AMC enrichment analyzed in this work [26] (Figure 1) which is most probably a representative of the basic community and not directly linked to AMC degradation.

In order to shed light on the hydrolytic activity and the contribution of individual phyla within this in situ enrichment culture, a thorough analysis of present CAZyme genes was done, highlighting the thermoacidophilic crenarchaeon *C. maquilignensis* as the single mayor contributor of glycoside hydrolase genes within the enrichment metagenome (Figure 2). *C. maquilignensis* was described in 1999 as a new genus of rod-shaped crenarchaeota from a hot spring in the Philippines and its full genome was published and up to date, *C. maquilingensis* IC- 167 is the only classified, validly published *C. maquilingensis* species, whereas 9 unclassified *Caldivirga* species are reported in the NCBI database [19,87]. Growth was reported at temperatures between 60- 92°C and pH values between 2.3 - 6.4 [19], whereas the utilization of polymeric carbohydrates such as cellulose or xylan has not been demonstrated, although 42 different glycosyl hydrolase entries are deposited for *C. maquilingensis* IC- 167 in the CAZY-database (cazy.org) which is the highest number among all Thermoproteales members with publically available genome sequences [88]. Despite the scarce overall characterization of *Caldivirga* species, some studies on its thermostable glycosdiases have been conducted [89–91]. Based on the high variety of glycosidase genes, especially from the cellulase containing family GH5 (32 identified hmm cluster, supplementary file 2) and due to the fact, that most glycosidases which could be succesfully labelled and quantified with the ABPP analysis could be assigned to *C. maquilignensis* species (Table 1), an important role of *C. maquilignensis* for the breakdown of cellulose within the enrichment culture described in this work was proposed. This is especially remarkable since the relative abundance of *Caldivirga* species in the enrichement culture was low (<1%).

Among other thermostable glycoside hydrolases quantified with the β-glucosidase ABP JJB111, the GH3 family β-glucosidase AMC3811_01245, originating from the Crenarchaeon *C. maquilingensis* was the most confident β-glucosidase hit, thus demonstrating activity *in situ*. Indeed, biochemical characterization of the heterologously expressed glucosidase demonstrated, that AMC3811_01245 is distinctly active under conditions, which resemble those of the spring where it was detected and moreover shows longtime stability for more than 4 hours (Figure 4). In addition to its ability to withstand high temperatures, its robustness against several commonly used solvents (Figure 8A) and divalent cations (Figure 8B) makes AMC3811_01245 a promising biocatalyst with decent β-glucosidase and exo-glucanase activity for the industrial processing of glucose-based poly- and disaccharides e.g. from pulp and plant waste material. It shoud be noted that among all GH3 entries in the CAZy database, archaeal representatives make up for only about 0.6% of total diversity including two characterized enzymes (according to Uniprot, both were not characterized biochemically,

though) and no enzymes with experimentally resolved tertiary structures. Thus, after ABPP has recently succesfully been applied for the detection of active serine hydrolases in hot environments (Ninck et al., 2022), we now enhanced the scoop of ABPP for the identification of active glycosidases in enrichment cultures directly in hot environments.

Data Availability statement

The mass spectrometry proteomics data for the digestions have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al. 2016) partner repository (https://www.ebi.ac.uk/pride/archive/) with the dataset identifier PXD025833. During the review process the data can be accessed via a reviewer account (Username: reviewer_pxd025833@ebi.ac.uk; Password: TRSLWcQX). The other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

BS, MK and IK acknowledge funding within the DFG-RSF Cooperation for joint German-Russian projects by the DFG (SI 642/12-1, KA 2894/6-1) and RSF (18-44-04024). Furtherly, this work was supported by the DFG (INST 20876/322-1 FUGG, to M. K. and F.K.).

Author contributions

T.K. and T.V.K. performed in-field experiments, such as setting up an enrichment culture and in situ labelling experiments, S.N., T.K. and L.S. performed the protein extraction and affinity enrichment experiments. DNA extraction and metagenome sequencing was done by T.V.K. and I. K., refinement of metagenome data, metaproteome assembly and creation of a CAZyme profile was done by S.E and A.P., S.N. and F.K. performed MS analyses, T.K. performed protein expression, biochemical characterization and moddeling, C.B, B.S. and M.K. supervised the study and T.K. wrote the manuscript.

References

1. Brock, T. D. Life at high temperatures. *Science (New York, N.Y.)* **230,** 132–138; 10.1126/science.230.4722.132 (1985).
2. Fiala, G. & Stetter, K. O. Pyrococcus furiosus sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100C. *Arch. Microbiol.* **145,** 56–61; 10.1007/BF00413027 (1986).
3. MARSH, C. L. & LARSEN, D. H. Characterization of some thermophilic bacteria from the Hot Springs of Yellowstone National Park. *Journal of bacteriology* **65,** 193–197; 10.1128/jb.65.2.193-197.1953 (1953).
4. Allen, E. E. & Banfield, J. F. Community genomics in microbial ecology and evolution. *Nature reviews. Microbiology* **3,** 489–498; 10.1038/nrmicro1157 (2005).
5. Li, S.-J. *et al.* Microbial communities evolve faster in extreme environments. *Scientific reports* **4,** 6205; 10.1038/srep06205 (2014).
6. Cioni, R., Fanelli, G., Guidi, M., Kinyariro, J. K. & Marini, L. Lake Bogoria hot springs (Kenya): geochemical features and geothermal implications. *Journal of Volcanology and Geothermal Research* **50,** 231–246; 10.1016/0377-0273(92)90095-U (1992).
7. Chernyh, N. A. *et al.* Microbial life in Bourlyashchy, the hottest thermal pool of Uzon Caldera, Kamchatka. *Extremophiles : life under extreme conditions* **19,** 1157–1171; 10.1007/s00792-015-0787-5 (2015).
8. Power, J. F. *et al.* Microbial biogeography of 925 geothermal springs in New Zealand. *Nature communications* **9,** 2876; 10.1038/s41467-018-05020-y (2018).
9. Lasa, I. & Berenguer, J. Thermophilic enzymes and their biotechnological potential. *Microbiologia (Madrid, Spain)* **9,** 77–89 (1993).
10. Galbe, M. & Wallberg, O. Pretreatment for biorefineries: a review of common methods for efficient utilisation of lignocellulosic materials. *Biotechnology for biofuels* **12,** 294; 10.1186/s13068-019-1634-1 (2019).
11. Littlechild, J. A. Enzymes from Extreme Environments and Their Industrial Applications. *Frontiers in bioengineering and biotechnology* **3,** 161; 10.3389/fbioe.2015.00161 (2015).
12. Cokgor, E. U., Oktay, S., Tas, D. O., Zengin, G. E. & Orhon, D. Influence of pH and temperature on soluble substrate generation with primary sludge fermentation. *Bioresource technology* **100,** 380–386; 10.1016/j.biortech.2008.05.025 (2009).
13. Schocke, L., Bräsen, C. & Siebers, B. Thermoacidophilic Sulfolobus species as source for extremozymes and as novel archaeal platform organisms. *Current Opinion in Biotechnology* **59,** 71–77; 10.1016/j.copbio.2019.02.012 (2019).
14. Mamo, G., Faryar, R. & Karlsson, E. N. Microbial Glycoside Hydrolases for Biomass Utilization in Biofuels Applications. In *Biofuel Technologies,* edited by V. K. Gupta & M. G. Tuohy (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), pp. 171–188.
15. Gupta, V. K. & Tuohy, M. G. (eds.). *Biofuel Technologies* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

16. Duchesne, L. C. & Larson, D. W. Cellulose and the Evolution of Plant Life. *BioScience* **39,** 238–241; 10.2307/1311160 (1989).

17. Blumer-Schuette, S. E., Kataeva, I., Westpheling, J., Adams, M. W. & Kelly, R. M. Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Current Opinion in Biotechnology* **19,** 210–217; 10.1016/j.copbio.2008.04.007 (2008).

18. Kublanov, I. V. *et al.* Biodiversity of thermophilic prokaryotes with hydrolytic activities in hot springs of Uzon Caldera, Kamchatka (Russia). *Applied and environmental microbiology* **75,** 286–291; 10.1128/AEM.00607-08 (2009).

19. Itoh, T., Suzuki, K., Sanchez, P. C. & Nakase, T. Caldivirga maquilingensis gen. nov., sp. nov., a new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. *International journal of systematic bacteriology* **49 Pt 3,** 1157–1163; 10.1099/00207713-49-3-1157 (1999).

20. Mead, D. A. *et al.* Complete Genome Sequence of Paenibacillus strain Y4.12MC10, a Novel Paenibacillus lautus strain Isolated from Obsidian Hot Spring in Yellowstone National Park. *Standards in genomic sciences* **6,** 381–400; 10.4056/sigs.2605792 (2012).

21. Gavrilov, S. N. *et al.* Isolation and Characterization of the First Xylanolytic Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1 and Its Unusual Multidomain Glycosidase. *Frontiers in microbiology* **7,** 552; 10.3389/fmicb.2016.00552 (2016).

22. Korzhenkov, A. A., Toshchakov, S. V., Podosokorskaya, O. A., Patrushev, M. V. & Kublanov, I. V. Data on draft genome sequence of Caldanaerobacter sp. strain 1523vc, a thermophilic bacterium, isolated from a hot spring of Uzon Caldera, (Kamchatka, Russia). *Data in brief* **33,** 106336; 10.1016/j.dib.2020.106336 (2020).

23. Zayulina, K. S. *et al.* Novel Hyperthermophilic Crenarchaeon Thermofilum adornatum sp. nov. Uses GH1, GH3, and Two Novel Glycosidases for Cellulose Hydrolysis. *Frontiers in microbiology* **10,** 2972; 10.3389/fmicb.2019.02972 (2019).

24. Zayulina, K. S., Elcheninov, A. G., Toshchakov, S. V. & Kublanov, I. V. Complete Genome Sequence of a Hyperthermophilic Archaeon, Thermosphaera sp. Strain 3507, Isolated from a Chilean Hot Spring. *Microbiology resource announcements* **9**; 10.1128/MRA.01262-20 (2020).

25. Reichart, N. J., Bowers, R. M., Woyke, T. & Hatzenpichler, R. High Potential for Biomass-Degrading Enzymes Revealed by Hot Spring Metagenomics. *Frontiers in microbiology* **12,** 668238; 10.3389/fmicb.2021.668238 (2021).

26. Burgess, E. A., Unrine, J. M., Mills, G. L., Romanek, C. S. & Wiegel, J. Comparative geochemical and microbiological characterization of two thermal pools in the Uzon Caldera, Kamchatka, Russia. *Microbial ecology* **63,** 471–489; 10.1007/s00248-011-9979-4 (2012).

27. Mardanov, A. V., Gumerov, V. M., Beletsky, A. V. & Ravin, N. V. Microbial diversity in acidic thermal pools in the Uzon Caldera, Kamchatka. *Antonie van Leeuwenhoek* **111,** 35–43; 10.1007/s10482-017-0924-5 (2018).

28. Zarafeta, D. *et al.* Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. *Scientific reports* **6,** 38886; 10.1038/srep38886 (2016).

29. Merkel, A. Y. *et al.* Microbial diversity and autotrophic activity in Kamchatka hot springs. *Extremophiles : life under extreme conditions* **21,** 307–317; 10.1007/s00792-016-0903-1 (2017).

30. Kochetkova, T. V., Podosokorskaya, O. A., Elcheninov, A. G. & Kublanov, I. V. Diversity of Thermophilic Prokaryotes Inhabiting Russian Natural Hot Springs. *Microbiology* **91,** 1–27; 10.1134/S0026261722010064 (2022).

31. Kaushal, G., Kumar, J., Sangwan, R. S. & Singh, S. P. Metagenomic analysis of geothermal water reservoir sites exploring carbohydrate-related thermozymes. *International journal of biological macromolecules* **119,** 882–895; 10.1016/j.ijbiomac.2018.07.196 (2018).

32. Prokofeva, M. I. *et al.* Cultivated anaerobic acidophilic/acidotolerant thermophiles from terrestrial and deep-sea hydrothermal habitats. *Extremophiles : life under extreme conditions* **9,** 437–448; 10.1007/s00792-005-0461-4 (2005).

33. Schröder, C., Elleuche, S., Blank, S. & Antranikian, G. Characterization of a heat-active archaeal β-glucosidase from a hydrothermal spring metagenome. *Enzyme and microbial technology* **57,** 48–54; 10.1016/j.enzmictec.2014.01.010 (2014).

34. Gupta, R., Govil, T., Capalash, N. & Sharma, P. Characterization of a glycoside hydrolase family 1 β-galactosidase from hot spring metagenome with transglycosylation activity. *Applied biochemistry and biotechnology* **168,** 1681–1693; 10.1007/s12010-012-9889-z (2012).

35. Nodwell, M. B. & Sieber, S. A. ABPP methodology: introduction and overview. *Topics in current chemistry* **324,** 1–41; 10.1007/128_2011_302 (2012).

36. Chen, X. *et al.* Target identification with quantitative activity based protein profiling (ABPP). *Proteomics* **17**; 10.1002/pmic.201600212 (2017).

37. Keller, L. J., Babin, B. M., Lakemeyer, M. & Bogyo, M. Activity-based protein profiling in bacteria: Applications for identification of therapeutic targets and characterization of microbial communities. *Current opinion in chemical biology* **54,** 45–53; 10.1016/j.cbpa.2019.10.007 (2020).

38. Adam, G. C., Cravatt, B. F. & Sorensen*, E. J. Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chemistry & Biology* **8,** 81–95; 10.1016/S1074-5521(00)90060-7 (2001).

39. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M. & Cravatt, B. F. Activity-based probes for the proteomic profiling of metalloproteases. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 10000–10005; 10.1073/pnas.0402784101 (2004).

40. Wu, L. *et al.* An overview of activity-based probes for glycosidases. *Current opinion in chemical biology* **53,** 25–36; 10.1016/j.cbpa.2019.05.030 (2019).

41. Fonović, M. & Bogyo, M. Activity-based probes as a tool for functional proteomic analysis of proteases. *Expert review of proteomics* **5,** 721–730; 10.1586/14789450.5.5.721 (2008).

42. Sadler, N. C. & Wright, A. T. Activity-based protein profiling of microbes. *Current opinion in chemical biology* **24,** 139–144; 10.1016/j.cbpa.2014.10.022 (2015).

43. Husaini, A. M. *et al.* Multiplex Fluorescent, Activity-Based Protein Profiling Identifies Active α-Glycosidases and Other Hydrolases in Plants. *Plant physiology* **177,** 24–37; 10.1104/pp.18.00250 (2018).

44. Whidbey, C. & Wright, A. T. Activity-Based Protein Profiling-Enabling Multimodal Functional Studies of Microbial Communities. *Current topics in microbiology and immunology* **420,** 1–21; 10.1007/82_2018_128 (2019).

45. Benns, H. J., Wincott, C. J., Tate, E. W. & Child, M. A. Activity- and reactivity-based proteomics: Recent technological advances and applications in drug discovery. *Current opinion in chemical biology* **60,** 20–29; 10.1016/j.cbpa.2020.06.011 (2021).

46. Zweerink, S. *et al.* Activity-based protein profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nature communications* **8,** 15352; 10.1038/ncomms15352 (2017).

47. Klaus, T. *et al.* Activity-Based Protein Profiling for the Identification of Novel Carbohydrate-Active Enzymes Involved in Xylan Degradation in the Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1E. *Frontiers in microbiology* **12**; 10.3389/fmicb.2021.734039 (2022).

48. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome research* **27,** 824–834; 10.1101/gr.213959.116 (2017).

49. Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2,** 26; 10.1186/2049-2618-2-26 (2014).

50. Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures. *Genome biology* **10,** R85; 10.1186/gb-2009-10-8-r85 (2009).

51. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523,** 208–211; 10.1038/nature14486 (2015).

52. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature microbiology* **3,** 836–843; 10.1038/s41564-018-0171-1 (2018).

53. Bornemann, T. L., Esser, S. P., Stach, T. L., Burg, T. & Probst, A. J. *uBin – a manual refining tool for metagenomic bins designed for educational purposes* (2020).

54. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25,** 1043–1055; 10.1101/gr.186072.114 (2015).

55. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11,** 119; 10.1186/1471-2105-11-119 (2010).

56. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215,** 403–410; 10.1016/S0022-2836(05)80360-2 (1990).

57. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics (Oxford, England)* **23,** 1282–1288; 10.1093/bioinformatics/btm098 (2007).

58. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics (Oxford, England)*; 10.1093/bioinformatics/btz848 (2019).

59. Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature biotechnology* **38,** 1079–1086; 10.1038/s41587-020-0501-8 (2020).

60. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS computational biology* **7,** e1002195; 10.1371/journal.pcbi.1002195 (2011).

61. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **46,** W95-W101; 10.1093/nar/gky418 (2018).

62. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic acids research* **32,** W20-5; 10.1093/nar/gkh435 (2004).

63. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39,** W29-37; 10.1093/nar/gkr367 (2011).

64. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology* **37,** 420–423; 10.1038/s41587-019-0036-z (2019).

65. Käll, L., Krogh, A. & Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic acids research* **35,** W429-32; 10.1093/nar/gkm256 (2007).

66. Krogh, A., Larsson, B., Heijne, G. von & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305,** 567–580; 10.1006/jmbi.2000.4315 (2001).

67. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46,** W296-W303; 10.1093/nar/gky427 (2018).

68. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25,** 1605–1612; 10.1002/jcc.20084 (2004).

69. McIlvaine, T. C. A BUFFER SOLUTION FOR COLORIMETRIC COMPARISON. *Journal of Biological Chemistry* **49,** 183–186; 10.1016/S0021-9258(18)86000-8 (1921).

70. Kallemeijn, W. W. *et al.* Novel activity-based probes for broad-spectrum profiling of retaining β-exoglucosidases in situ and in vivo. *Angewandte Chemie (International ed. in English)* **51,** 12529–12533; 10.1002/anie.201207771 (2012).

71. Little, E., Bork, P. & Doolittle, R. F. Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *Journal of molecular evolution* **39,** 631–643; 10.1007/BF00160409 (1994).

72. Kataeva, I. A. *et al.* The fibronectin type 3-like repeat from the Clostridium thermocellum cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *Applied and environmental microbiology* **68,** 4292–4300; 10.1128/AEM.68.9.4292-4300.2002 (2002).

73. D'Auria, S. *et al.* beta-glycosidase from the hyperthermophilic archaeon Sulfolobus solfataricus: structure and activity in the presence of alcohols. *Journal of biochemistry* **126,** 545–552; 10.1093/oxfordjournals.jbchem.a022484 (1999).

74. Faulds, C. B., Pérez-Boada, M. & Martínez, A. T. Influence of organic co-solvents on the activity and substrate specificity of feruloyl esterases. *Bioresource technology* **102,** 4962–4967; 10.1016/j.biortech.2011.01.088 (2011).

75. Varrot, A. *et al.* NAD+ and metal-ion dependent hydrolysis by family 4 glycosidases: structural insight into specificity for phospho-beta-D-glucosides. *Journal of molecular biology* **346,** 423–435; 10.1016/j.jmb.2004.11.058 (2005).

76. Nazmi, A. R., Reinisch, T. & Hinz, H.-J. Ca-binding to Bacillus licheniformis alpha-amylase (BLA). *Archives of biochemistry and biophysics* **453,** 18–25; 10.1016/j.abb.2006.04.004 (2006).

77. Srivastava, N. *et al.* Microbial Beta Glucosidase Enzymes: Recent Advances in Biomass Conversation for Biofuels Application. *Biomolecules* **9**; 10.3390/biom9060220 (2019).

78. Ge, X. *et al.* Conversion of Lignocellulosic Biomass Into Platform Chemicals for Biobased Polyurethane Application (Elsevier2018), Vol. 3, pp. 161–213.

79. Albaser, A. *et al.* Discovery of a Bacterial Glycoside Hydrolase Family 3 (GH3) β-Glucosidase with Myrosinase Activity from a Citrobacter Strain Isolated from Soil. *Journal of agricultural and food chemistry* **64,** 1520–1527; 10.1021/acs.jafc.5b05381 (2016).

80. Busch, P., Suleiman, M., Schäfers, C. & Antranikian, G. A multi-omic screening approach for the discovery of thermoactive glycoside hydrolases. *Extremophiles : life under extreme conditions* **25,** 101–114; 10.1007/s00792-020-01214-9 (2021).

81. Sakai, H. D. & Kurosawa, N. Exploration and isolation of novel thermophiles in frozen enrichment cultures derived from a terrestrial acidic hot spring. *Extremophiles : life under extreme conditions* **20,** 207–214; 10.1007/s00792-016-0815-0 (2016).

82. Reigstad, L. J., Jorgensen, S. L. & Schleper, C. Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka. *The ISME journal* **4,** 346–356; 10.1038/ismej.2009.126 (2010).

83. Wilkins, L. G. E., Ettinger, C. L., Jospin, G. & Eisen, J. A. Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Scientific reports* **9,** 3059; 10.1038/s41598-019-39576-6 (2019).

84. Frolov, E. N. *et al.* Diversity and Activity of Sulfate-Reducing Prokaryotes in Kamchatka Hot Springs. *Microorganisms* **9**; 10.3390/microorganisms9102072 (2021).

85. Dobretsov, N. L. *et al.* Geological, hydrogeochemical, and microbiological characteristics of the Oil site of the Uzon caldera (Kamchatka). *Russian Geology and Geophysics* **56,** 39–63; 10.1016/j.rgg.2015.01.003 (2015).

86. Rozanov, A. S. *et al.* Molecular analysis of the benthos microbial community in Zavarzin thermal spring (Uzon Caldera, Kamchatka, Russia). *BMC genomics* **15 Suppl 12,** S12; 10.1186/1471-2164-15-S12-S12 (2014).

87. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database: The Journal of Biological Databases and Curation* **2020**; 10.1093/database/baaa062 (2020).

88. Zayulina, K. S. *et al.* Novel hyperthermophilic crenarchaeon Infirmifilum lucidum gen. nov. sp. nov., reclassification of Thermofilum uzonense as Infirmifilum uzonense comb. nov. and assignment of the family Thermofilaceae to the order Thermofilales ord. nov. *Systematic and applied microbiology* **44,** 126230; 10.1016/j.syapm.2021.126230 (2021).

89. Letsididi, R. *et al.* Characterization of a thermostable glycoside hydrolase (CMbg0408) from the hyperthermophilic archaeon Caldivirga maquilingensis IC-167. *Journal of the science of food and agriculture* **97,** 2132–2140; 10.1002/jsfa.8019 (2017).

90. Letsididi, R. *et al.* Lactulose production by a thermostable glycoside hydrolase from the hyperthermophilic archaeon Caldivirga maquilingensis IC-167. *Journal of the science of food and agriculture* **98,** 928–937; 10.1002/jsfa.8539 (2018).

91. Letsididi, R., S. Letsididi, K., Zhang, T., Jiang, B. & Mu, W. Molecular Modelling of a Thermostable Glycoside Hydrolase from Caldivirga maquilingensis and Its Substrate Docking Mechanism for Galactooligosaccharides Synthesis. *BB* **6,** 1–7; 10.12691/bb-6-1-1 (2018).

4. Summary

Exploiting natures biodiversity, which is continously becoming more accesible is crucious for the detection of valuable biocatalysts for industrial and environmental applications. Elaborated (meta)genomic tools using Next Generation Sequencing methods, advanced analytical workflows for proteomic analyses and increasing bioinformatical and computational advances supplies researchers with vast numbers of sequence based biological information such as gen- and protein sequences. However, these information obtained from environmental samples and cultures grown in controlled laboratory conditions oftenly lack information about the activity state at the sampling site under the respective growth conditions and many hidden protein functions cannot be derived from sequence information alone. Thus the main focus of this thesis was the establishment of ABPP for the identification of thermophilic hydrolases in pure cultures, enrichment cultures as well as directly in the environment to close this gap. Glycosyl hydrolases, which were mainly targeted in the chapters 3.1 and 3.3 represent a family of enzymes, which is abundant in all living cells and serves for several vital cell functions, such as aquiring nutrition from complex polysaccharides as well as the synthesis of cell walls and secondary metabolites. Not suprisingly, these enzymes are highly demanded for industrial purposes, especially if they possess thermostable traits like the described glycosidases which where characterized in chapter 3.1. In this project, a recently discovered *Thermococcus* strain with unusual carbohydrate degrading abilities was described and analyzed with respect to its CAZymes and its growth capabilities on complex carbohydrates, such as xylan, with a novel combined ABPP approach. Noteworthy, in this work, the sucessfull application of glycoside hydrolase specific ABPs in a hyperthermophilic host was described for the first time. In chapter 3.2. the application scope of this approach was broadened to more complex and heterogenous environmental samples from thermophilic springs. An elaborated protocol, combining advanced metagenomic and metaproteomic analyses with well established fluorophosphonated probes for the specific labelling of active serine hydrolases was sucessfully applied, prooving that this concept can also be applied for bioprospecting of thermoactive hydrolases directly in the environment without the need of cultivating desired microbes. In chapter 3.3 the potential of thermophilic β-glucosidases of a hot spring was addressed with combining the before mentioned techniques, such as metagenomic sequencing, metaproteomic analysis and ABPP with a glucosidase specific ABP on a enrichment culture with amorphous cellulose as substrate. The cells which were enriched in this culture, were succesfully labelled with the probe and the best targeted glucosidase was finally heterologously expressed, purified and characterized as a

thermostable and highly active β-glucosidase. In summary, this work demonstrates that ABPP in combination with other microbial, bioinformatic and biochemical methods can serve as a valuable tool for the identification and characterization of novel biocatalysts in pure cell culture, enrichment cultures or directly in the environment even under extreme conditions.

5. Zusammenfassung

Die Nutzung der biologischen Vielfalt der Natur, die immer leichter zugänglich wird, ist von entscheidender Bedeutung für die Entdeckung wertvoller Biokatalysatoren für industrielle und umwelttechnische Anwendungen. Ausgefeilte (meta)genomische Werkzeuge unter Verwendung von Next Generation Sequencing-Methoden, fortgeschrittene analytische Arbeitsabläufe für proteomische Analysen und zunehmende bioinformatische und technische Fortschritte versorgen die Forscher mit einer großen Anzahl sequenzbasierter biologischer Informationen wie Gen- und Proteinsequenzen. Diesen Informationen, die aus Umweltproben und unter kontrollierten Laborbedingungen gezüchteten Kulturen gewonnen werden, mangelt es jedoch häufig an Informationen über den Aktivitätszustand an der Entnahmestelle unter den jeweiligen Wachstumsbedingungen, und viele verborgene Proteinfunktionen lassen sich nicht allein aus Sequenzinformationen ableiten. Daher lag der Schwerpunkt dieser Arbeit auf der Etablierung von ABPP zur Identifizierung von thermophilen Hydrolasen in Reinkulturen, Anreicherungskulturen sowie direkt in der Umwelt, um diese Lücke zu schließen. Glykosylhydrolasen, die in den Kapiteln 3.1 und 3.3 im Mittelpunkt standen, stellen eine Enzymfamilie dar, die in allen lebenden Zellen reichlich vorhanden ist und mehreren lebenswichtigen Zellfunktionen dient, wie z. B. der Gewinnung von Nährstoffen aus komplexen Polysacchariden sowie der Synthese von Zellwänden und Sekundärmetaboliten. Es überrascht nicht, dass diese Enzyme für industrielle Zwecke sehr gefragt sind, insbesondere wenn sie thermostabile Eigenschaften besitzen, wie die in Kapitel 3.1 beschriebenen Glycosidasen. In diesem Projekt wurde ein kürzlich entdeckter *Thermococcus*-Stamm mit ungewöhnlichen Fähigkeiten zum Abbau von Kohlenhydraten beschrieben und im Hinblick auf seine CAZyme und seine Wachstumsfähigkeit auf komplexen Kohlenhydraten, wie beispielsweise Xylan, mit einem neuartigen kombinierten ABPP-Ansatz analysiert. Bemerkenswert ist, dass in dieser Arbeit zum ersten Mal die erfolgreiche Anwendung von Glykosidhydrolase-spezifischen ABPs in einem hyperthermophilen Wirt beschrieben wurde. In Kapitel 3.2. wurde der Anwendungsbereich dieses Ansatzes auf komplexere und heterogene Umweltproben aus thermophilen Quellen ausgeweitet. Ein ausgearbeitetes Protokoll, das

fortgeschrittene metagenomische und metaproteomische Analysen mit gut etablierten fluorophosphonierten Sonden für die spezifische Markierung aktiver Serinhydrolasen kombiniert, wurde erfolgreich angewandt und bewies, dass dieses Konzept auch für die Bioprospektion thermoaktiver Hydrolasen direkt in der Umwelt angewendet werden kann, ohne dass die gewünschten Mikroben kultiviert werden müssen. In Kapitel 3.3 wurde das Potenzial der thermophilen β-Glucosidasen einer heißen Quelle durch die Kombination der zuvor genannten Techniken, wie metagenomische Sequenzierung, metaproteomische Analyse und ABPP mit einem glucosidasespezifischen ABP auf einer Anreicherungskultur mit amorpher Zellulose als Substrat untersucht. Die Zellen, die in dieser Kultur angereichert wurden, wurden erfolgreich mit der Sonde markiert, und die beste zielgerichtete Glucosidase wurde schließlich heterolog exprimiert, gereinigt und als thermostabile und hochaktive β-Glucosidase charakterisiert. Zusammenfassend zeigt diese Arbeit, dass ABPP in Kombination mit anderen mikrobiellen, bioinformatischen und biochemischen Methoden als wertvolles Werkzeug für die Identifizierung und Charakterisierung neuartiger Biokatalysatoren in reinen Zellkulturen, Anreicherungskulturen oder direkt in der Umwelt selbst unter extremen Bedingungen dienen kann.

6. References

1. Brock, T. D. & Freeze, H. Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of bacteriology* **98,** 289–297; 10.1128/jb.98.1.289-297.1969 (1969).
2. Rampelotto, P. H. Extremophiles and extreme environments. *Life (Basel, Switzerland)* **3,** 482–485; 10.3390/life3030482 (2013).
3. Pikuta, E. V., Hoover, R. B. & Tang, J. Microbial extremophiles at the limits of life. *Critical reviews in microbiology* **33,** 183–209; 10.1080/10408410701451948 (2007).
4. Martin, W., Baross, J., Kelley, D. & Russell, M. J. Hydrothermal vents and the origin of life. *Nature reviews. Microbiology* **6,** 805–814; 10.1038/nrmicro1991 (2008).
5. Reed, C. J., Lewis, H., Trejo, E., Winston, V. & Evilia, C. Protein adaptations in archaeal extremophiles. *Archaea (Vancouver, B.C.)* **2013,** 373275; 10.1155/2013/373275 (2013).
6. Brininger, C., Spradlin, S., Cobani, L. & Evilia, C. The more adaptive to change, the more likely you are to survive: Protein adaptation in extremophiles. *Seminars in cell & developmental biology* **84,** 158–169; 10.1016/j.semcdb.2017.12.016 (2018).
7. Siliakus, M. F., van der Oost, J. & Kengen, S. W. M. Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles : life under extreme conditions* **21,** 651–670; 10.1007/s00792-017-0939-x (2017).
8. Pflüger, K. & Müller, V. Transport of compatible solutes in extremophiles. *Journal of bioenergetics and biomembranes* **36,** 17–24; 10.1023/b:jobb.0000019594.43450.c5 (2004).
9. Siebers, B. & Schönheit, P. Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Current Opinion in Microbiology* **8,** 695–705; 10.1016/j.mib.2005.10.014 (2005).
10. Tse, C. & Ma, K. Growth and Metabolism of Extremophilic Microorganisms. In *Biotechnology of Extremophiles*, edited by P. H. Rampelotto (Springer International Publishing, Cham, 2016), Vol. 1, pp. 1–46.
11. Krulwich, T. A., Sachs, G. & Padan, E. Molecular aspects of bacterial pH sensing and homeostasis. *Nature reviews. Microbiology* **9,** 330–343; 10.1038/nrmicro2549 (2011).
12. van den Burg, B. Extremophiles as a source for novel enzymes. *Current Opinion in Microbiology* **6,** 213–218; 10.1016/S1369-5274(03)00060-2 (2003).
13. Coker, J. A. Extremophiles and biotechnology: current uses and prospects. *F1000Research* **5**; 10.12688/f1000research.7432.1 (2016).
14. Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of bacteriology* **127,** 1550–1557; 10.1128/jb.127.3.1550-1557.1976 (1976).
15. Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science (New York, N.Y.)* **239,** 487–491; 10.1126/science.2448875 (1988).
16. Ferrera, I. & Reysenbach, A.-L. Thermophiles. In *eLS*, edited by L. John Wiley & Sons (Wiley 2001).
17. Maheshwari, R., Bharadwaj, G. & Bhat, M. K. Thermophilic fungi: their physiology and enzymes. *Microbiology and molecular biology reviews : MMBR* **64,** 461–488; 10.1128/MMBR.64.3.461-488.2000 (2000).

18. Saini, N., Pal, K., Sujata, Deepak, B. & Mona, S. Thermophilic algae: A new prospect towards environmental sustainability. *Journal of Cleaner Production* **324,** 129277; 10.1016/j.jclepro.2021.129277 (2021).

19. Tansey, M. R. & Brock, T. D. The upper temperature limit for eukaryotic organisms. *Proceedings of the National Academy of Sciences of the United States of America* **69,** 2426–2428; 10.1073/pnas.69.9.2426 (1972).

20. Archana S. Rao, Ajay Nair, Veena S. More, K.S Anantharaju, Sunil S. More: Chapter 11 - Extremophiles for sustainable agriculture, Editor(s): Harikesh Bahadur Singh, Anukool Vaishnav, *New and Future Developments in Microbial Biotechnology and Bioengineering, 243-264* (Elsevier, 2021).

21. Patel, H. & Rawat, S. Thermophilic fungi: Diversity, physiology, genetics, and applications. In *New and Future Developments in Microbial Biotechnology and Bioengineering* (Elsevier 2021), pp. 69–93.

22. Forterre, P. A Hot Topic: The Origin of Hyperthermophiles. *Cell* **85,** 789–792; 10.1016/S0092-8674(00)81262-3 (1996).

23. Huber, H. & Stetter, K. O. Hyperthermophiles and their possible potential in biotechnology. *Journal of Biotechnology* **64,** 39–52; 10.1016/S0168-1656(98)00102-3 (1998).

24. Stetter, K. O. Hyperthermophiles in the history of life. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **361,** 1837-42; discussion 1842-3; 10.1098/rstb.2006.1907 (2006).

25. Pace, N. R. Origin of life-facing up to the physical setting. *Cell* **65,** 531–533; 10.1016/0092-8674(91)90082-A (1991).

26. Blöchl, E. *et al.* Pyrolobus fumarii, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles : life under extreme conditions* **1,** 14–21; 10.1007/s007920050010 (1997).

27. Takai, K. *et al.* Cell proliferation at 122 degrees C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 10949–10954; 10.1073/pnas.0712334105 (2008).

28. Kashefi, K. & Lovley, D. R. Extending the upper temperature limit for life. *Science (New York, N.Y.)* **301,** 934; 10.1126/science.1086823 (2003).

29. Saiki, T., Kimura, R. & Arima, K. Isolation and Characterization of Extremely Thermophilic Bacteria from Hot Springs. *Agricultural and Biological Chemistry* **36,** 2357–2366; 10.1080/00021369.1972.10860589 (1972).

30. Segerer, A., Langworthy, T. A. & Stetter, K. O. Thermoplasma acidophilum and Thermoplasma volcanium sp. nov. from Solfatara Fields. *Systematic and Applied Microbiology* **10,** 161–171; 10.1016/S0723-2020(88)80031-6 (1988).

31. Aanniz, T. *et al.* Thermophilic bacteria in Moroccan hot springs, salt marshes and desert soils. *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]* **46,** 443–453; 10.1590/S1517-838246220140219 (2015).

32. Mladenovska & Ahring. Growth kinetics of thermophilic Methanosarcina spp. isolated from full-scale biogas plants treating animal manures. *FEMS microbiology ecology* **31,** 225–229; 10.1111/j.1574-6941.2000.tb00687.x (2000).

33. Kleeberg, I., Hetz, C., Kroppenstedt, R. M., Müller, R. J. & Deckwer, W. D. Biodegradation of aliphatic-aromatic copolyesters by Thermomonospora fusca and other thermophilic compost isolates. *Applied and environmental microbiology* **64,** 1731–1735; 10.1128/AEM.64.5.1731-1735.1998 (1998).

34. Miroshnichenko, M. L. & Bonch-Osmolovskaya, E. A. Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. *Extremophiles : life under extreme conditions* **10,** 85–96; 10.1007/s00792-005-0489-5 (2006).

35. Schmerling, C., Kouril, T., Snoep, J., Bräsen, C. & Siebers, B. Enhanced underground metabolism challenges life at high temperature–metabolic thermoadaptation in hyperthermophilic Archaea. *Current Opinion in Systems Biology* **30,** 100423; 10.1016/j.coisb.2022.100423 (2022).

36. Adam, P. S., Borrel, G., Brochier-Armanet, C. & Gribaldo, S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *The ISME journal* **11,** 2407–2425; 10.1038/ismej.2017.122 (2017).

37. Zou, D., Liu, H. & Li, M. Community, Distribution, and Ecological Roles of Estuarine Archaea. *Frontiers in microbiology* **11,** 2060; 10.3389/fmicb.2020.02060 (2020).

38. Bang, C. & Schmitz, R. A. Archaea associated with human surfaces: not to be underestimated. *FEMS microbiology reviews* **39,** 631–648; 10.1093/femsre/fuv010 (2015).

39. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* **74,** 5088–5090; 10.1073/pnas.74.11.5088 (1977).

40. Doolittle, W. F. Evolution: Two Domains of Life or Three? *Current biology : CB* **30,** R177-R179; 10.1016/j.cub.2020.01.010 (2020).

41. Williams, T. A., Cox, C. J., Foster, P. G., Szöllősi, G. J. & Embley, T. M. Author Correction: Phylogenomics provides robust support for a two-domains tree of life. *Nature ecology & evolution* **4,** 1568; 10.1038/s41559-020-01347-2 (2020).

42. Guy, L. & Ettema, T. J. G. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in Microbiology* **19,** 580–587; 10.1016/j.tim.2011.09.002 (2011).

43. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499,** 431–437; 10.1038/nature12352 (2013).

44. Castelle, C. J. *et al.* Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Current biology : CB* **25,** 690–701; 10.1016/j.cub.2015.01.014 (2015).

45. Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science (New York, N.Y.)* **357**; 10.1126/science.aaf3883 (2017).

46. Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A. & Forterre, P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS genetics* **13,** e1006810; 10.1371/journal.pgen.1006810 (2017).

47. Raymann, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 6670–6675; 10.1073/pnas.1420858112 (2015).

48. Dombrowski, N., Lee, J.-H., Williams, T. A., Offre, P. & Spang, A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS microbiology letters* **366**; 10.1093/femsle/fnz008 (2019).

49. Bell, S. D. & Jackson, S. P. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends in Microbiology* **6,** 222–228; 10.1016/S0966-842X(98)01281-5 (1998).

50. Grohmann, D. & Werner, F. Recent advances in the understanding of archaeal transcription. *Current Opinion in Microbiology* **14,** 328–334; 10.1016/j.mib.2011.04.012 (2011).

51. Jain, S., Caforio, A. & Driessen, A. J. M. Biosynthesis of archaeal membrane ether lipids. *Frontiers in microbiology* **5,** 641; 10.3389/fmicb.2014.00641 (2014).

52. Klingl, A., Pickl, C. & Flechsler, J. Archaeal Cell Walls. *Sub-cellular biochemistry* **92,** 471–493; 10.1007/978-3-030-18768-2_14 (2019).

53. Rodrigues-Oliveira, T., Belmok, A., Vasconcellos, D., Schuster, B. & Kyaw, C. M. Archaeal S-Layers: Overview and Current State of the Art. *Frontiers in microbiology* **8,** 2597; 10.3389/fmicb.2017.02597 (2017).

54. Albers, S. V., van de Vossenberg, J. L., Driessen, A. J. & Konings, W. N. Adaptations of the archaeal cell membrane to heat stress. *Frontiers in bioscience : a journal and virtual library* **5,** D813-20; 10.2741/albers (2000).

55. Sato, T. & Atomi, H. Novel metabolic pathways in Archaea. *Current Opinion in Microbiology* **14,** 307–314; 10.1016/j.mib.2011.04.014 (2011).

56. Qin, W. *et al.* Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. *The ISME journal* **14,** 2595–2609; 10.1038/s41396-020-0710-7 (2020).

57. Kengen, S. W. *et al.* Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by Pyrococcus furiosus. *Journal of Biological Chemistry* **269,** 17537–17541; 10.1016/S0021-9258(17)32474-2 (1994).

58. Ronimus, R. S., Heus, E. de & Morgan, H. W. Sequencing, expression, characterisation and phylogeny of the ADP-dependent phosphofructokinase from the hyperthermophilic, euryarchaeal Thermococcus zilligii. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1517,** 384–391; 10.1016/s0167-4781(00)00301-8 (2001).

59. van der Oost, J. *et al.* The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the hyperthermophilic archaeon Pyrococcus furiosus represents a novel site of glycolytic regulation. *Journal of Biological Chemistry* **273,** 28149–28154; 10.1074/jbc.273.43.28149 (1998).

60. Say, R. F. & Fuchs, G. Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* **464,** 1077–1081; 10.1038/nature08884 (2010).

61. Soderberg, T. Biosynthesis of ribose-5-phosphate and erythrose-4-phosphate in archaea: a phylogenetic analysis of archaeal genomes. *Archaea (Vancouver, B.C.)* **1,** 347–352; 10.1155/2005/314760 (2005).

62. Orita, I. *et al.* The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon Thermococcus kodakaraensis. *Journal of bacteriology* **188,** 4698–4704; 10.1128/JB.00492-06 (2006).

63. Buan, N. R. Methanogens: pushing the boundaries of biology. *Emerging topics in life sciences* **2,** 629–646; 10.1042/ETLS20180031 (2018).

64. Leigh, J. A., Albers, S.-V., Atomi, H. & Allers, T. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS microbiology reviews* **35,** 577–608; 10.1111/j.1574-6976.2011.00265.x (2011).

65. Amend, J. P. & Shock, E. L. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and bacteria. *FEMS microbiology reviews* **25,** 175–243; 10.1111/j.1574-6976.2001.tb00576.x (2001).

66. Schut, G. J. *et al.* The Order Thermococcales and the Family Thermococcaceae. In *The Prokaryotes*, edited by E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt & F. Thompson (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014), pp. 363–383.

67. Ronimus, R. S., Reysenbach, A., Musgrave, D. R. & Morgan, H. W. The phylogenetic position of the Thermococcus isolate AN1 based on 16S rRNA gene sequence analysis: a proposal that AN1 represents a new species, Thermococcus zilligii sp. nov. *Archives of microbiology* **168,** 245–248; 10.1007/s002030050495 (1997).

68. Holden, J. F. *et al.* Diversity among three novel groups of hyperthermophilic deep-sea Thermococcus species from three sites in the northeastern Pacific Ocean. *FEMS microbiology ecology* **36,** 51–60; 10.1111/j.1574-6941.2001.tb00825.x (2001).

69. Bertoldo, C. & Antranikian, G. The Order Thermococcales. In *The Prokaryotes*, edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt (Springer New York, New York, NY, 2006), pp. 69–81.

70. Marteinsson, V. T. *et al.* Thermococcus barophilus sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *International journal of systematic bacteriology* **49 Pt 2,** 351–359; 10.1099/00207713-49-2-351 (1999).

71. Jolivet, E., Corre, E., L'Haridon, S., Forterre, P. & Prieur, D. Thermococcus marinus sp. nov. and Thermococcus radiotolerans sp. nov., two hyperthermophilic archaea from deep-sea hydrothermal vents that resist ionizing radiation. *Extremophiles : life under extreme conditions* **8,** 219–227; 10.1007/s00792-004-0380-9 (2004).

72. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of Thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as Pyrococcus sp. KOD1. *Archaea (Vancouver, B.C.)* **1,** 263–267; 10.1155/2004/204953 (2004).

73. Gavrilov, S. N. *et al.* Isolation and Characterization of the First Xylanolytic Hyperthermophilic Euryarchaeon Thermococcus sp. Strain 2319x1 and Its Unusual Multidomain Glycosidase. *Frontiers in microbiology* **7,** 552; 10.3389/fmicb.2016.00552 (2016).

74. Santos, C. R. *et al.* Structural basis for branching-enzyme activity of glycoside hydrolase family 57: structure and stability studies of a novel branching enzyme from the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *Proteins* **79,** 547–557; 10.1002/prot.22902 (2011).

75. Zona, R., Chang-Pi-Hin, F., O'Donohue, M. J. & Janecek, S. Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from Thermococcus hydrothermalis. *European journal of biochemistry* **271,** 2863–2872; 10.1111/j.1432-1033.2004.04144.x (2004).

76. Singleton, M. R. & Littlechild, J. A. Pyrrolidone carboxylpeptidase from Thermococcus litoralis. In *Hyperthermophilic Enzymes Part A* (Elsevier2001), Vol. 330, pp. 394–403.

77. Ellens, K. W. *et al.* Confronting the catalytic dark matter encoded by sequenced genomes. *Nucleic acids research* **45,** 11495–11514; 10.1093/nar/gkx937 (2017).

78. Sime, J. T. Applications of Biocatalysis to Industrial Processes. *J. Chem. Educ.* **76,** 1658; 10.1021/ed076p1658 (1999).

79. Karam, J. & Nicell, J. A. Potential Applications of Enzymes in Waste Treatment. *J. Chem. Technol. Biotechnol.* **69,** 141–153; 10.1002/(SICI)1097-4660(199706)69:2<141::AID-JCTB694>3.0.CO;2-U (1997).

80. Bhat, M. K. Cellulases and related enzymes in biotechnology. *Biotechnology Advances* **18,** 355–383; 10.1016/S0734-9750(00)00041-0 (2000).

81. Singhania, R. R. *et al.* Industrial Enzymes. In *Industrial Biorefineries & White Biotechnology* (Elsevier2015), pp. 473–497.

82. Hughes, G. & Lewis, J. C. Introduction: Biocatalysis in Industry. *Chemical reviews* **118,** 1–3; 10.1021/acs.chemrev.7b00741 (2018).

83. Taylor, I. N. *et al.* Application of thermophilic enzymes in commercial biotransformation processes. *Biochemical Society transactions* **32,** 290–292; 10.1042/bst0320290 (2004).

84. Zamost, B. L., Nielsen, H. K. & Starnes, R. L. Thermostable enzymes for industrial applications. *Journal of Industrial Microbiology* **8,** 71–81; 10.1007/BF01578757 (1991).

85. Unsworth, L. D., van der Oost, J. & Koutsopoulos, S. Hyperthermophilic enzymes--stability, activity and implementation strategies for high temperature applications. *The FEBS journal* **274,** 4044–4056; 10.1111/j.1742-4658.2007.05954.x (2007).

86. Vieille, C. & Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and molecular biology reviews : MMBR* **65,** 1–43; 10.1128/MMBR.65.1.1-43.2001 (2001).

87. Ramos, A. *et al.* Stabilization of Enzymes against Thermal Stress and Freeze-Drying by Mannosylglycerate. *Applied and environmental microbiology* **63,** 4020–4025; 10.1128/aem.63.10.4020-4025.1997 (1997).

88. Kaushik, J. K. & Bhat, R. Why is trehalose an exceptional protein stabilizer? An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose. *Journal of Biological Chemistry* **278,** 26458–26465; 10.1074/jbc.M300815200 (2003).

89. Ellis, R. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Current Opinion in Structural Biology* **11,** 114–119; 10.1016/S0959-440X(00)00172-X (2001).

90. Matthews, B. W., Nicholson, H. & Becktel, W. J. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proceedings of the National Academy of Sciences of the United States of America* **84,** 6663–6667; 10.1073/pnas.84.19.6663 (1987).

91. Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. Directed evolution of a thermostable esterase. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 12809–12813; 10.1073/pnas.95.22.12809 (1998).

92. Lehmann, M. & Wyss, M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current Opinion in Biotechnology* **12,** 371–375; 10.1016/S0958-1669(00)00229-9 (2001).

93. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3,** 853–859; 10.1016/S0969-2126(01)00220-9 (1995).

94. Webb, E. C. (ed.). *Enzyme nomenclature. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and molecular biology on the nomenclature and classification of enzymes.* 1984th ed. (Acad. Press, San Diego, Calif., 1992).

95. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal* **280 (Pt 2),** 309–316; 10.1042/bj2800309 (1991).

96. Henrissat, B. & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal* **293 (Pt 3),** 781–788; 10.1042/bj2930781 (1993).

97. Stracke, C. *et al.* Salt Stress Response of Sulfolobus acidocaldarius Involves Complex Trehalose Metabolism Utilizing a Novel Trehalose-6-Phosphate Synthase (TPS)/Trehalose-6-Phosphate

Phosphatase (TPP) Pathway. *Applied and environmental microbiology* **86**; 10.1128/AEM.01565-20 (2020).

98. Wei, S. The application of biotechnology on the enhancing of biogas production from lignocellulosic waste. *Applied microbiology and biotechnology* **100,** 9821–9836; 10.1007/s00253-016-7926-5 (2016).

99. Shrivastava, S. *Industrial Applications of Glycoside Hydrolases* (Springer Singapore, Singapore, 2020).

100. Bachovchin, D. A. & Cravatt, B. F. The pharmacological landscape and therapeutic potential of serine hydrolases. *Nature reviews. Drug discovery* **11,** 52–68; 10.1038/nrd3620 (2012).

101. Dalmaso, G. Z. L., Ferreira, D. & Vermelho, A. B. Marine extremophiles: a source of hydrolases for biotechnological applications. *Marine drugs* **13,** 1925–1965; 10.3390/md13041925 (2015).

102. Patricelli, M. P., Giang, D. K., Stamp, L. M. & Burbaum, J. J. Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes. *Proteomics* **1,** 1067–1071; 10.1002/1615-9861(200109)1:9<1067::AID-PROT1067>3.0.CO;2-4 (2001).

103. Casida, J. E. & Quistad, G. B. Serine hydrolase targets of organophosphorus toxicants. *Chemico-biological interactions* **157-158,** 277–283; 10.1016/j.cbi.2005.10.036 (2005).

104. McKinney, M. K. & Cravatt, B. F. Evidence for distinct roles in catalysis for residues of the serine-serine-lysine catalytic triad of fatty acid amide hydrolase. *Journal of Biological Chemistry* **278,** 37393–37399; 10.1074/jbc.M303922200 (2003).

105. Page, M. J. & Di Cera, E. Serine peptidases: classification, structure and function. *Cellular and molecular life sciences : CMLS* **65,** 1220–1236; 10.1007/s00018-008-7565-9 (2008).

106. Ekici, O. D., Paetzel, M. & Dalbey, R. E. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein science : a publication of the Protein Society* **17,** 2023–2037; 10.1110/ps.035436.108 (2008).

107. Chistoserdovai, L. Functional metagenomics: recent advances and future challenges. *Biotechnology & genetic engineering reviews* **26,** 335–352; Review (2010).

108. Ferrer, M. *et al.* Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial biotechnology* **9,** 22–34; 10.1111/1751-7915.12309 (2016).

109. Park, S.-J., Kang, C.-H., Chae, J.-C. & Rhee, S.-K. Metagenome microarray for screening of fosmid clones containing specific genes. *FEMS microbiology letters* **284,** 28–34; 10.1111/j.1574-6968.2008.01180.x (2008).

110. Prakash, T. & Taylor, T. D. Functional assignment of metagenomic data: challenges and applications. *Briefings in bioinformatics* **13,** 711–727; 10.1093/bib/bbs033 (2012).

111. Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D. & Charles, T. C. Current and future resources for functional metagenomics. *Frontiers in microbiology* **6,** 1196; 10.3389/fmicb.2015.01196 (2015).

112. Gilbert, J. *et al.* Bioprospecting metagenomics for new glycoside hydrolases. *Methods in molecular biology (Clifton, N.J.)* **908,** 141–151; 10.1007/978-1-61779-956-3_14 (2012).

113. Thankappan, S. *et al.* Bioprospecting thermophilic glycosyl hydrolases, from hot springs of Himachal Pradesh, for biomass valorization. *AMB Express* **8,** 168; 10.1186/s13568-018-0690-4 (2018).

114. Takeda, M. *et al.* Metagenomic mining and structure-function studies of a hyper-thermostable cellobiohydrolase from hot spring sediment. *Communications biology* **5,** 247; 10.1038/s42003-022-03195-1 (2022).

115. Sørensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology* **115,** 113–128; 10.1016/j.jbiotec.2004.08.004 (2005).

116. Warren, R. L. *et al.* Transcription of foreign DNA in Escherichia coli. *Genome research* **18,** 1798–1805; 10.1101/gr.080358.108 (2008).

117. Sakuraba, H. & Ohshima, T. Heterologous Production of Thermostable Proteins and Enzymes. In *Thermophilic Microbes in Environmental and Industrial Biotechnology*, edited by T. Satyanarayana, J. Littlechild & Y. Kawarabayasi (Springer Netherlands, Dordrecht, 2013), pp. 395–412.

118. Mo, X., Chen, C., Pang, H., Feng, Y. & Feng, J. Identification and characterization of a novel xylanase derived from a rice straw degrading enrichment culture. *Applied microbiology and biotechnology* **87,** 2137–2146; 10.1007/s00253-010-2712-2 (2010).

119. Patil, R. Isolation of polyvinyl chloride degrading bacterial strains from environmental samples using enrichment culture technique. *Afr. J. Biotechnol.* **11**; 10.5897/AJB11.3630 (2012).

120. Cravatt, B. Chemical strategies for the global analysis of protein function. *Current Opinion in Chemical Biology* **4,** 663–668; 10.1016/s1367-5931(00)00147-2 (2000).

121. Liu, Y., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 14694–14699; 10.1073/pnas.96.26.14694 (1999).

122. Faucher, F., Bennett, J. M., Bogyo, M. & Lovell, S. Strategies for Tuning the Selectivity of Chemical Probes that Target Serine Hydrolases. *Cell chemical biology* **27,** 937–952; 10.1016/j.chembiol.2020.07.008 (2020).

123. Jessani, N. & Cravatt, B. F. The development and application of methods for activity-based protein profiling. *Current Opinion in Chemical Biology* **8,** 54–59; 10.1016/j.cbpa.2003.11.004 (2004).

124. Powers, J. C., Asgian, J. L., Ekici, O. D. & James, K. E. Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chemical reviews* **102,** 4639–4750; 10.1021/cr010182v (2002).

125. Speers, A. E. & Cravatt, B. F. Profiling enzyme activities in vivo using click chemistry methods. *Chemistry & biology* **11,** 535–546; 10.1016/j.chembiol.2004.03.012 (2004).

126. Speers, A. E. & Cravatt, B. F. Activity-Based Protein Profiling (ABPP) and Click Chemistry (CC)-ABPP by MudPIT Mass Spectrometry. *Current protocols in chemical biology* **1,** 29–41; 10.1002/9780470559277.ch090138 (2009).

127. Stephen Dahms, A. 3-Deoxy-D-pentulosonic acid aldolase and its role in a new pathway of D-xylose degradation. *Biochemical and Biophysical Research Communications* **60,** 1433–1439; 10.1016/0006-291X(74)90358-1 (1974).

128. Bruinenberg, P. M., Bot, P. H. M., Dijken, J. P. & Scheffers, W. A. The role of redox balances in the anaerobic fermentation of xylose by yeasts. *European J. Appl. Microbiol. Biotechnol.* **18,** 287–292; 10.1007/BF00500493 (1983).

129. Hochster, R. M. & Watson, R. W. Enzymatic isomerization of d-xylose to d-xylulose. *Archives of Biochemistry and Biophysics* **48,** 120–129; 10.1016/0003-9861(54)90313-6 (1954).

130. Sutter, J.-M., Johnsen, U., Reinhardt, A. & Schönheit, P. Pentose degradation in archaea: Halorhabdus species degrade D-xylose, L-arabinose and D-ribose via bacterial-type pathways. *Extremophiles : life under extreme conditions* **24,** 759–772; 10.1007/s00792-020-01192-y (2020).

131. Almqvist, H., Jonsdottir Glaser, S., Tufvegren, C., Wasserstrom, L. & Lidén, G. Characterization of the Weimberg Pathway in Caulobacter crescentus. *Fermentation* **4,** 44; 10.3390/FERMENTATION4020044 (2018).

132. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46,** W296-W303; 10.1093/nar/gky427 (2018).

133. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science : a publication of the Protein Society* **30,** 70–82; 10.1002/pro.3943 (2020).

134. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39,** W29-37; 10.1093/nar/gkr367 (2011).

135. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **31,** 455–461; 10.1002/jcc.21334 (2010).

136. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering* **8,** 127–134; 10.1093/protein/8.2.127 (1995).

I. Danksagung

Allen voran möchte ich mich bei meiner Doktormutter Frau Bettina Siebers für die Möglichkeit in ihrer Gruppe an diesem spannenden Projekt arbeiten können. Ich bin dankbar für den wissenschaftlichen Input, die wohlwollende Betreuung, als auch den vielen Freiraum, den ich zur Bearbeitung meiner Fragestellungen hatte. Gleichermaßen bedanke ich mich für wissenschaftliche und technische Ratschläge von Herrn Dr. Christopher Bräsen, der jederzeit ein offenenes Ohr hatte und mit Rat zur Seite stand. Beiden bin ich darüber hinaus dankbar für unermüdliches Korrekturlesen von mir entworfener Manuskripte, sowie dieser Thesis. Herrn Prof. Dr. Markus Kaiser bin ich ebenfalls dankbar für sein Mitwirken und Korrekturlesen meiner Manuskripte, sowie für seine Bereitschaft das Zweitgutachten meiner Thesis zu übernehmen. Der Deutschen Forschungsgesellschaft (DFG) gebührt ebenfalls Dank für die Bereitstellung der Gelder für das ExoCarb Projekt, in dessen Rahmen diese Arbeit durchgeführt wurde. An dieser Stelle möchte ich mich bei allen Mitwirkenden Projektpartnern bedanken. Ich danke Frau Dr. Sabrina Ninck dafür, dass sie mit ihren Kenntnissen und Fähigkeiten insbesondere bei den ABPP und Massenspektrometrie- Experimenten einen essentiellen Beitrag für diese Arbeit geleistet hat und durch ihr akribisches Arbeiten stets für eine hohe Qualität der erhobenen Daten gesorgt hat. Herrn Dr. Ilya Kublanov möchte ich meinen Dank ausdrücken für zahlreiche bioinformatische und mikrobiologische Arbeit, sowohl für die einmalige Möglichkeit spannende Feldversuche in der Uzon-Caldera Region in Kamtschatka durchzuführen, bei denen ich von Frau Dr. Tatjana Kochetkova betreut wurde. Weiterer Dank gilt Frau Sarah P. Esser, sowie ihrem Doktorvater Prof. Dr. Alexander Probst, die mit ihrer herrausragenden Expertise in der Bioinformatik das *in situ*-ABPP Projekt auf ein neues Level gehoben haben. Die tägliche Arbeit in den Laboren der MEB-Gruppe wurde sehr bereichert durch alle AG-Mitglieder, inklusive Frau Agathe Materla, Thomas Knurra and Sabiene Dietl, die stets mit administrativen und technischen Hilfestellungen zur Seite standen. In besonderer Art bin ich Christian Schmerling für wissenschaftliche Anregungen dankbar, die sich oft als überaus hilfreich herausgestellt haben. Ebenso danke ich Dr. Benjamin Meyer für die gute Büro-Nachbarschaft und angenehme Gespräche, sowie dem Rest der Gruppe für eine stests angenehme und lustige Atmosphäre; Dr. Christina Stracke, Dr. Lu Shen, Dr. Xiaoxiao Zhu, Svenja Höfmann, Laura Kuschmierz, Alexander Wagner, Larissa Schocke, Astrid Neu, Sabine Krevet, Carmen Peraglie und Carsten Schröder. Darüber hinaus möchte ich auch meinen ehemaligen Bachelor und Masterstudentinnen Lobna Eltoukhy, Ying Jia Low und Michaela Bojara für ihre Arbeit im Labor und am Computer danken. Großer Dank gilt auch meiner Familie, allen voran meinen Eltern und Omas, die mich auf dem Weg zur Promotion stets unterstützt und den Rücken frei gehalten haben. Außerdem danke ich allen Freunden die mich auf diesem Weg begleitet haben, vor allem den Einhörnern Dr. Dominik Karrer, Nikolai Huwa und Kai Krämer für spannende, stets spaßige, oft hilfreiche und teils absurde Diskussionen.

II. Curriculum Vitae

III. Verfassungserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit:

Establishing Activity Based Protein Profiling as a robust Method for Bioprospecting of Active Hydrolases in (Hyper)thermophilic Environments

ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht. Ich bestätige, dass die eingereichten Arbeiten meine eigenen sind, es sei denn, die Arbeiten waren ein Teil von gemeinsam verfassten Veröffentlichungen. Ich versichere außerdem, dass ich die vorliegende Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Ort, Datum (Thomas Klaus)

IV. Appendix

Identification of potential D-xylose degradation pathway enzymes in *Thermococcus* sp. 2319x1E

As described in chapter 3.1, *Thermococcus* sp. 2319x1E is able to grow on beechwood xylan or its main degradation product D-xylose as sole carbon and energy source. Consequently, the cells should be capable to metabolize D-xylose and to channel the respecitve metabolites into the central carbon metabolism pathway, such as glycolysis, pentose phosphate pathway or citric acid cycle. Commonly in bacteria, D-xylose and other pentoses can be metabolized by the oxidative Weimberg pathway, in which D-xylose is oxidized via D-xylono-lactone to D-xylonic acid, which is furtherly dehydrated to 2-keto-3-deoxy-xylonate. In a second dehydration step, α-ketoglutarate semialdehyde is formed and subsequently converted to 2-ketoglutarate which enters the citric acid cycle as a key intermediate. In some bacteria, a variation of the Weimberg pathway, the so-called Dahms pathway, is used, in which 2-keto-3 deoxy-xylonate is split to pyruvate and glycolaldehyde by an aldolase [35,127]. Another oxidative pathway present e.g. in yeast, is the oxido-reductase pathway; D-xylose is reduced to xylitol by a NAD(P)H dependent xylose reductase, then xylitol is oxidized by a NAD^+ dependent xylitol dehydrogenase to xylulose, which then enters the pentose phosphate pathway after phosphorylation to xylulose-5-phosphate by an ATP dependent xylulose kinase [128]. Additionally to these patways another common bacterial pathway is fully non-oxidative an generates xylulose by isomerization from D-xylose and proceeds further via a xylulose kinase, as described above [129]. While pentose degradation in *Haloferax volcanii,* in *Haloarcula* and *Sulfolobus* species proceedes via oxidative Weimberg/Dahms pathways, a non-oxidative bacterial isomerase-type pathway via xylose isomerase and xylulose kinase was reported for halophilic *Halorhabdus* species [130]. Yet in the genome of *Thermococcus* sp. 2319x1E no homologs of any common enzyme involved in D-xylose degradation could be easily identified among the annotated proteins. In comparative proteomic studies of *Thermococcus* sp. 2319x1E cells grown on D-xylose compared to cells grown on D-glucose, several enzymes with putative role in sugar metabolism were identified and depicted with their respective logfoldchange in table 1. Noteworthy, there are 3 upregulated kinases, whereas EGIDFPOO_00820 was the kinase with the highest log-fold change and was subjected to thorough bioinformetical analysis.

Table 1: Selected proteins with putative function in sugar/energy metabolism which were upregulated in* Thermococcus *sp. 2319x1E cells grown on D-xylose compared to cells grown on D-glucose.

Gene number	annotation	-log(P) value
EGIDFPOO_01390	Ribokinase	1.411
EGIDFPOO_02004	putative2-dehydro-3-deoxy-D-pentonatealdolaseYjhH	1.818
EGIDFPOO_01938	NADP-dependentglyceraldehyde-3-phosphatedehydrogenase	1.825
EGIDFPOO_01960	Ribulose-bisphosphate carboxylase largechain	2.161
EGIDFPOO_00819	Glycerophosphodiesterphosphodiesterase	2.533
EGIDFPOO_00035	Ribose-phosphatepyrophosphokinase	2.537
EGIDFPOO_00820	Glycerolkinase	2.703
EGIDFPOO_01680	Fructose-1,6-bisphosphatealdolase/phosphatase	3.573

Beside the two kinases EGIDFPOO_00820 and EGIDFPO1390, which were found as significantly upregulated during growth on D-xylose, three enzymes annotated as endonuclease 4 (EGIDFPOO_00142 and EGIDFPOO_01100) or hypothetical protein (EGIDFPOO_00419) were selected for detailed bioinformatical analysis, revealing that they contain xylose isomerase (XI) domains and show significant structural similarity to the the xylose isomerase from psychrophilic bacterium *Paenibacillus* sp. R4 (see table 2). Thus, these potential isomerases are valid candidates for xylose isomerases in *Thermococcus* sp. 2319x1E. For the ribokinase EGIDFPOO_01390 and the glycerolkinase EGIDFPOO_00820 have been found carbohydrate kinase domains (PFAM) and most interestingly, for EGIDFPOO_00820 there was found a distinct close structural homology to the xylulokinase of the gramnegative bacterium *Brucella ovis* (PDB ID 5VM1_B).

Table 2: Analysis of PFAM domains and structural homologies of selected enzymes with putative involvment on D-xylose degradation in Thermococcus sp. 2319x1E.

Gene number	annotation	proteome	PFAM	E-value	HHPRED	E-value
EGIDFPOO_00820	Glycerolkinase	up	FGGY family of carbohydrate kinases	1.5e-87 (N), 2.2e-50 (C)	Xylulokinase 5VM1_B	2.00E-66
EGIDFPOO_01390	Ribokinase	up	pfkB family carbohydrate kinase	1.30E-47	sugar kinase, ribokinase family 2RBC_A	5.00E-35
EGIDFPOO_00142	endonuclease 4	ND	XI -like TIM barrel	1.10E-35	XI 7VPF_A	1.50E-34
EGIDFPOO_01100	endonuclease 4	ND	XI-like TIM barrel	2.20E-31	XI 7VPF_A	6.00E-33
EGIDFPOO_00419	hypothetical protein	ND	XI-like TIM barrel	1.20E-21	XI 7VPF_A	4.50E-32

Reported D-xylose degradation pathways, including the isomerase patway (xylose isomerase XI highlighted with a light red box, xylulose kinase XK hinglighted with a blue box) is depicted in Fig. 1. Since the *in silico* analysis of EGIDFPOO_00142, EGIDFPOO_00419 and EGIDFPOO_01100 indicates the presence of potential xylose isomerases and EGIDFPOO_00820 is a promising candidate for an xylulose kinase, an isomerase pathway for the degradation of D-xylose can be proposed for *Thermococcus* sp. 2319x1E.

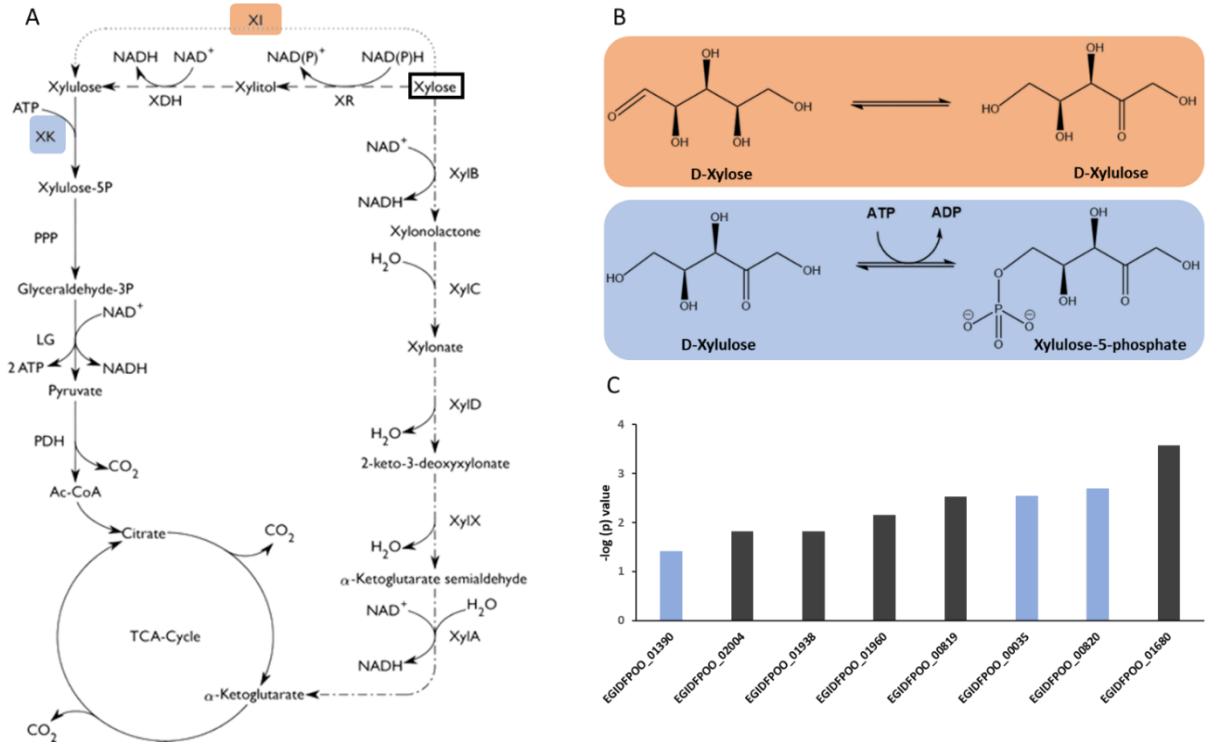


Figure 6: D-Xylose metabolism pathways. (A) *was adapted from Almquist et al* [131] *and shows known routes of D-xylose metabolism. Xylose isomerase (XI) and Xylulose kinase (XK) are highlighted with a light red and light blue box.* ***(B)*** *Reactions catalyzed by XI and XK are the isomerization of D-xylulose to D-Xylulose (XI) and the subsequent phosphorylation of D-xylulose to xylulose-5-phosphate, which then enters the pentose phosphate pathway (PPP) to yield pyruvate in the end.* ***(C)*** *in comparative proteomic studies of* Thermococcus *sp. 2319x1E cells grown on either D-xylose or D-glucose, a set of enzymes with putative involvement in sugar metabolism was upregulated in proteome samples of the cells grown on D-xylose with some enzymes annotated as putative kinases (light blue bars). These potential kinases are considered as candidates for XK activity.*

Due to the upregulation of EGIDFPOO_00820 in cells grown on D-xylose and its structural homology to a reported xylulose kinase, structural moddeling and subsequent substrate docking experiments have been done, to provide supplementary indications for the ability of EGIDFPOO_00820 to bind the relevant substrates (D-xylose, D-xylulose, xylitol and xylulose-5-phosphate). These *in silico* experiments were performed by Ying Jia Low (Masterstudent 2020) under my supervision. SWISS-MODEL [132] was used to identify the structure 2itm of the xylulose kinase from *E. coli* as a suitable template and subsequent model construction (p-score 7.01e-16; QMEAN score 0.64; QMQE score -3.73). With the resulting PDB file, a protein

structure was visualized using UCSF Chimera [133] (see Fig. 2A). The aspartat residue D239 was predicted as a catalytic and substrate docking was done in the surrounding cleft, using using AutoDock Vina 4.2.6 (Trott and Olson 2010) with the substrates D-xylose (compound ID 135191), xylitol (compound ID 6912), D-xylulose (compound ID 5289590) and xylulose-5-phosphate (compound ID 5459820), which where obtained from the NCBI Pubchem database. Positive binding results were obtained for xylitol (-4.0 kcal/mol, hydrogen bridges to D9, T12 and D239), D-xylose (-4.8 kcal/mol, hydrogen bridges to D9, T12 and D239) and D-xylulose (-4 kcal/mol, hydrogen bridges to T12, T13 and D239), but not to xylulose-5-phosphate. The docking result with D-xylulose was visualized with LigPlot+ (Wallace et al. 1995) (Fig. 4B).

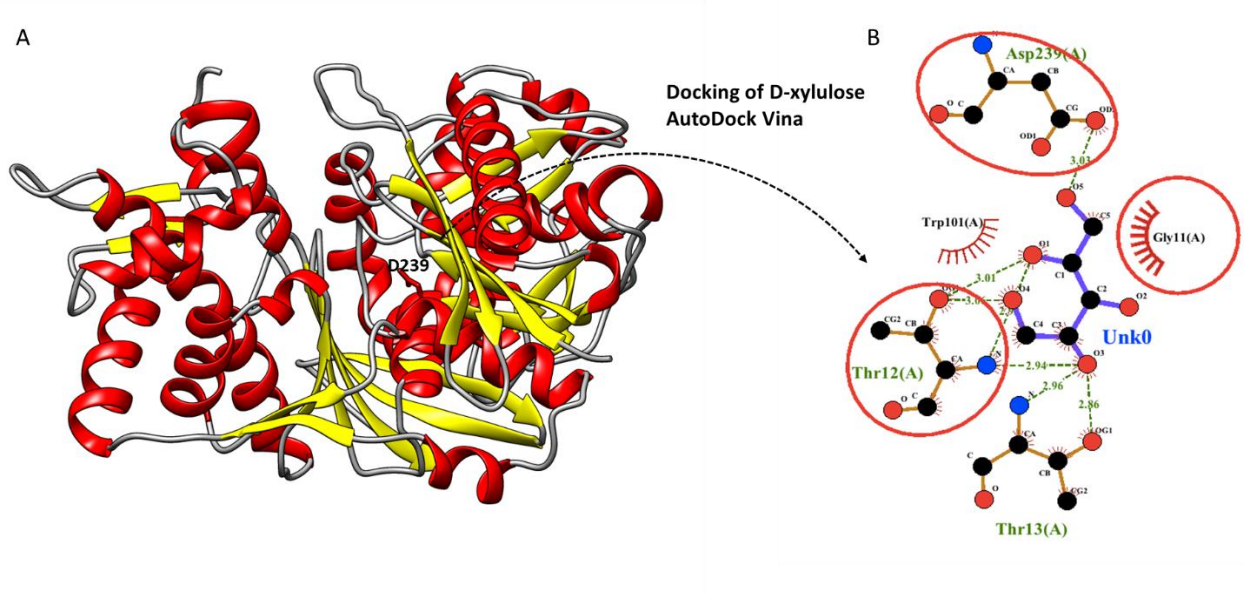


Figure 7: Structural model of the putative kinase EGIDFPOO_00820 and proposed binding of D-Xylulose. ***(A)*** *A structural model was constructed using SWISS-MODEL [132] with the template 2itm (p-score 7.01e-16; QMEAN score 0.64; QMQE score -3.73). D239 was predicted as a catalytic residue by HMMER [134]. The model was visualized with UCSF Chimera [133].* ***(B)*** *Substrate docking into the moddeled structure was done using AutoDock Vina 4.2.6 [135] with the substrates xylose, xylitol, xylulose and xylulose-5-phosphate, which where obtained from the NCBI Pubchem database. Docking was done within the binding groove in which the residue D239 was identified. The docking result with xylulose (compound ID: 5289590) gave a binding affinity of -4.0 kcal/mol. The respective docking model was imported to Ligplot+ [136] to visualize the binding situation and involved residues.*

Complementary to the comparative proteomic results, which schow at least three different upregulated carbohydrate kinases, comparative biochemical assays with *Thermococcus* sp. 2319x1E crude extract have been performed. Crude extract for the kinase assay was obtained freshly from cultures grown anaerobically overnight at 85° C in 200 ml bottles as described in the methods section in the manuscript 3.1 in modified Pfennig medium

containing either 0.1 % (w/v) D-glucose or D-xylose as carbon source. Cells were obtained by centrifugation at 7000 x g for 20 min at room temperature and were resuspended in Phosphate-buffered saline (PBS) with protease inhibitor cOmplete ULTRA and 1 mM of the redox agent DTT. Cell disruption was done using a Precellys 14 homogeniser (Bertini Techologies, Darmstadt, Germany) for 15 secs at 4000 rpm x 3 times with subsequent centrifugation of the cell lysate at 12,000 x g for 40 min at 4°C. Proteinconcentration of the supernatant (crude extract) was done using Bradford assay. The kinase reaction was determined photometrically on a Specord 210 spectrophotometer (Analytik Jena, Germany) in 0.5 ml glass cuvettes containing 0.1 M 2-(N-morpholino) ethanesulfonic acid (MES), 20 mM MgCl2, 0.1 M KCl, 10 mM ATP, 10 mM phosphoenolpyruvate, 5 mM NADH, 1 mM DTT and 24.5 - 27.4 µg *Thermococcus* crude extract. 7.0 U ml^{-1} pyruvate kinase and 15.3 U ml^{-1} lactate dehydrogenase were used as auxiliary enzymes. The reaction was started by the addition of 10 mM D-xylulose/ D-xylose/ xylit or glycerol and kinase activity was calculated from the decrease in absorption at 340 nm due to the oxidation of NADH to NAD^+. Samples were measured in triplicates, revealing significant differences in the phosphorylation of D-xylulose or glycerol in the crude extracts and non-significant differences for D-xylose and D-xylitol (Fig. 5). Interestingly, the specific kinase activity (U per mg of total protein in the sample) on D-xylulose in cell extract obtained from cells grown on D-xylose was 0.024 U mg^{-1} and thus roughly 3.7 times higher than in the crude extracts obtained from cells grown on D-glucose (0.006 U mg^{-1}). The kinase activity on glycerol on the other hand was overall significantly higher in both crude extracts, although the activity in D-glucose grown cells was roughly 1.8 times higher than in D-xylose grown cells (0.154 U mg^{-1} compared to 0.084 U mg^{-1}).

This experiment demonstrates the presence of kinases in *Thermococcus sp.* 2319x1E, that are capable of phosphorylating D-xylulose, which is a key step of D-xylose metabolism via the isomerase pathway or the oxidoreductase pathway. Moreover, it was demonstrated, that D-xylulose phosphorylation is induced, when the cells grow on D-xylose as their sole carbon source, compared to cells grown on D-glucose. This observation coincedes with the observation of upregulated kinases in the proteome of respective cells – most importantly to be mentioned here is the enzyme EGIDFPOO_00820, which is annotated as glycerol kinase, but shows distinct structural similarity to xylulose kinases (table 2). Though, the upregulation of glycerol kinase activity in cells grown on D-glucose compared to cells grown on D-xylose does not fit to the observation of significantly less expression of the resepctive enzyme. However, this could also be due to other factors, such as post-translational modification or simply the presence of other enzymes enabling the phosphorylation of glycerol. However, the putative xylulose kinase

EGIDFPOO_00820 and the putative xylose isomerase EGIDFPOO_00142 have been cloned into pET28b vectors but no sufficient heterologous expression of active enzymes could be obtained in *E. coli*. In order to furtherly explore D-xylose degradation in *Thermococcus sp.* 2319x1E and other *Thermococcales*, proper biochemical characterization of EGIDFPOO_00820 as well as growth experiments with knock-out strains, lacking *EGIDFPOO_00820* and beforementioned putative xylose isomerase genes *EGIDFPOO_00142*, *EGIDFPOO_01100* and *EGIDFPOO_00419* should be performed.

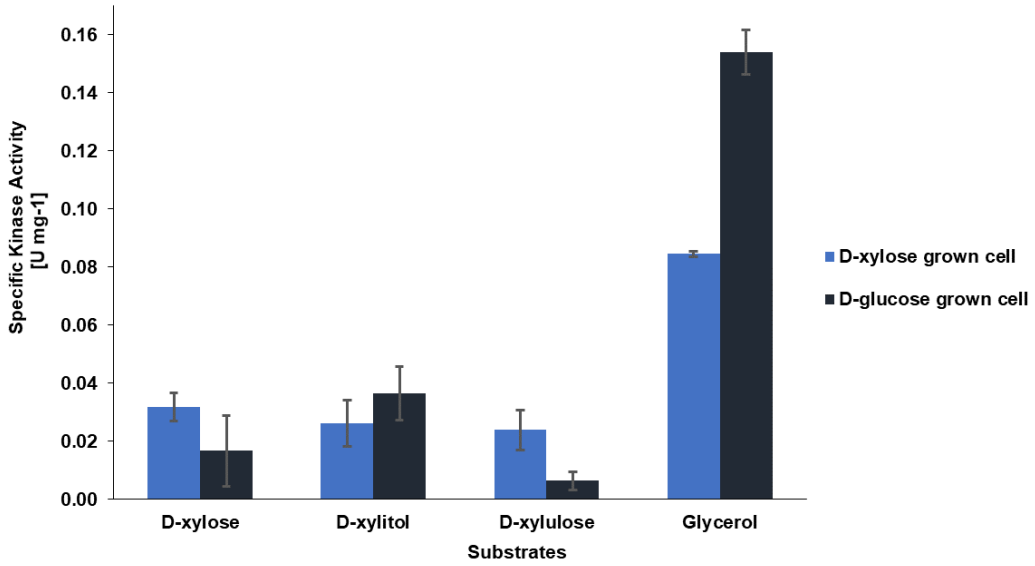


Figure 8: Specific kinase activity in crude extract from* Thermococcus *sp. 2319x1E cells. *The crude extract was freshly prepared from cells grown either on D-xylose (bright blue) or D-glucose (dark blue).*

V. List of Figures

VI. List of Abbrevations

ABPP	**Activity Based Protein Profiling**
ABPs	**Activity Based Probes**
ADP	**Adenosine diphosphate**
ATP	**Adenosine triphosphate**
CAZyme	**Carbohydrate active enzyme**
CBM	**Carbohydrate binding domaine**
DNA	**Deoxyribonucleic acid**
DTT	**Dithiothreitol**
EC	**Enzyme Commision**
e.g.	**for example**
E. coli	***Escherichia coli***
Fig.	**figure**
GH	**Glycoside hydrolase**
GT	**Glycoside transferase**
h	**hour(s)**
kcal	**kilo calories**
kDa	**kilodalton**
KEGG	**Kyoto Enzyclopedia of Genes and Genomes**
LC	**Liquid Chromatography**
LUCA	**Last Universal Common Ancestor**
MS	**Mass Spectronomy**
mg	**milligramm**
MES	**2-ethanesulfonic acid**
M	**molar**
mM	**millimolar**
ml	**milliliter**
NADH	**Nicotinamide adenine dinucleotide (reduced form)**
NAD^+	**Nicotinamide adenine dinucleotide (oxidized form)**
NADPH	**Nicotinamide adenine dinucleotide phosphate (reduced form)**

NCBI	**National Center for Biotechnology Information**
nm	**nanometer**
pH	**potential of hydrogen, acidity scale in aqueous solution**
PCR	**polymerase chain reaction**
PBS	**phosphate buffered saline**
PFAM	**Protein Families**
rpm	**rounds per minute**
RNA	**ribonucleic acid**
rRNA	**ribosomal ribonucleic acid**
S	**sedimentation coefficient**
sp.	**species**
U	**(enzyme) units**
° C	**degree celsius**
µg	**mikrogramm**
x g	**earth acceleration/ gravity**
XI	**Xylose Isomerase**